Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління

Кафедра Комп'ютерних інтелектуальних технологій та систем

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти другий (магістерський)

Прогнозування часових рядів на основі нейронних мереж прямого поширення

Виконав:

студент 2 курсу, групи КІТм-21-2

Гончаренко Д.В.

(прізвище, ініціали)

Спеціальність 123 Комп'ютерна інженерія

Тип програми освітньо-професійна

Освітня програма Комп'ютерні інтелектуальні технології

Керівник проф. О.Г. Руденко

(посада, ініціали, прізвище)

Допускається до захисту _____

(підпис)

Зав.кафедри _____ О.Г. Руденко

(підпис)

2022 р.

Харківський національний університет радіоелектроніки

Факультет	Комп'ютерної інженерії та управління
Кафедра	Комп'ютерних інтелектуальних технологій та систем
Рівень вищої освіти	другий (магістерський)
Спеціальність	123 Комп'ютерна інженерія
Тип програми	освітньо-професійна
Освітня програма	Комп'ютерні інтелектуальні технології

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«_____»_____2022 р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Гончаренко Д.В.
(прізвище, ініціали)

1. Тема роботи (проекту) «Прогнозування часових рядів на основі нейронних мереж прямого поширення нейронних мереж прямого поширення»
затверджена наказом університету від «16» листопада 2022р. № _____
2. Термін подання студентом роботи до екзаменаційної комісії ___ _____2022р.
3. Вихідні дані до роботи (проекту) Модель нейронної мережі прямого поширення для прогнозування часових рядів
4. Перелік питань, що потрібно опрацювати в роботі 1. Основні методи і моделі прогнозування. 2. Математичні моделі машинного навчання для прогнозування часових рядів. 3. Математичні моделі пронозування.

5. Перелік графічного матеріалу із зазначенням кресленików, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням кафедри)
Слайд-презентація, 12 слайдів, Додаток А

6. Консультанти розділів роботи (п.6 включається до завдання за наявністю консультантів згідно до наказу, зазначеному у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Огляд літератури за темою роботи	8.11.2022 – 29.11.2022	Виконано
2	Огляд основних методів і моделей	29.11.2022 – 07.12.2022	Виконано
	прогнозування		
3	Розгляд математичних моделей машинного	07.12.2022 – 12.12.2022	Виконано
	навчання для прогнозування часових рядів		
4	Розробка математичної моделі прогнозування	12.12.2022 – 16.12.2022	Виконано
5	Оформлення матеріалів кваліфікаційної	12.12.2022 – 17.12.2022	Виконано
	роботи		
6	Подання кваліфікаційної роботи керівникові	18.12.2022	Виконано
7	Подання кваліфікаційної роботи на рецензу-		
	вання	19.12.2022	Виконано

Дата видачі завдання 8 листопада 2022 р.

Студент _____
(підпис)

Керівник роботи _____ проф. О.Г.Руденко
(підпис) (посада, ініціали,прізвище)

РЕФЕРАТ

Магістерська кваліфікаційна робота містить 82 сторінок, на яких розміщено 34 рисунки та 4 таблиці. Перелік використаної літератури містить 19 найменувань.

ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ, НЕЙРОННА МЕРЕЖА, МЕНЕДЖЕР, ШТУЧНИЙ ІНТЕЛЕКТ, JAVA, SPRING

Метою написання магістерської дисертації є аналіз методів прогнозування часових рядів з застосуванням нейронних мереж прямого поширення.

Об’єктом дослідження магістерської дисертації є процес прогнозування часових рядів.

Предметом дослідження є часові ряди.

При написанні роботи були застосовані методи аналізу, порівняння, дослідження.

В роботі розглянуто можливість застосування нейронної мережі для прогнозування часових рядів. Виявлено основні переваги застосування нейронних мереж для задач прогнозування. Розглянуто основні архітектури нейронних мереж та обрано метод навчання та структуру мережі для прогнозування часових рядів.

Експериментально доведено, що застосування нейронних мереж прямого поширення для задач прогнозування часових рядів є ефективним.

ABSTRACT

Master's qualification work contains 82 pages, which contains 34 figures and 4 tables. The list of used literature contains 19 titles.

INTELLIGENT DECISION SUPPORT SYSTEMS, NEURAL NETWORK, MANAGER, ARTIFICIAL INTELLIGENCE, JAVA, SPRING.

The purpose of writing the master's thesis is to analyze the methods of forecasting time series using feed forward neural networks.

The object of study of the master's thesis is the process of forecasting time series.

The subject of research is time series.

Methods of analysis, comparison, research were used in writing the work.

The paper considers the possibility of using a neural network for time series forecasting. The main advantages of using neural networks for forecasting tasks are revealed. The basic architectures of neural networks are considered and the training method and network structure for time series forecasting are chosen.

It is experimentally proved that the use of feed forward neural networks for time series forecasting tasks is effective.

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління
Кафедра Комп'ютерних інтелектуальних технологій та систем

АНОТАЦІЯ
КВАЛІФІКАЦІЙНОЇ РОБОТИ

рівень вищої освіти другий (магістерський)

Прогнозування часових рядів на основі нейронних мереж прямого поширення

Виконав:
студент 2 курсу, групи КІТм-21-2
Гончаренко Д.В.
(прізвище, ініціали)

Спеціальність 123 Комп'ютерна інженерія
Тип програми освітньо-професійна
Освітня програма Комп'ютерні інтелектуальні технології
Керівник _____проф. О.Г. Руденко_____
(посада, ініціали, прізвище)

2022 р.

АНОТАЦІЯ

Гончаренко Д.В. Тема кваліфікаційної роботи – «Прогнозування часових рядів на основі нейронних мереж прямого поширення». Магістерська кваліфікаційна робота.

У роботі вирішено актуально проблему прогнозування часових рядів з застосування нейронних мереж прямого поширення.

Метою написання магістерської дисертації є аналіз методів прогнозування часових рядів з застосуванням нейронних мережі прямого поширення.

Задачі, які були поставлені при написанні роботи:

1. Виконати огляд основних методів і моделей прогнозування.

2. Розглянути математичні моделі машинного навчання для прогнозування часових рядів.

3. Розробити математичну модель прогнозування та навести приклад реалізації моделі для часових рядів на основі нейронної мережі прямого поширення.

Об’єктом дослідження магістерської дисертації є процес прогнозування часових рядів.

Предметом дослідження є часові ряди.

Розробка ефективних алгоритмів прогнозування часових рядів є актуальним напрямком досліджень. При цьому одними із найефективніших методів аналізу і прогнозування часових рядів є методи, які засновані на застосуванні нейронних мереж.

У першому розділі роботи виконано огляд основних методів і моделей прогнозування.

Встановлено, що метою аналізу часового ряду є побудова прогнозу його значень на майбутній період, а завданнями – виявлення компонент, під

впливом яких формуються значення часового ряду і побудова математичної моделі кожної компоненти чи його сукупності.

Встановлено основні методи і моделі прогнозування. Наприклад, для регресійних моделей та методів перевагами є: простота обчислень, гнучкість, прозорість моделювання (доступність для аналізу всіх проміжних обчислень), однаковість аналізу та проектування, наочність результатів прогнозування.

При цьому недоліками є: складність визначення функціональної залежності, трудомісткість знаходження коефіцієнтів залежності, відсутність можливості моделювання нелінійних процесів (для нелінійної регресії).

Перевагами авторегресійних моделей та методів є: простота, прозорість моделювання, однаковість аналізу та проектування, безліч прикладів застосування. Перевагами є: трудомісткість та ресурсомісткість ідентифікації моделей, неможливість моделювання нелінійностей, низька адаптивність.

Перевагами методів та моделей експоненціального згладжування є: простота моделювання, однаковість аналізу та проектування. При цьому серед недоліків виділяється недостатня гнучкість та вузька застосовність моделей.

Перевагами застосування нейромережевих методів та моделей є: нелінійність моделей, масштабованість, висока адаптивність, однаковість аналізу та проектування, безліч прикладів застосування. Серед переваг – складність вибору архітектури та жорсткі вимоги до навчальної вибірки.

Для ланцюгів Маркова виділяються наступні переваги: простота моделювання, однаковість аналізу та проектування методу, можливість простеження та вивчення причин зміни станів. Недоліками є: неможливість моделювання процесів із довгою пам'яттю, невисока точність прогнозування, вузька застосовність моделі.

За результатами огляду встановлено, що кожен метод прогнозування унікальний у своєму виконанні та реалізації. Найменшою кількістю

недоліків у порівнянні з перевагами мають нейронні мережі, саме тому їх актуально застосовувати для прогнозування часових рядів.

У другому розділі роботи розглянуто математичні моделі машинного навчання для прогнозування часових рядів.

Встановлено, що будь-яка нейронна мережа складається як мінімум із трьох шарів:

- вхідний шар – служить для розподілу даних по мережі та не робить жодних обчислень. Виходи цього шару передають тільки сигнали на входи наступного шару (прихованого або вихідного). Число вхідних елементів дорівнює числу значень, що вводяться в структуру;

- прихований шар – шар звичайних нейронів, які передають сигнали від входу до виходу. Їхнім входом служить вихід попереднього шару, а вихід - входом наступного шару. Прихованих шарів може бути і кілька, якщо цього вимагає завдання. Прихований шар називається так тому, що його елементи не одержують дані від зовнішнього середовища безпосередньо і не надсилають дані безпосередньо до зовнішнього середовища;

- вихідний шар – шар, який видає результат розрахунків усієї нейронної мережі. Число вихідних елементів дорівнює числу значень, що подаються цією структурою на вихід у зовнішнє середовище.

Принцип навчання нейронної мережі - кожен шар посилає вихідний сигнал лише на наступні шари та приймає вхідні сигнали лише від попередніх шарів. Вхідний вектор подається на нижній шар, а вихідний вектор визначається шляхом почергового обчислення рівнів активності елементів кожного шару (знизу-вгору) з використанням відомих значень активності елементів попередніх шарів. Отриманий вихідний вектор порівнюється з потрібним. Якщо вони збігаються, навчання не відбувається. В іншому випадку обчислюється різниця між фактичними та необхідними вихідними значеннями, яка передається послідовно від вихідного шару до вхідного. Розбіжність між тим, що дасть мережа, і тим, що для цього

навчального набору має бути отримано насправді, є помилкою, яка може використовуватися для коригування ваг.

Для проведення порівняльного аналізу розглянутих алгоритмів навчання штучних нейронних мереж використовувалося три критерії: середня абсолютна похибка навчання, кількість епох навчання та кількість нейронів у прихованому шарі.

За результатами цього дослідження найменша похибка навчання була виявлена при навчанні нейронної мережі методом Левенберга-Маркварда. Саме тому цей метод був обраний для навчання нейронної мережі.

Виходячи з завдання на виконання магістерської дисертації, остаточно прийнято рішення застосовувати багатошаровий персептрон, тому що ~~так як~~ він не має недоліків, які унеможливлюють застосовувати даний тип архітектури для прогнозування часових рядів.

В третьому розділі роботи представлено розробку математичної моделі прогнозування та наведено приклад реалізації моделі для часових рядів на основі нейронної мережі прямого поширення.

В якості вихідних даних для прогнозування було використано кількість отриманих повідомлень, що надійшли на обслуговуючий пристрій.

Проведене нормування вхідних даних та встановлено, що у разі лінійного, нелінійних перетворень розподілу даних належать до нормального закону. Отже, нормовані дані можна використовувати для прогнозування.

Також встановлено, що вибір початкових вагових коефіцієнтів ґрунтується на мінімізації кількості епох навчання нейронної мережі та нейронів прихованого шару. Значення початкових вагових коефіцієнтів варіюються в проміжку (-1;1).

В якості функції активації прихованого шару обрано логістичну сигмоїдальну функцію, оскільки ця функція диференційована на всій осі абсцис і має властивість посилювати слабкі сигнали краще, ніж сильні, тим самим запобігаючи насиченню від сильних сигналів. В якості функції

активації вихідного шару використано лінійну функцію, у зв'язку з тим, що вона дає змогу виконувати дії з безперервними сигналами в діапазоні (-∞, ∞).

Визначено, що найменше значення похибки навчання спостерігається за значення k=0.1 і кількості нейронів у вихідному шарі, яка дорівнює 11.

Прихований шар нейронної мережі складається з 24 нейронів (за кількістю входів (за значеннями часового ряду, години).

Для реалізації нейронної мережі з метою прогнозування використовувалася універсальна інтегрована система STATISTICA Neural Networks. За значенням коефіцієнта детермінації 0,993 можна зробити висновок, що прогноз є високоточним.

Зроблено висновок, що прогнозування з використанням штучних нейронних мереж є найбільш достовірним. Проте було також відзначено, що при цьому витрачається велика кількість ресурсів. У висновку вказано, що в роботі розглянуто класифікацію методів прогнозування часових рядів, описано найпоширеніші моделі прогнозування, а також способи їх реалізації. Проведено аналіз описаних моделей прогнозування часових рядів, унаслідок чого визначено переваги та недоліки кожної з моделей.

Викладено етапи побудови нейронних мереж, правила вибору архітектури мережі та обробки даних, методи навчання, а також функції активації нейронів для прогнозованої мережі. Встановлено, що використання в прогнозуванні штучних нейронних мереж цілком обґрунтоване високим значенням коефіцієнта детермінації, порівняно з іншими поширеними методами прогнозування.

НЕЙРОННА МЕРЕЖА, МЕРЕЖА ПРЯМОГО ПОШИРЕННЯ, АРХІТЕКТУРА, ПРИХОВАНИЙ ШАР, ВИХІДНИЙ ШАР, МАШИННЕ НАВЧАННЯ, АЛГОРИТМ, ЧАСОВІ РЯДИ, ПРОГНОЗУВАННЯ, МАТЕМАТИЧНА МОДЕЛЬ, КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ.

Публікації здобувача за темою роботи:

1. Гончаренко Д.В. Властивості штучних нейромереж / Денис Валентинович Гончаренко// Вінниця: *ГО «Європейська наукова платформа»*. – 2022. – С. 31–32.

Використані в роботі публікації керівника та співробітників кафедри:

1. Руденко О. Г. Искусственные нейронные сети: архитектуры, обучение, применения / О. Г. Руденко, Е. В. Бодянский. – Харьков: ТЕЛЕТЕХ, 2004. – 370 с.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 14
ВСТУП 15
1 ОГЛЯД ОСНОВНИХ МЕТОДІВ ТА МОДЕЛЕЙ ПРОГНОЗУВАННЯ 16
1.1 Аналіз і методи прогнозування часових рядів 16
1.2 Регресійні моделі 19
1.3 Авторегресійні моделі 21
1.4 Моделі експоненціального згладжування 24
1.5 Моделі на основі нейронних мереж 24
1.6 Ланцюги Маркова 25
1.7 Переваги та недоліки методів прогнозування. Переваги застосування нейронних мереж для прогнозування часових рядів 27
2 МАТЕМАТИЧНІ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ 30
2.1 Архітектура нейронної мережі 30
2.2 Методи навчання нейронних мереж 35
2.3 Вибір архітектури і методу навчання нейронної мережі для прогнозування часових рядів 39
3 РОЗРОБКА НЕЙРОННОЇ МЕРЕЖІ ПРЯМОГО ПОШИРЕННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ 60
3.1 Математична модель нейронної мережі 60
3.2 Структура нейронної мережі прямого поширення для прогнозування часових рядів 63
3.3 Приклад реалізації моделі 72
ВИСНОВКИ 77
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ 78
ДОДАТОК А 79

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ЕС – Експертна система

ІА – Інтелектуальний агент

ІСППР – Інтелектуальна система підтримки прийняття рішень

НМ – Нейронна мережа

ОПР – Особа, яка приймає рішення

СППР – Система підтримки прийняття рішень

ШІ – Штучний інтелект

FF – Feed-Forward Neural Network, нейронна мережа із прямим зв’язком

ВСТУП

Проблема дослідження часових рядів є актуальною для всіх областей науки та практики. Основне завдання при цьому полягає в порівняльному аналізі методик виявлення основних компонентів часового ряду, щоб більш точно описати вихідну статистичну інформацію динамічного ряду.

Статистика займається вивченням змін явищ у часі, аналізуючи часові лави. Аналіз часових рядів – це математичний розділ статистики, у якому здійснюється аналіз даних, отриманих при послідовному вимірі будь-якої величини, та зроблених через певні проміжки часу.

Розробка ефективних алгоритмів прогнозування часових рядів є актуальним напрямком досліджень. При цьому одними із найефективніших методів аналізу і прогнозування часових рядів є методи, які засновані на застосуванні нейронних мереж.

Метою написання магістерської дисертації є аналіз методів прогнозування часових рядів з застосуванням нейронних мережі прямого поширення.

При написанні роботи були поставлені наступні задачі:

1. Виконати огляд основних методів і моделей прогнозування.

2. Розглянути математичні моделі машинного навчання для прогнозування часових рядів.

3. Розробити математичну модель прогнозування та навести приклад реалізації моделі для часових рядів на основі нейронної мережі прямого поширення.

Об’єктом дослідження магістерської дисертації є процес прогнозування часових рядів.

Предметом дослідження є часові ряди.

При написанні роботи були застосовані методи аналізу, порівняння, дослідження.

1 ОГЛЯД ОСНОВНИХ МЕТОДІВ ТА МОДЕЛЕЙ ПРОГНОЗУВАННЯ

1.1 Аналіз і методи прогнозування часових рядів

Часовий ряд – це сукупність числових значень певного статистичного показника у послідовні моменти чи періоди часу.

Елементи, з яких складається часовий ряд [1]:

- моменти або періоди часу, до яких відносяться наведені статистичні дані про явище, що вивчається;

- статистичні показники, що характеризують це явище за певний період часу.

Рівні ряду - це статистичні показники, які характеризують явище, що досліджується. Вид ряду динаміки залежить не тільки від характеру показників, які оцінюють досліджуване явище, а й від того, наводиться показник у період чи момент часу. Статистичні показники можуть бути абсолютними, середніми або відносними [2].

Метою аналізу часового ряду є побудова прогнозу його значень на майбутній період, а завданнями – виявлення компонент, під впливом яких формуються значення часового ряду і побудова математичної моделі кожної компоненти чи його сукупності.

Метод прогнозування – певний набір дій, у результаті якого, визначається модель прогнозування конкретного часового ряду.

Модель прогнозування – функціональне уявлення, яке адекватно описує досліджуваний процес і є основою отримання його майбутніх значень.

Прийом прогнозування – математична чи логічна операція, яка дозволяє отримати конкретні результати у процесі розробки прогнозів.

Методи прогнозування, як показано на рисунку 1.1, умовно можна розділити на дві групи: інтуїтивні та формалізовані.

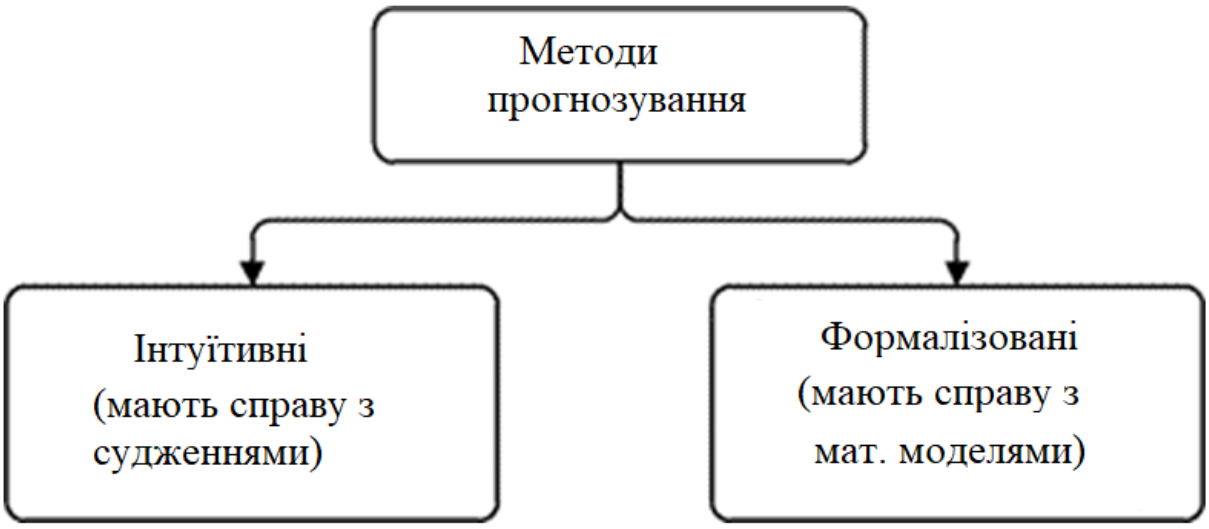


Рисунок 1.1 – Класифікація методів прогнозування

Інтуїтивні методи прогнозування – це судження та оцінка експертів. Якщо об'єкт занадто простий або навпаки занадто складний та такий, для якого аналітично неможливо врахувати вплив багатьох чинників, застосовується інтуїтивне прогнозування. Тобто опитують експертів та отримані експертні оцінки як індивідуальні, так і колективні використовуються як кінцеві прогнози або як вихідні дані в комплексних системах прогнозування. Цей метод активно застосовується у маркетингу, економіці, політиці [3].

Методи ґрунтуються на виявленні та узагальненні різних думок експертів (фахівців, науковців, споживачів та інших людей, які мають достатні знання та кваліфікацію у сфері, що вивчається, а також можуть інтуїтивно відчувати правильне рішення). Експертні оцінки, звичайно, суб'єктивні, оскільки засновані на суб'єктивних методах пізнання.

Рівень точності, надійності та достовірності підсумкових оцінок багато в чому залежить від кількості опитаних експертів, ступеня їхньої компетентності у сфері, методів проведення експертизи, узгодженості різних думок та інших факторів. Методи експертних оцінок виявляються найбільш

ефективними, коли досліджуваний процес важко формалізувати та описати кількісно, а також коли тимчасові та фінансові ресурси дуже обмежені.

Експертні оцінки поділяються на індивідуальні та колективні, залежно від кількості учасників.

Індивідуальну оцінку виконує один фахівець, причому обсяг роботи може бути досить великим і кропітким.

Колективну оцінку виконує група фахівців, що дає об'єктивніші результати, але виникає проблема сумісності висновків експертів та питання про отримання загального висновку комісії експертів.

Формалізовані методи прогнозування – це методи прогнозування, у яких будують моделі прогнозування, тобто визначають таку математичну залежність, яка дозволяє зробити прогноз, тобто обчислити майбутнє значення процесу [3].

Моделі для прогнозування часових рядів поділяють на статистичні та структурні.

Функціональна залежність у статистичних моделях між фактичними та майбутніми значеннями часового ряду, а також зовнішніми факторами задається аналітично.

До статистичних моделей належать:

- регресійні моделі;
- авторегресійні моделі;
- моделі експонентного згладжування

Функціональна залежність у структурних моделях між фактичними та майбутніми значеннями часового ряду, а також зовнішніми факторами задається структурно.

До структурних моделей відносять:

- нейромережеві моделі;
- моделі на базі ланцюгів Маркова;
- моделі на базі класифікаційно-регресійних дерев.

Нижче вказано та коротко описано основні, найбільш поширені моделі прогнозування часових рядів.

1.2 Регресійні моделі

Регресійні моделі дозволяють вирішити завдання, які вимагають вивчення відношень між двома та більш змінними.

Мета даного аналізу: визначення залежності між вихідною змінною та множиною зовнішніх факторів (регресорів). При цьому коефіцієнти регресії можуть визначатися як за методом найменших квадратів, так і методом максимальної правдоподібності [4].

Лінійна регресія є найпростішим варіантом регресійної моделі. В основу моделі покладено припущення, що існує дискретний зовнішній фактор *X(t)*, який впливає на досліджуваний процес *Z(t)* і при цьому зв'язок між процесом та зовнішнім фактором є лінійним.

Модель прогнозування описується рівнянням:

$$Z(t) = \alpha_0 + \alpha_1 X(t) + \varepsilon_t, \tag{1.1}$$

де: α_0, α_1 – коефіцієнти регресії;

ε – помилка моделі прогнозування;

X(t) – значення зовнішнього чинника на момент часу t.

Для отримання прогнозних значень *Z(t)* у момент часу t необхідно мати значення *X(t)* у той самий момент часу *t*, що рідко можна здійснити на практиці.

Множинна регресійна модель дозволяє провести аналіз зв'язку між декількома незалежними змінними (регресорами) та залежною змінною [5].

Модель прогнозування описується рівнянням:

$$Z(t) = \alpha_0 + \alpha_1 X_1(t) + \alpha_2 X_2(t) + \cdots + \alpha_s X_s(t) + \varepsilon_t, \tag{1.2}$$

де α – коефіцієнти регресії;

ε – помилка моделі прогнозування;

X(t) – значення зовнішнього чинника на момент часу t.

Головним недоліком даної моделі є те, що для обчислення майбутнього значення процесу *Z(t)* потрібно знати майбутні значення всіх чинників $X_1(t),\ldots,X_s(t)$, що на практиці майже неможливо [5].

Нелінійна регресійна модель дозволяє описати залежність між вихідним процесом *Z(t)* та зовнішнім фактором *X(t)* за допомогою рівняння (1.3):

$$Z(t) = F(X(t), A). \tag{1.3}$$

Щоб побудувати модель прогнозування, необхідно визначити параметри функції А.

Приклад нелінійної регресійної моделі рівняння:

$$Z(t) = \alpha_0 + \alpha_1 \cos(X(t)) + \varepsilon_t, \tag{1.4}$$

де α_0, α_1 – коефіцієнти регресії;

ε – помилка моделі прогнозування;

$X(t)$ – значення зовнішнього чинника на момент часу t;

$Z(t)$ – вихідний процес.

Для того, щоб побудувати модель, достатньо визначити параметри $A=[\alpha_0, \alpha_1]$ але на практиці рідко зустрічаються процеси, для яких вид функціональної залежності між процесом *Z(t)* та зовнішнім фактором *X(t)* наперед відомий. Саме тому нелінійні регресійні моделі на практиці застосовуються рідко.

1.3 Авторегресійні моделі

В основу моделі покладено припущення, що значення процесу *Z(t)* лінійно залежить від деякої кількості попередніх значень того ж процесу *Z(t-1),…,Z(t_p)*.

Рівняння для авторегресійної моделі:

$$Z(t) = \alpha_\infty + \alpha_1 Z(t-1) + \alpha_2 Z(t-2) + \cdots + \alpha_p Z(tp) + \varepsilon_1, \quad (1.5)$$

де α – коефіцієнти регресії;

Z(t_p) – значення процесу *Z(t)* у минулих періодах регресії;

ε – помилка моделі прогнозування.

Основні види авторегресійних моделей прогнозування:

- модель ковзного середнього;
- модель з умовною гетероскедастичністю;
- модель із розподіленим лагом.

Модель авторегресії (autoregressive, AR) і модель ковзного середнього (moving average, MA) в області аналізу часових рядів є одними з найбільш використовуваних [6].

Часто, для того, щоб досягти більшої гнучкості, з'єднують в одній моделі авторегресію і ковзне середнього. Така модель позначається ARMA(*p*, *q*) і з'єднує у собі фільтр як ковзного середнього порядку «*q*» і авторегресію фільтрованих значень процесу порядку «*p*».

Якщо в якості вхідних даних використовується різниця значень часового ряду *d*-того порядку, то така модель називається авторегресією проінтегрованого ковзного середнього *ARIMA (p, d, q)*.

Модель *ARIMAX (p, d, q)* є розвитком моделі *ARIMA (p, d, q)*, визначається за рівнянням:

$$Z(t) = AR(p) + \alpha_1 X_1(t) + \cdots + \alpha_s X_s(t), \quad (1.6)$$

де $\alpha_1,\ldots,\alpha_s$ – коефіцієнти зовнішніх факторів $X_1(t),\ldots,X_s(t)$

Процес $Z(t)$ є наслідком моделі $MA(q)$, тобто відфільтрованими значеннями вихідного процесу. Для подальшого прогнозування $Z(t)$ застосовують модель авторегресії, в яку вводять додаткові регресори зовнішніх факторів $X_1(t),\ldots,X_s(t)$ [5].

В основу моделі з умовною гетероскедастичністю покладено припущення, що умовне значення дисперсії залежить як від попередніх значень ряду, так і від попередніх значень дисперсії.

Абревіатура ARCH перекладається як авторегресійна умовна гетероскедастичність, тобто зміна в часі дисперсії випадкових величин (збурень), які формалізовано можна представити таким чином:

$$a_{t+1} = y_{t+1} - \mu_y(y_1, y_2, y_3, \ldots y_t), \tag{1.7}$$

де a_{t+1} – збурення (випадкова величина) в (t+1) період часу;

y_{t+1}– значення показника в (t+1) період часу;

$\mu_y(y_1, y_2, y_3, \ldots y_t)$ – середнє значення показника, оцінене на основі даних, які передують моменту (t + 1).

Абревіатура GARCH перекладається як узагальнена авторегресійна модель з умовною гетероскедастичністю.

У GARCH моделях на відміну від ARCH моделей у рівнянні дисперсії враховано, крім лагових змінних випадкових величин, ще й лагові змінні умовної дисперсії.

Оскільки ARCH/GARCH вважаються моделями дисперсії, середнє значення показника можна моделювати на основі різних підходів, зокрема, простим середнім, лінійною регресійною моделлю, ARIMA, SARIMA моделями тощо.

Власне загальну ARCH/ GARCH модель можемо представити як послідовність фільтрів, яка представлена на рисунку 1.2 [6].

Загалом, ARCH/GARCH модель можна вважати «доповненням» як до одно-, багатофакторної лінійної регресії, так і до ARIMA моделей [54]. Однак, оцінювання ACRH/GARCH моделей неможливо здійснити методом найменших квадратів (МНК).

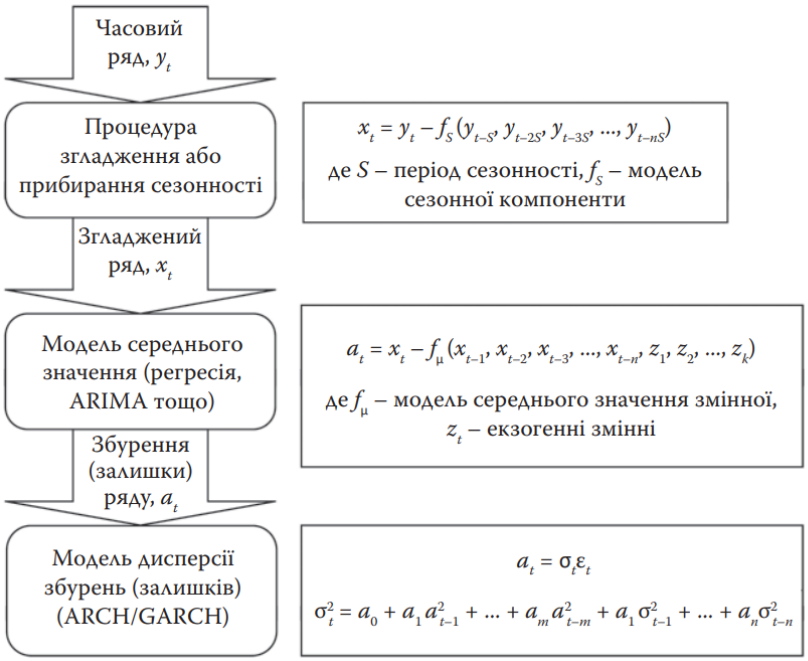


Рисунок 1.2 – Узагальнений алгоритм послідовної реалізації ARCH/GARCH моделі

Для їх оцінювання розроблено спеціальні процедури, зокрема метод максимальної вірогідності [6].

При моделюванні процесів на змінну, що вивчається, часто впливають не тільки поточні значення процесу, але і значення часового ряду, що передують моменту часу, тобто лаги.

ARDLM(p,l) – модель авторегресії розподіленого лага, визначається за рівнянням [7]:

$$Z(t) = \varphi_0 + \varphi_1 Z(t-l-1) + \varphi_2 Z(t-l-2) + \cdots + \varphi_p Z(t-l-p) + \varepsilon_t, \quad (1.8)$$

де $\varphi_0, \ldots, \varphi_p$ – коефіцієнти моделі

l-величина лага

Найчастіше ця модель застосовується для моделювання економічних процесів.

1.4 Моделі експоненціального згладжування

Ідея експоненціального згладжування полягає в тому, що відбувається постійний перегляд спрогнозованих значень у міру надходження фактичних. Ця модель привласнює ваги спостереженням у міру їх старіння. Тобто останні спостереження мають більший вплив на прогнозне значення, ніж старші спостереження.

Функція має вигляд:

$$Z(t) = S(t) + \varepsilon_t; \tag{1.9}$$

$$S(t) = \alpha * Z(t-1) + (1+\alpha) * S(t-1) + \varepsilon_t, \tag{1.10}$$

де α - коефіцієнт згладжування, $0<\alpha<1$,

$S(1)=Z(0)$.

Кожне згладжене значення $S(t)$ є зваженим середнім між попереднім значенням часового ряду $Z(t)$ і попереднього значення $S(t-1)$.

1.5 Моделі на основі нейронних мереж

Система, що складається з нейронів і здатна змінювати свою структуру під впливом зовнішніх факторів, називається нейронною мережею.

Нейрон є перемикачем, який отримує та передає імпульси або сигнали. Активований нейрон - це той нейрон, який отримав досить сильний імпульс та передає його зв'язаним нейронам. Ще не активований нейрон - це той нейрон, який не передає імпульс і залишається в стані спокою.

Компоненти нейрона:

- синапси – з'єднують нейрони з іншими нейронами та одержують імпульси від сусідніх нейронів;

- аксон – передає імпульс іншим нейронам;

- дендрит – отримує сигнали з різних джерел та від синапсів у тому числі.

Кожен імпульс, що надходить, множиться на відповідну вагу з вектора ваг $W(w_1,w_2,\ldots,w_m)$, аналогічний синаптичній силі. Далі підсумовуються всі добутки, визначаючи рівень активації нейрона.

Нейрон або не активується, або активується і передає сигнал далі іншим нейронам у мережі, це залежить від функції активації та порога активації. Поєднуючи такі нейрони, одержують нейронну мережу.

Нейронні мережі дозволяють моделювати нелінійну залежність майбутнього значення тимчасового ряду від його фактичних значень та від значень зовнішніх факторів [8], за допомогою навчання її на вже наявних даних.

1.6 Ланцюги Маркова

Більшість складних систем, що розвиваються в часі, доцільно розцінювати випадковими процесами, розвиток яких залежить від ряду випадкових факторів, що супроводжують ці зміни.

Для обчислення числових параметрів, що характеризують такі системи, зручно побудувати ймовірну модель явища, що враховує випадкові фактори, що впливають на нього.

Нехай є деяка система *S*, що змінюється протягом часу. Якщо стан системи S змінюється у часі мимовільним чином, то системі протікає випадковий процес [9].

Оптимально описувати виникнення випадкових подій у вигляді ймовірностей переходів з одного стану системи до іншого, оскільки при

переході між станами система не повинна враховувати, як потрапила до цього стану.

Випадковий процес вважається Марківським, якщо у кожний момент часу t ймовірність довільного стану системи в майбутньому пов'язана лише з її станом у теперішньому і не враховує, як система прийшла у цей стан. Такий процес представляється у формі графів переходу з одного стану (S_n) в інший стан S_{n+1}.

Однією з форм Марківських процесів зі стаціонарними ймовірностями переходу з дискретним часом на кінцевому фазовому просторі називають однорідним ланцюгом Маркова. Характеризується наявністю декількох ймовірностей переходу з одного стану до іншого, які складають матрицю ймовірностей переходу.

На рисунку 1.3 зображено ланцюг Маркова для процесу з трьома станами.

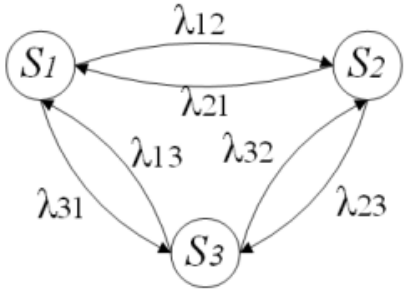


Рисунок 1.3 – Ланцюг Маркова з трьома станами

S_1,S_2,S_3 – події в яких може виявитися процес Z(t);

λ_{13} – імовірність переходу процесу Z(t) з події S_1 в подію S_2.

Сума всіх ймовірностей переходів повинна дорівнювати 1. Далі на підставі значень ймовірностей переходів будується матриця перехідних ймовірностей.

Стани змінюються від кроку до кроку, тобто система з часом переходить у новий стан з певною ймовірністю. Зміни станів відбуваються випадково, тобто має місце випадковий процес у дискретному часі. Таким чином, ланцюг Маркова визначає залежність між поточним значенням процесу та його майбутнім значенням

1.7 Переваги та недоліки методів прогнозування. Переваги застосування нейронних мереж для прогнозування часових рядів

Основні переваги та недоліки описаних методів прогнозування часових рядів зведемо в таблицю 1.1.

Таблиця 1.1 – Переваги та недоліки методів та моделей прогнозування

Модель та метод прогнозування	Переваги	Недоліки
1	2	3
Регресійні моделі та методи	1. Простота обчислень 2. Гнучкість 3. Прозорість моделювання (доступність для аналізу всіх проміжних обчислень) 4. Однаковість аналізу та проектування 5. Наочність результатів Прогнозування	1. Складність визначення функціональної залежності 2. Трудомісткість знаходження коефіцієнтів залежності; 3. Відсутність можливості моделювання нелінійних процесів (для нелінійної регресії)
Авторегресійні моделі та методи	1. Простота 2. Прозорість моделювання	1. Трудомісткість та ресурсомісткість

	3. Однаковість аналізу та проектування 4. Безліч прикладів застосування	ідентифікації моделей 2. Неможливість моделювання нелінійностей 3. Низька адаптивність

Продовження таблиці 1.1

Методи та моделі експоненціального згладжування	1. Простота моделювання 2. Однаковість аналізу та проектування	1. Недостатня гнучкість 2. Вузька застосовність моделей
Нейромережні методи та моделі	1. Нелінійність моделей 2. Масштабованість 3. Висока адаптивність 4. Однаковість аналізу та проектування 5. Безліч прикладів застосування	1. Складність вибору архітектури 2. Жорсткі вимоги до навчальної вибірки
Ланцюги Маркова	1. Простота моделювання 2. Однаковість аналізу та проектування методу 3. Можливість простеження та вивчення причин зміни станів	1. Неможливість моделювання процесів із довгою пам'яттю 2. Невисока точність прогнозування 3. вузька застосовність моделі

Таким чином, кожен метод прогнозування унікальний у своєму виконанні та реалізації. Найменшою кількістю недоліків у порівнянні з

перевагами мають нейронні мережі, саме тому їх актуально застосовувати для прогнозування часових рядів.

2 МАТЕМАТИЧНІ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

2.1 Архітектура нейронної мережі

Нейронна мережа - математична модель, і навіть її програмне чи апаратне втілення, побудоване за принципом організації та функціонування біологічних нейронних мереж - мереж нервових клітин живого організму.

Один з найпривабливіших аспектів використання нейронних мереж полягає в тому, що хоча елементи такої мережі мають дуже обмежені обчислювальні можливості, вся мережа в цілому, поєднуючи велику кількість таких елементів, виявляється здатною виконувати досить складні завдання [11].

Нейронна мережа є сукупністю елементів, з'єднаних так, щоб між ними забезпечувалася взаємодія. Ці елементи, звані нейронами або вузлами, є простими процесорами, обчислювальні можливості яких зазвичай обмежуються деяким правилом комбінування вхідних сигналів і правилом активізації, що дозволяє обчислити вихідний сигнал по сукупності вхідних сигналів, як показано на рисунку 2.1.

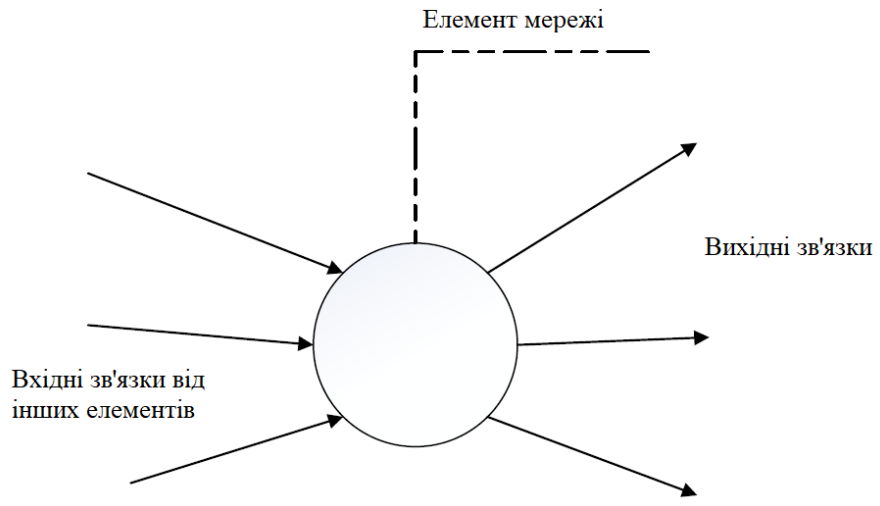


Рисунок 2.1 – Окремий елемент нейронної мережі

Вихідний сигнал елемента може посилатися іншим елементам зваженими зв'язками, з кожним з яких пов'язаний ваговий коефіцієнт або вага. Залежно від коефіцієнта сигнал, що передається, або посилюється, або послаблюється.

Структурна схема нейрона наведена на рисунку 2.2.

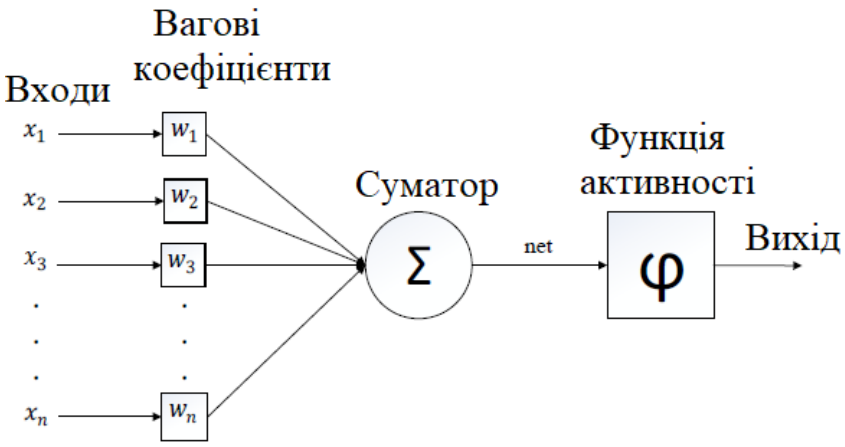


Рисунок 2.2 – Структурна схема нейрона

Структурно нейрон складається з наступних елементів:

- вхідні сигнали x_n;
- вагові коефіцієнти w_n;
- суматор Σ та його вихід net;
- функція активності φ;
- вихідний сигнал.

На вхід нейрона подається набір вхідних сигналів x_1, x_2, ..., x_n, кожен вхідний сигнал множиться на відповідну вагу зв'язку w_1, w_2, …, w_n. Вага зв'язку є позитивною для збудливого та негативною для гальмівного зв'язку нейрону. Вхідні сигнали, зважені за допомогою ваг зв'язків, надходять на блок підсумовування, де обчислюється їх сума і визначається рівень збудження нейрона (активація) S.

Структура зв'язків представляється у вигляді вагової матриці, в якій кожен елемент є величиною вагового коефіцієнта для зв'язку, що йде від одного елемента до іншого.

Для опису структури зв'язків можна використовувати не одну, а кілька вагових матриць, якщо елементи мережі виявляються згрупованими шарами.

Матриця ваг є пам'яттю мережі, що зберігає інформацію про те, як має виконуватися завдання.

Функція активності - це нелінійна функція, що обчислює вихідний сигнал формального нейрона.

Часто використовуються такі функції: гранична функція; лінійна; сигмоїдальна. Вибір функції визначається специфікою поставленого завдання чи обмеженнями, накладаними деякими алгоритмами навчання [12].

Види функцій активації.

1. Тотожна функція.

Функція активності для вхідних елементів може бути тотожною функцією, і це означає, що значення активності, тобто сигнал на виході, виявляється точно таким же, як і сигнал на вході нейрона. Вхідні елементи зазвичай призначені для розподілу сигналів, що вводяться між іншими елементами мережі, тому для вхідних елементів зазвичай потрібно, щоб вихідний від елемента сигнал був таким же, як і вхідний. На відміну від інших елементів мережі, вхідні елементи мають лише одне вхідне значення. Кожен з цих елементів зв'язується з багатьма іншими елементами мережі, тому дані, отримані від одного датчика, виявляються розподіленими між багатьма елементами мережі. Найчастіше, оскільки вхідні елементи призначені виключно для того, щоб розподіляти сигнали, що отримуються із зовнішнього середовища, багато дослідників взагалі не вважають вхідні елементи частиною нейронної мережі [13].

2. Порогова функція.

У більшості моделей нейронних мереж використовують нелінійні функції активності. Порогова функція обмежує активність значеннями 1 або

0 в залежності від значення вхідних даних у порівнянні з деякою граничною величиною θ.

Для зручності можна відняти граничне значення (зване зміщенням або зрушенням) вхідних даних і розглянути граничну функцію в її математично еквівалентній формі.

Зсув W_0 буде негативним, а значення на вході обчислюватиметься за формулою [14]:

$$net_j = \omega_0 + \sum_{i=1}^{n} x_i \omega_{ij}. \quad (2.1)$$

Зсув зазвичай означає зв'язок, що походить від елемента, активність якого завжди дорівнює 1. Вхідні дані можна подати у вигляді:

$$net_j = \sum_{i=1}^{n} x_i \omega_{ij}. \quad (2.2)$$

де x_0 завжди вважається рівним 1.

Сигмоїдальна функція. Вихідні значення такої функції безперервно заповнюють діапазон від 0 до 1. Прикладом може бути логістична функція:

$$f(net) = \frac{1}{1+\exp(-net)}. \quad (2.3)$$

Області вихідних значень логістичних функцій можуть різнитися.

Наприклад, для біполярного сигмоїда областю вихідних значень є діапазон (-1; 1).

Зазвичай, будь-яка нейронна мережа складається як мінімум із трьох шарів:

- вхідний шар – служить для розподілу даних по мережі та не робить жодних обчислень. Виходи цього шару передають тільки сигнали на входи наступного шару (прихованого або вихідного). Число вхідних елементів дорівнює числу значень, що вводяться в структуру;

- прихований шар – шар звичайних нейронів, які передають сигнали від входу до виходу. Їхнім входом служить вихід попереднього шару, а вихід - входом наступного шару. Прихованих шарів може бути і кілька, якщо цього вимагає завдання. Прихований шар називається так тому, що його елементи не одержують дані від зовнішнього середовища безпосередньо і не надсилають дані безпосередньо до зовнішнього середовища;

- вихідний шар – шар, який видає результат розрахунків усієї нейронної мережі. Число вихідних елементів дорівнює числу значень, що подаються цією структурою на вихід у зовнішнє середовище.

В якості прикладу на рисунку 2.3 представлена структурна схема багатошарової нейронної мережі [13].

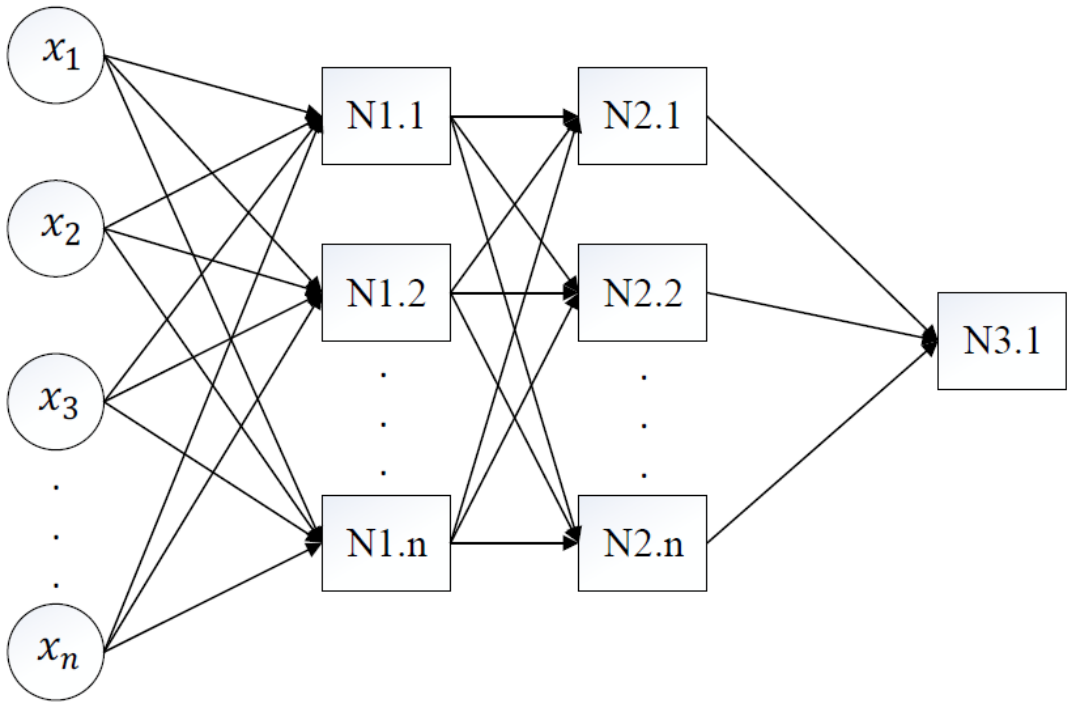


Рисунок 2.3 – Структурна схема багатошарової нейронної мережі

Перевага використання нейронної мережі полягає в тому, що нейромережа не обмежена знаннями, які заклав у неї програміст. Вони мають

змогу навчатися на попередніх подіях. За рахунок цього досягається висока ефективність та адаптивність.

2.2 Методи навчання нейронних мереж

Для того, щоб мережа успішно виконувала поставлені перед нею завдання, потрібно її навчити.

Принцип навчання нейронної мережі - кожен шар посилає вихідний сигнал лише на наступні шари та приймає вхідні сигнали лише від попередніх шарів. Вхідний вектор подається на нижній шар, а вихідний вектор визначається шляхом почергового обчислення рівнів активності елементів кожного шару (знизу-вгору) з використанням відомих значень активності елементів попередніх шарів. Отриманий вихідний вектор порівнюється з потрібним. Якщо вони збігаються, навчання не відбувається. В іншому випадку обчислюється різниця між фактичними та необхідними вихідними значеннями, яка передається послідовно від вихідного шару до вхідного. Розбіжність між тим, що дасть мережа, і тим, що для цього навчального набору має бути отримано насправді, є помилкою, яка може використовуватися для коригування ваг [13].

Прикладом правила корекції помилок є дельта-правило, яке називається також правилом Відроу-Хоффа. Припустимо, що вихідний елемент має активність y, а справжній вивід має дорівнювати t. Тоді помилка δ задається такою формулою:

$$\delta = t - y. \tag{2.4}$$

Таким чином, процес навчання полягає у підстроюванні її внутрішніх параметрів під конкретне завдання за певним алгоритмом [13].

За допомогою функції помилок можна оцінити якість роботи нейронної мережі під час навчання. Від якості навчання нейронної мережі залежить її здатність вирішувати поставлені перед нею завдання.

Іноді необхідно виконувати попередню підготовку даних.

Така попередня підготовка може бути дуже простою, - наприклад, переклад за допомогою масштабування значень всіх ознак (тобто змінних) в діапазон від 0 до 1, - а може включати використання і складніших статистичних процедур. Однак варто наголосити, що довгостроковою метою розробки нейронних мереж є мінімізація необхідності прямого впливу розробника на процес знаходження рішення, оскільки головною перевагою нейронних мереж є їх потенційна можливість виробляти власні рішення.

Насправді найкращі результати виходять тоді, коли є чітке розуміння аналізованої проблемної галузі знань і концептуальне розуміння проблем побудови нейронної мережі.

Дані, що використовуються для навчання нейронної мережі, зазвичай поділяються на дві категорії: одні використовуються для навчання, інші – для тестування. Реальні якості нейронної мережі виявляються лише стадії тестування, оскільки успішне завершення навчання мережі має означати відсутність ознак неправильної роботи мережі під час її тестування.

Процес тестування розробляється те щоб у його ході для цієї мережі можна було б оцінити її здатність узагальнювати отримані знання. Узагальнення в даному випадку означає здатність мережі правильно виконувати завдання з даними, які виявляються хоча і аналогічними даними, що пред'являлися мережі в процесі навчання, але все ж таки відмінними від них.

Навчання нейронної мережі з учителем передбачає, що для кожного вхідного вектора з навчальної множини існує необхідне значення вихідного вектора, що називається цільовим. Ці вектори утворюють навчальну пару. Ваги мережі змінюють доти, доки для кожного вхідного вектора буде отримано прийнятний рівень відхилення вихідного вектора від цільового.

Навчання нейронної мережі без вчителя є набагато правдоподібнішою моделлю навчання з точки зору біологічного коріння штучних нейронних мереж. Навчальна множина складається лише з вхідних векторів. Алгоритм

навчання нейронної мережі підлаштовує ваги мережі так, щоб виходили узгоджені вихідні вектори, тобто, щоб представлення досить близьких векторів вхідних давало однакові виходи.

Навчання передбачає два обчислювальні проходи по всіх шарах мережі: прямого та зворотного, це означає, що сигнал про помилку з виходу перцептрону знову надійде на входи мережі.

При прямому проході вхідний вектор подається до вхідних вузлів мережі, після чого поширюється від шару до шару. В результаті генерується набір вихідних сигналів, який є фактичною реакцією мережі.

Під час прямого проходу всі синаптичні ваги мережі фіксовані; під час зворотного проходу синаптичні ваги налаштовуються відповідно до правила корекції помилок, а саме: фактичний вихід мережі віднімається від бажаного відгуку, у результаті формується сигнал помилки. Цей сигнал згодом поширюється по мережі у напрямку, зворотному напрямку синаптичних зв'язків. Синаптичні ваги налаштовуються для максимального наближення вихідного сигналу мережі до бажаного. Цей метод підвищує точність результатів. Схематично процес навчання нейронної мережі представлений рисунку 2.4.

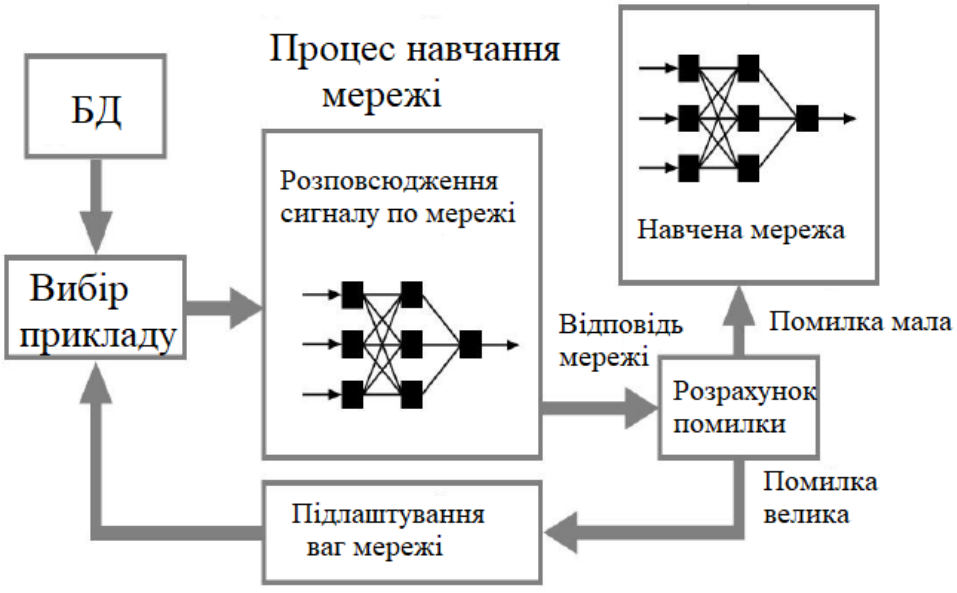


Рисунок 2.4 – Процес навчання нейронної мережі

По архітектурі зв'язків нейронні мережі поділяються на два класи: мережі прямого поширення (графи не мають петель) та рекурентні мережі (мережі зі зворотними зв'язками). До першого класу відносяться одношарові та багатошарові персептрони. До другого класу належать мережі Хопфілда, Кохонена.

Якість роботи нейронної мережі залежить від пред'явленого у процесі навчання набору навчальних даних. Навчальні дані мають бути типовими для завдання, для вирішення якого навчається мережа. Навчання мережі - процес унікальний, оскільки найчастіше прийнятні розв'язання проблеми можуть бути отримані лише у процесі численних експериментів [14].

Кожен клас мереж призначений для вирішення певних типів проблем. Доведено, що під час вирішення завдань прогнозування ефективно працюють багатошарові персептрони (MLP). Теоретично кількість шарів та число нейронів у кожному шарі може бути довільним.

Порядок дій, необхідних для вирішення задач прогнозування з використанням нейронних мереж, представлений на рисунку 2.5.

На першому етапі визначаються базові статистичні характеристики. На другому кроці проводиться формування образів, що подаються безпосередньо на входи нейронної мережі, тобто створюються навчальні та тестові множини.

Далі з використанням обраного алгоритму навчання проводиться навчання нейронної мережі та її тестування на тестовій множині.

При задовільній точності роботи нейронної мережі на тестовій множині можна вважати, що нейронна мережа побудована і може використовуватись для прогнозування [14].

Рисунок 2.5 – Алгоритм дій, необхідних для вирішення задач прогнозування з використанням нейронних мереж

2.3 Вибір архітектури та методу навчання нейронної мережі для прогнозування часових рядів

Під час проектування штучної нейронної мережі необхідно визначити такі основні параметри:

- вибір вхідних даних;
- вибір методу навчання;
- вибір архітектури нейронної мережі;
- вибір функції активації.

Процес побудови моделі штучної нейронної мережі можна умовно розділити на 5 основних етапів:

- збір даних для аналізу (формування таблиці залежних і незалежних вхідних змінних);

- опрацювання та підготовка даних (формування навчальної та тестової вибірок вхідних змінних, визначення діапазону зміни параметрів, нормування даних);

- вибір типу й архітектури нейронної мережі, придатного для розв'язання завдання;

- побудова і навчання нейромережі;

- перевірка моделі на тестовій вибірці (оцінка похибки моделі, інтерпретація результатів).

Повноцінний процес проектування нейронної мережі для конкретно заданої мети і в особливій предметній області є складнішим завданням, ніж випадковий підбір параметрів і побудова на їхній основі деякої кількості мереж, з яких вибирається найкраща.

Вибір вхідних змінних є важливим критерієм, оскільки неправильна специфікація вхідних даних призводить до погіршення продуктивності штучної нейронної мережі. Також взаємодія між вхідними змінними призводить до критичних відмінностей між вихідними значеннями.

Під час підбору вхідних змінних насамперед визначається базовий набір змінних, тобто основні фактори, що безпосередньо впливають на прогноз. З базового набору варто прибрати дані, які є "шумовими", інакше кажучи, виключити дані, за допомогою яких знижується ефективність узагальнення цієї мережі.

Дані для прогнозу не повинні бути корельованими. Корельовані змінні знижують продуктивність нейронної мережі, взаємодіючи одна з одною, вони створюють зміщення, що впливає на результат прогнозування. Виявлення "шумових" змінних полягає у знаходженні корельованих пар змінних, тобто, якщо дві змінні є сильно корельованими, то одну з них можна

прибрати з базового набору вхідних змінних для забезпечення високої точності прогнозування [15].

Велика кількість вхідних змінних також не забезпечує високої продуктивності нейронної мережі, а навпаки, сприяє її погіршенню. Скорочення кількості вхідних змінних досягається шляхом вибору однієї або декількох більш ліквідних змінних із базового набору, що також дає змогу усувати корельовані змінні [15].

Таким чином, продуктивність проєктованої нейронної мережі може бути збільшена за рахунок скорочення корельованих змінних з набору вхідних даних. Але, не варто забувати, що вхідні дані мають бути достатніми для прогнозу, тобто надмірне скорочення вхідних змінних може призвести до підвищення неточності прогнозу.

Перед введенням даних у штучну нейронну мережу необхідна їхня попередня обробка. Обробка даних полягає в нормуванні значень за певними правилами і перевірці, що дані розподілені за нормальним законом. Нормування даних дає змогу уніфікувати подання інформації всередині нейронної мережі, визначити значення вхідних змінних у певному діапазоні, що необхідно для деяких алгоритмів навчання нейронних мереж.

Після нормування даних необхідно упевнитися, що значення є рівномірними, тобто кожному інтервалу змінних належить приблизно однакова кількість значень, інакше на виході штучної нейронної мережі будуть значення, які належать до одного інтервалу змінних, що призведе до однозначного прогнозу, в якому не враховано всіх можливих варіантів.

Під вибором архітектури нейронної мережі розуміється вибір способу організації нейронів, виділення прихованих шарів, побудови зв'язків між ними. Для конкретного завдання підбирається мережа певної архітектури [14].

Основою для вибору архітектури майбутньої нейронної мережі є найменша помилка прогнозування. Із множини навчених і протестованих

нейронних мереж різної архітектури обирають мережу, яка показала найвищі результати прогнозу, тобто найменше значення помилки прогнозування [14].

Наразі не існує строго визначеної процедури вибору кількості нейронів і шарів у штучній нейронній мережі.

Кількість нейронів і шарів нейронної мережі пов'язана:

- зі складністю розв'язуваної задачі;

- з кількістю даних для навчання;

- з необхідною кількістю входів і виходів нейронної мережі;

- з наявними ресурсами: пам'яттю і швидкодією машини, на якій моделюється мережа.

Якщо в нейронній мережі занадто мало нейронів або шарів, виникають труднощі з навчанням мережі або величиною помилки, отриманої в результаті прогнозування. Перевищення необхідної кількості нейронів також заважає роботі нейронної мережі, виникають проблеми, такі як:

- низька швидкодія;

- надмірне використання пам'яті ЕОМ;

- мережа перенавчиться, тобто вихідний вектор передаватиме незначні та несуттєві деталі, наприклад, шум або помилкові дані;

- залежність виходу від входу виявиться різко нелінійною;

- здатність мережі до узагальнення буде знижена.

За архітектурою зв'язків нейронні мережі ділять на два класи (таблиця 2.1) – мережі прямого поширення та рекурентні мережі зі зворотним зв'язком.

Таблиця 2.1 – Основні архітектури нейронних мереж

Вид нейронної мережі	Модель нейронної мережі
Мережі прямого поширення	Одношаровий персептрон
	Багатошаровий персептрон
	Мережі радіальних базисних функцій

Мережі зворотного поширення	Мережа Кохонена
	Мережа Хопфілда

Одношаровий персептрон являє собою найпростішу модель штучної нейронної мережі. У такій нейронній мережі сигнали з вхідного шару відразу подаються на вихідний шар. У вихідному шарі проводяться необхідні обчислення, результати яких відразу подаються на виходи нейронної мережі [15].

Тип вхідних сигналів для одношарового персептрона - це бінарні або аналогові послідовності. Розмірності входу і виходу обмежуються тільки можливостями обчислювальної системи, на якій моделюється нейронна мережа. Ємність мережі збігається з числом нейронів [16].

Вхідний вектор *X* складається з m компонент, кожна з яких являє собою чисельну характеристику аналізованого нейронною мережею об'єкта, що аналізується нейронною мережею. Одношаровий персептрон складається з групи нейронів, що утворюють один шар, як показано на рисунку 2.6.

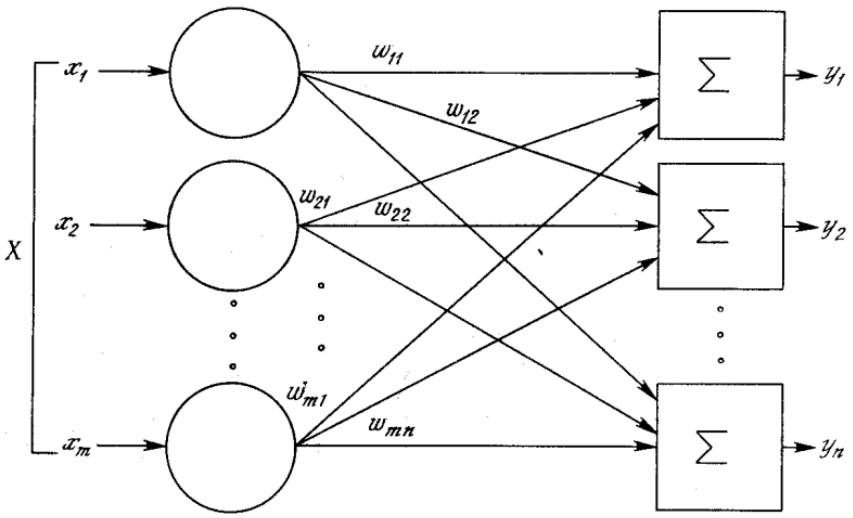


Рисунок 2.6 – Архітектура одношарового персептрона

Одношаровий персептрон має простий алгоритм навчання і здатний розв'язувати лише найпростіші завдання. В одношаровому персептроні кожен елемент вхідного вектора з'єднаний з усіма входами нейронів вихідного шару, з'єднання яких задається матрицею ваг W, що має вигляд:

$$W = \begin{pmatrix} \omega_{11} & \cdots & \omega_{1n} \\ \vdots & \ddots & \vdots \\ \omega_{n1} & \cdots & \omega_{nn} \end{pmatrix}. \quad (2.5)$$

Виходи шару нейронів формують вектор-стовпець y(t), і, таким чином, опис шару нейронів має вигляд:

$$y(t) = f(\sum_{i=1}^{N} \omega_i(t) x_i(t) - b), \quad (2.6)$$

де $w_i(t)$- вагові коефіцієнти на вході нейронів;

$x_i(t)$ - значення на вході нейронів;

b - значення зміщення.

Перевага одношарового персептрона полягає в простоті його реалізації. Основним недоліком одношарового персептрона є те, що персептрон класифікує тільки об'єкти, що лінійно розділяються. Саме тому дана архітектура не підходить для реалізації задачі прогнозування часових рядів в цілому.

Архітектура багатошарового персептрона, яка показана на рисунку 2.7, являє собою мережу, що складається з множини шарів нейронних елементів.

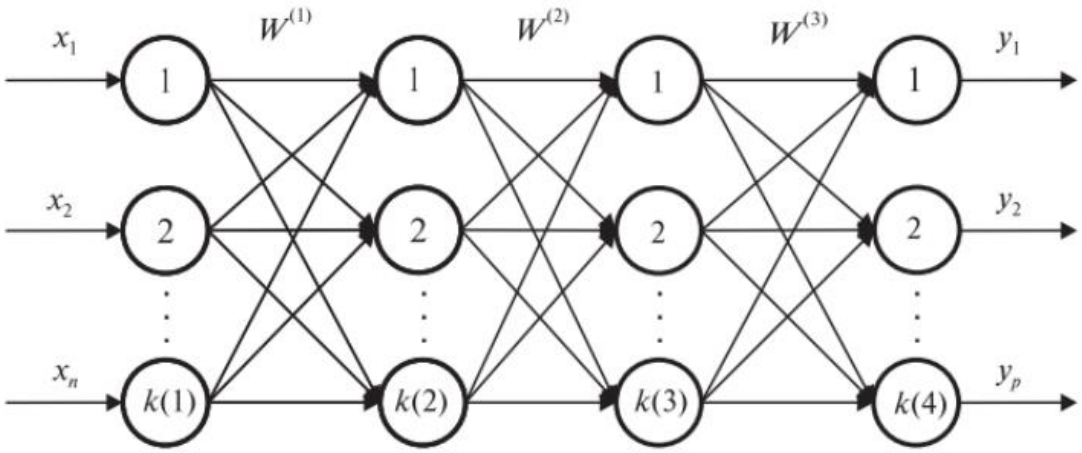


Рисунок 2.7 – Архітектура багатошарового персептрона

Вхідний шар виконує розподільчі функції. Вихідний шар необхідний для обробки інформації, отриманої від попередніх шарів, і видачі остаточних результатів. Шари нейронних елементів, розташованих між вхідним і вихідним шарами, називаються прихованими.

Аналогічно вихідному шару, основна функція прихованих шарів – це обробка інформації. Вихід кожного нейронного елемента попереднього шару нейронної мережі з'єднаний синаптичними зв'язками з усіма входами нейронних елементів наступного шару. Таким чином, топологія багатошарової нейронної мережі є однорідною та регулярною [16].

Нейронна мережа багатошарового персептрона може моделювати функцію практично будь-якої складності, причому кількість шарів і кількість нейронів у кожному шарі визначають складність функції. Визначення числа проміжних шарів і числа нейронів у них є важливим критерієм під час моделювання штучної нейронної мережі. Більшість дослідників та інженерів, застосовуючи архітектуру до певних проблем, використовують загальні правила [16]:

- кількість входів і виходів мережі визначаються кількістю вхідних і вихідних параметрів досліджуваного об'єкта, явища, процесу;

- якщо складність у відношенні між отриманими і бажаними даними на виході збільшується, кількість нейронів прихованого шару також збільшується;

- якщо модельований процес може розділятися на багато етапів, потрібен додатковий прихований шар (шари).

Кожен шар розраховує нелінійне перетворення від лінійної комбінації сигналів попереднього шару. Багатошарова нейронна мережа може формувати на виході довільну багатовимірну функцію за умови відповідного вибору кількості шарів, діапазону зміни сигналів і параметрів нейронів.

Багатошарові нейронні мережі є універсальними апроксиматорами функцій. Загальна формула, що описує багатошаровий персептрон

$$f(x) = (\sum_{i_N} \omega_{i_n j_{nN} \dots \omega_{i_2 j_{22}}} \cdot F(\omega_{i_1 j_{11}} \cdot x_{i_1 j_{11} - b_{j_1 1}}) - b_{j22} - \dots - b_{jnN}), \qquad (2.7)$$

де $w_i(t)$- вагові коефіцієнти на вході нейронів;

$x_i(t)$ - значення на вході нейронів;

b - значення зміщення.

За рахунок почергового розрахунку лінійних комбінацій і нелінійних перетворень досягається апроксимація довільної багатовимірної функції за відповідного вибору параметрів мережі.

Нейронна мережа радіальних базисних функцій містить у найбільш простій формі три шари: вхідний шар, що виконує розподіл даних зразка для першого шару ваг; вихідний шар і шар прихованих нейронів із радіально-симетричною функцією активації, кожен j-й з яких призначений для зберігання окремого еталонного вектора у вигляді вектора ваг.

Навчання шару нейронів-зразків мережі передбачає попереднє проведення кластеризації для знаходження еталонних векторів і певних евристик для визначення еталонних значень.

На рисунку 2.8 представлено структурну схему нейронної мережі з радіальними базисними функціями.

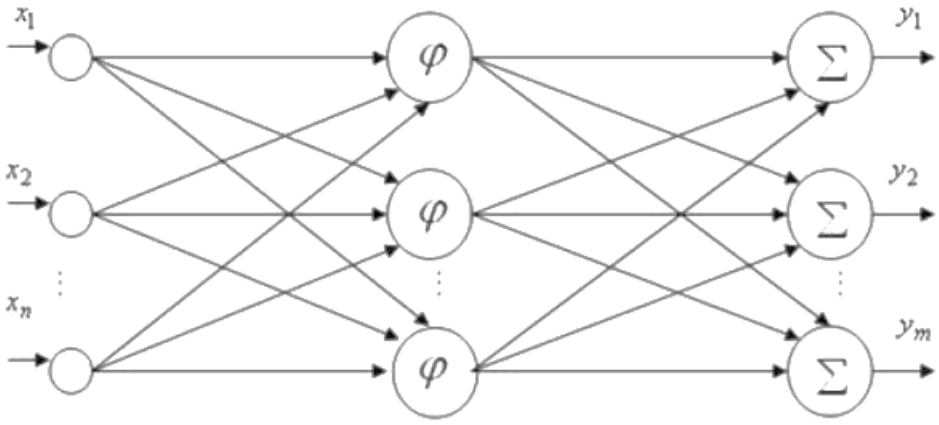


Рисунок 2.8 – Архітектура нейронної мережі з радіально-базисними функціями

Нейрони прихованого шару з'єднані за повнозв'язною схемою з нейронами вихідного шару. У моделях радіально-базисних функцій можуть бути використані різні способи вимірювання відстані між векторами, а також функції активації нейронів прихованого шару. Мережі з радіально базисними функціями мають низку переваг:

- мережі з радіально-базисними функціями здатні змоделювати довільну нелінійну функцію за допомогою всього одного проміжного шару;

- параметри лінійної комбінації у вихідному шарі можна повністю оптимізувати за допомогою добре відомих методів лінійної оптимізації;

Недоліки мереж РБФ:

- ці нейронні мережі мають погані екстраполювальні властивості і виходять вельми громіздкими при великій розмірності вектора входів;

- заздалегідь має бути відома кількість еталонів, а також евристики для побудови активаційних функцій нейронів прихованого шару.

Типова нейронна мережа зі зворотним поширенням помилки має вхідний шар, вихідний шар і як мінімум один прихований шар. Нейрони організовані в пошарову структуру з прямою передачею сигналу. Кожен

нейрон мережі продукує зважену суму своїх входів, пропускає цю величину через передавальну функцію і видає вихідне значення.

Алгоритм роботи нейронної мережі зі зворотним поширенням помилок полягає в тому, що після визначення числа шарів і нейронів, у будь-якому з них, організовується пошук значення для синаптичних ваг, здатних мінімізувати похибку прогнозованого результату. Саме для цього існують алгоритми навчання, де відбувається підгонка моделі мережі до наявних навчальних даних.

Відштовхуючись від початкової конфігурації ваг і порогів, результатом роботи алгоритму навчання є знаходження глобального мінімуму в просторі похибок. Обчислюється вектор градієнта похибок, який вказує напрямок найкоротшого спуску із заданої точки. Після закінчення навчання, алгоритм зупиняється в нижній точці, яка може виявитися як локальним, так і глобальним мінімумом. На рисунку 2.9 зображена архітектура мережі зворотного поширення помилки.

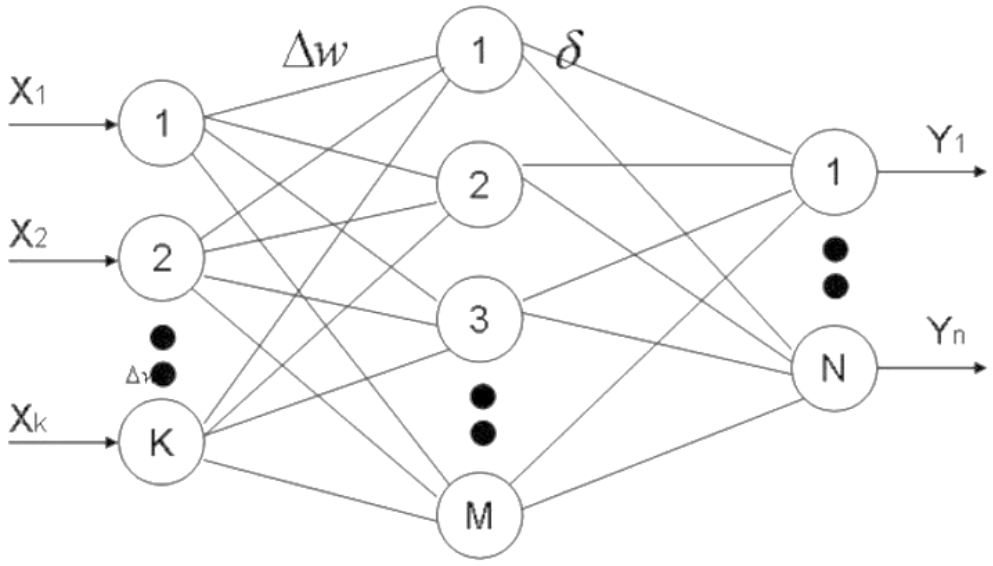


Рисунок 2.9 - Архітектура мережі зворотного поширення помилки

На кожній епосі на вхід нейронної мережі по черзі подаються всі навчальні вибірки, вихідні значення мережі порівнюються з бажаними

значеннями і обчислюється значення похибки. Значення похибки використовуються для корекції ваг, і дії повторюються. Процес навчання припиняється в тому разі, якщо пройдено певну кількість епох або значення похибки досягло певного наперед заданого рівня [17].

До переваг методу зворотного поширення належать:

- досить висока ефективність прогнозування;

- через кожен нейрон проходить інформація тільки про пов'язані з ним нейрони.

Недоліки методу зворотного поширення:

- є можливість застрягання процесу градієнтного спуску в областях численних локальних мінімумів;

- необхідно фіксувати кількість нейронів прихованого шару, що є критичним параметром складності мережі, від якого може суттєво залежати якість навчання і швидкість збіжності нейронної мережі [17].

Нейронна мережа Гопфілда являє собою динамічну систему, стани якої характеризуються фазовим простором.

Функціонування мережі можна наочно представити як рух до деякої точки рівноваги.

Нейронна мережа Гопфілда реалізує суттєву властивість автоасоціативної пам'яті - відновлення за спотвореним (зашумленим) образом найближчого до нього еталонного.

Вхідний вектор використовується як початковий стан мережі, і далі мережа еволюціонує відповідно до своєї динаміки. Причому будь-який приклад, що перебуває в області тяжіння збереженого зразка, може бути використаний як покажчик для його відновлення. Вихідний (відновлений) зразок встановлюється, коли мережа досягає рівноваги [17].

Нейронна мережа складається з єдиного шару нейронів, кількість яких одночасно дорівнює кількості входів і кількості виходів. Вихід кожного нейрона з'єднаний із входами всіх інших нейронів.

У процесі функціонування мережі Гопфілда можна виділити два режими: навчання і класифікації. У режимі навчання на основі відомих навчальних вибірок підбираються вагові коефіцієнти. У режимі класифікації при зафіксованих значеннях ваг і введенні конкретного початкового стану нейронів виникає перехідний процес, що завершується в одному з локальних мінімумів.

Завдання асоціативної пам'яті, розв'язуване за допомогою цієї мережі, зазвичай має такий вигляд – є деякий набір двійкових сигналів, які вважаються еталонними. Мережа повинна вміти з довільного вхідного сигналу виділити той еталонний зразок, який найбільш схожий на поданий сигнал або ж видати повідомлення про те, що поданий сигнал ні з чим не асоціюється. Вхідний сигнал, у загальному випадку, може бути описаний вхідним вектором. Кожен елемент вектора дорівнює або +1, або -1.

Архітектура нейронної мережі Гопфілда представлена на рисунку 2.10.

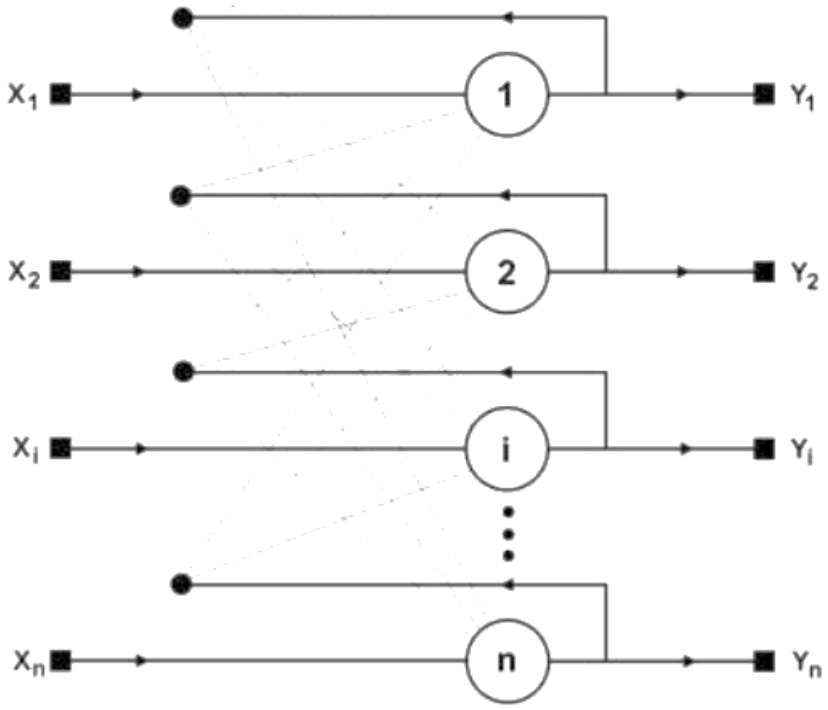


Рисунок 2.10 – Архітектура нейронної мережі Гопфілда

Різні види нейронних мереж мають свої переваги та недоліки. Наприклад, до переваг мереж прямого поширення належить простота їх реалізації, а також гарантована математична збіжність обчислень. При цьому до недоліків відносять швидке зростання числа нейронів зі збільшенням складності завдання [17].

Виходячи з завдання на виконання магістерської дисертації, остаточно прийнято рішення застосовувати багатошаровий персептрон, тому що він не має недоліків, які роблять неможливим застосовувати даний тип архітектури для прогнозування часових рядів.

Найпоширенішими методами навчання ШНМ є [17]:

- градієнтні алгоритми навчання;
- алгоритми методу спряжених градієнтів;
- квазіньютонові алгоритми навчання.

До градієнтних алгоритмів навчання належать:

- алгоритм градієнтного спуску зі збуренням;
- алгоритм градієнтного спуску з вибором параметра швидкості налаштування;
- пороговий алгоритм зворотного поширення помилки.

Алгоритм градієнтного спуску зі збуренням призначений для налаштування і навчання штучних нейронних мереж.

Цей алгоритм дає змогу долати локальні нерівності поверхні помилки і не зупинятися в локальних мінімумах.

У процесі роботи алгоритму підсумовуються градієнти, розраховані на кожному кроці навчання, а параметри модифікуються після того, як усі навчальні дані будуть представлені.

Алгоритм градієнтного спуску з вибором параметра швидкості налаштування використовує евристичну стратегію зміни цього параметра протягом усього процесу навчання.

У процесі навчання обчислюються вихід і похибка ініціалізованої нейронної мережі. Потім на кожному циклі навчання обчислюються нові значення параметрів, що налаштовуються, і нові значення виходів і похибок.

Ця стратегія сприяє збільшенню швидкості та скороченню тривалості навчання нейронної мережі загалом.

Пороговий алгоритм зворотного поширення помилки використовує евристичну стратегію зміни кроку приросту параметрів.

Метою цього алгоритму є підвищення чутливості методу при великих значень входу функції активації. Значення приросту для кожного параметра, що налаштовується, збільшується щоразу, коли похідна функція помилки за цим параметром зберігає знак для двох послідовних ітерацій.

На рисунках 2.11 – 2.13 зображено приклади навчання штучної нейронної мережі з використанням градієнтних методів навчання при кількості нейронів у прихованому шарі, що дорівнює 10.

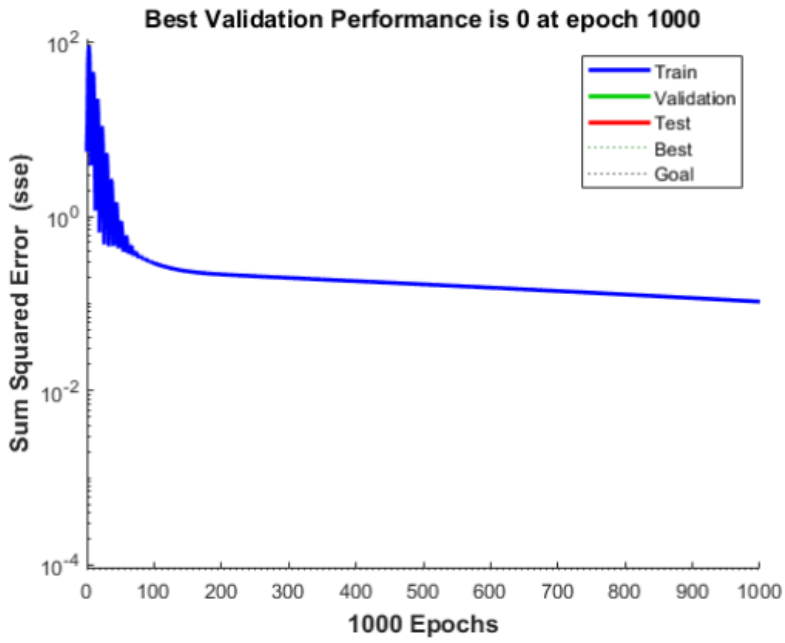


Рисунок 2.11 - Демонстрація навчання нейронної мережі з використанням алгоритму градієнтного спуску зі збуренням

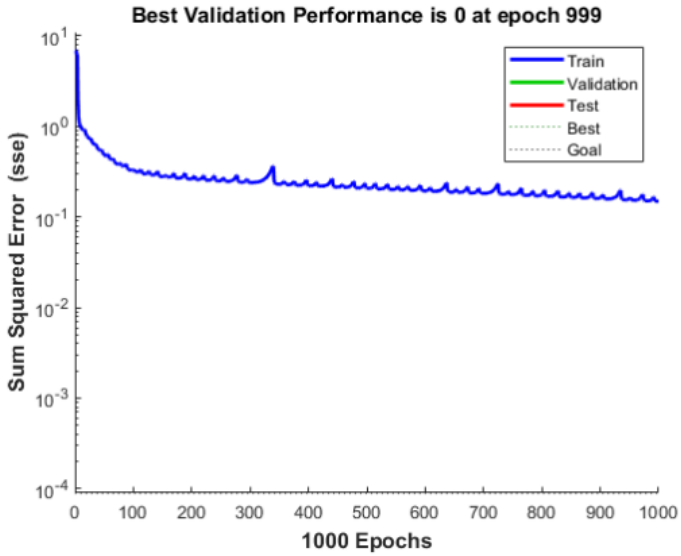


Рисунок 2.12 - Демонстрація навчання нейронної мережі з використанням алгоритму градієнтного спуску з вибором параметра швидкості налаштування

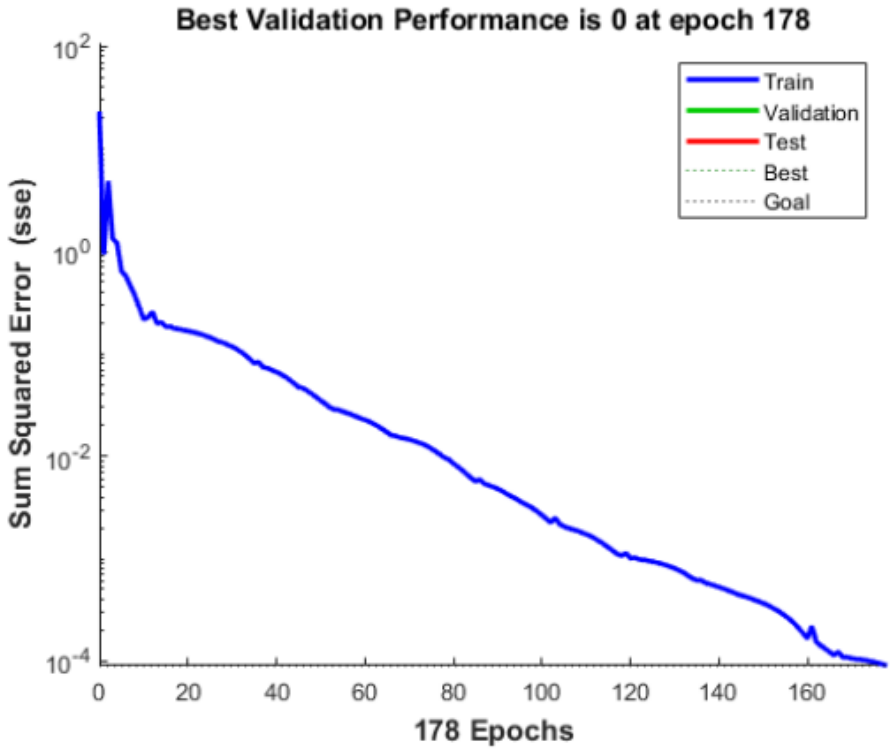


Рисунок 2.13 - Демонстрація навчання нейронної мережі з використанням порогового алгоритму зворотного поширення помилки

До алгоритмів методу спряжених градієнтів належать:

- метод Флетчера-Рівса;

- алгоритм Полока-Рібейри;

- алгоритм Моллера.

Під час навчання з використанням алгоритмів методу спряжених градієнтів градієнтів розмір кроку коригується на кожній ітерації, порівняно з градієнтними методами навчання. Для визначення розміру кроку вздовж спряженого напрямку виконуються процедури знаходження мінімуму.

Усі алгоритми методу спряжених градієнтів на першій ітерації здійснюють пошук глобального мінімуму в напрямку антиградієнта.

Після вибору напрямку визначається оптимальна відстань, на величину якої необхідно змінити параметри нейронної мережі. В результаті визначається наступний напрямок пошуку найшвидшого спуску і вектора руху в спряженому напрямку.

У випадку методу Флетчера-Рівса константа β дорівнює відношенню квадрата норми градієнта на поточній до квадрата норми градієнта на попередній ітерації.

Для алгоритму Полока-Рібейри константа β дорівнює скалярному добутку приросту градієнта на поточний градієнт, поділеному на квадрат норми градієнта на попередній ітерації.

Алгоритм Моллера поєднує в собі ідеї методу спряжених градієнтів із квазіньютоновими методами, даючи змогу уникнути витрат часу ресурсів на обчислення реакції мережі на кожній ітерації навчання. Цей метод вимагає більшої кількості ітерацій, ніж інші алгоритми спряжених градієнтів, але при цьому загальна кількість обчислень на кожному кроці істотно скорочується [17].

На рисунках 2.14 – 2.16 зображено приклади навчання штучної нейронної мережі з використанням методів навчання спряжених градієнтів за кількості нейронів у прихованому шарі, що дорівнює 10.

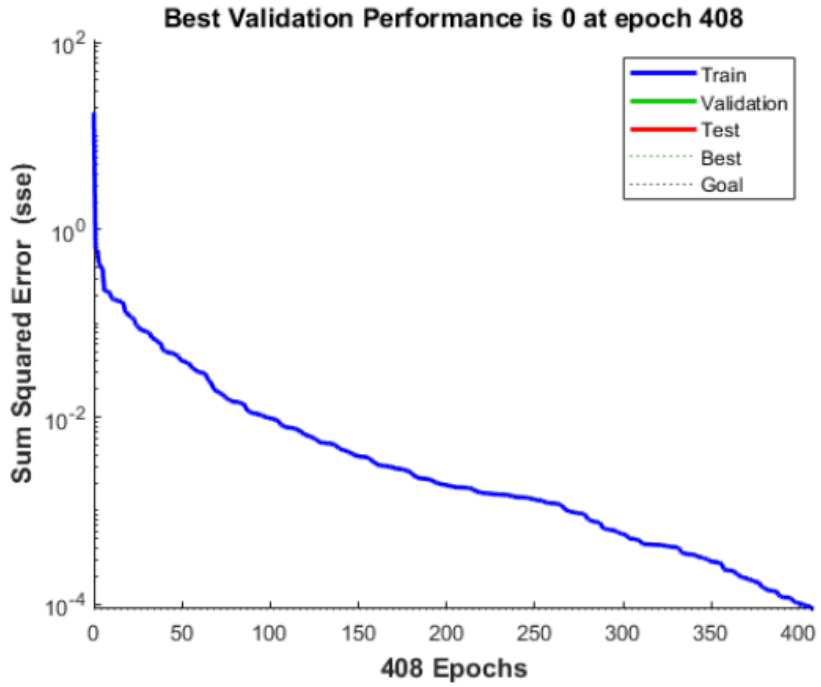


Рисунок 2.14 - Демонстрація навчання нейронної мережі з використанням алгоритму Флетчера-Рівса

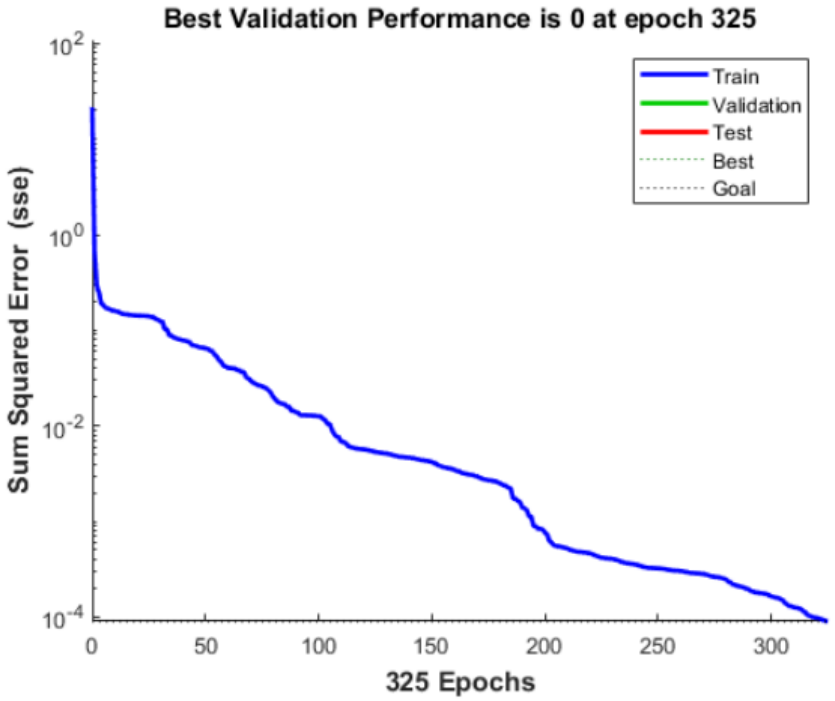


Рисунок 2.15 - Демонстрація навчання нейронної мережі з використанням алгоритму Полока-Рібейри

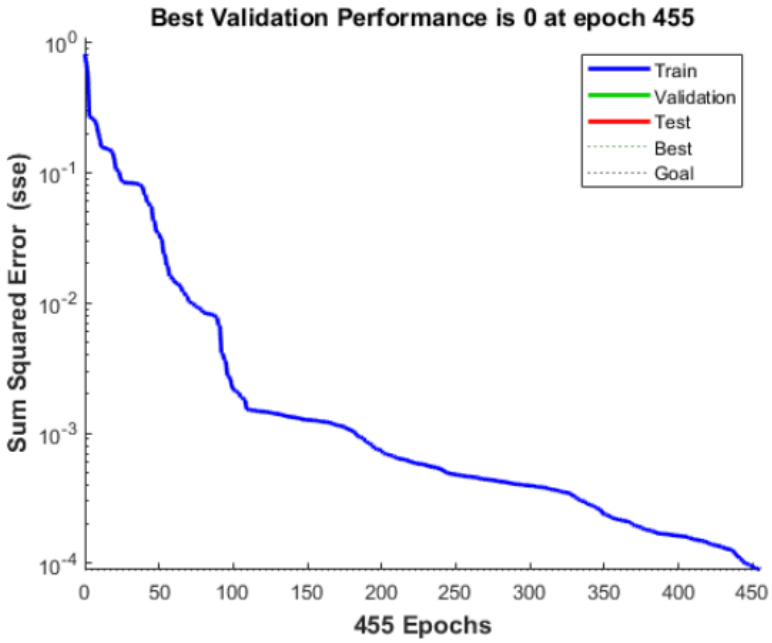


Рисунок 2.16 - Демонстрація навчання нейронної мережі з використанням алгоритму Моллера

До квазіньютонових алгоритмів навчання належать:

- алгоритм BFGS (Бройден, Флетчер, Гольдфарб, Шанно);
- однокроковий алгоритм методу січних площин;
- метод Левенберга-Маркварда.

Алгоритм BFGS є альтернативою методу сполучених градієнтів для більш прискореного навчання штучних нейронних мереж.

Основний крок алгоритму BFGS визначається наступним співвідношенням:

$$x(k+1) = x(k) - H(k)^{-1} g(k), \quad (2.8)$$

де *x(k)* - вектор параметрів, що налаштовуються;

H(k) - матриця других приватних похідних функції помилки;

g(k) - вектор градієнта функції помилки.

Цей алгоритм потребує більшої кількості обчислень на кожній ітерації, ніж алгоритми методу спряжених градієнтів. На кожній ітерації необхідно

зберігати матрицю других частинних похідних функції помилки, внаслідок чого, цей алгоритм неефективний для нейронних мереж із великою кількістю входів.

Однокроковий алгоритм методу січних площин ідентичний методу BFGS, але в цьому алгоритмі матриця других приватних похідних функції помилки не вираховується на кожній ітерації навчання, а визначається як одинична матриця. Цей алгоритм потребує менших обсягів пам'яті та обчислювальних ресурсів, порівняно з методом BFGS.

Алгоритм Левенберга-Маркварда є нелінійним методом найменших квадратів. Алгоритм розроблено для оптимізації нелінійних моделей, таким чином, щоб мінімізувати середньоквадратичну функцію помилок. Алгоритм навчання Левенберга-Маркварда полягає в послідовному наближенні заданих початкових даних до шуканого локального мінімуму. Метод передбачає знаходження компромісу між лінійною моделлю і градієнтним спуском. Таким чином, крок робиться тільки в тому разі, якщо він зменшує помилку. Для забезпечення просування використовується градієнтний спуск із досить малим кроком.

На рисунках 2.17 – 2.19 зображено приклади навчання штучної нейронної мережі з використанням квазіньютонових методів навчання при кількості нейронів у прихованому шарі, що дорівнює 10.

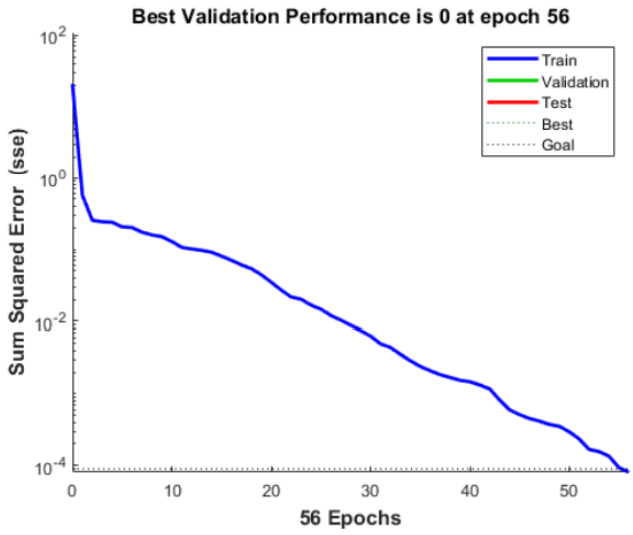


Рисунок 2.17 - Демонстрація навчання нейронної мережі з використанням методу BFGS

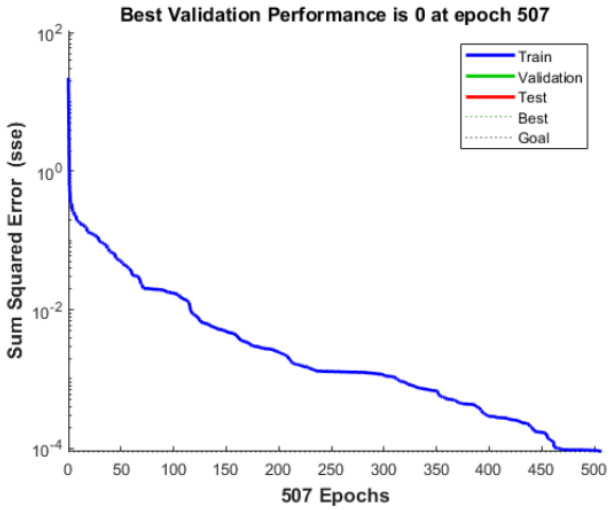


Рисунок 2.18 – Навчання мережі за алгоритмом січних площин

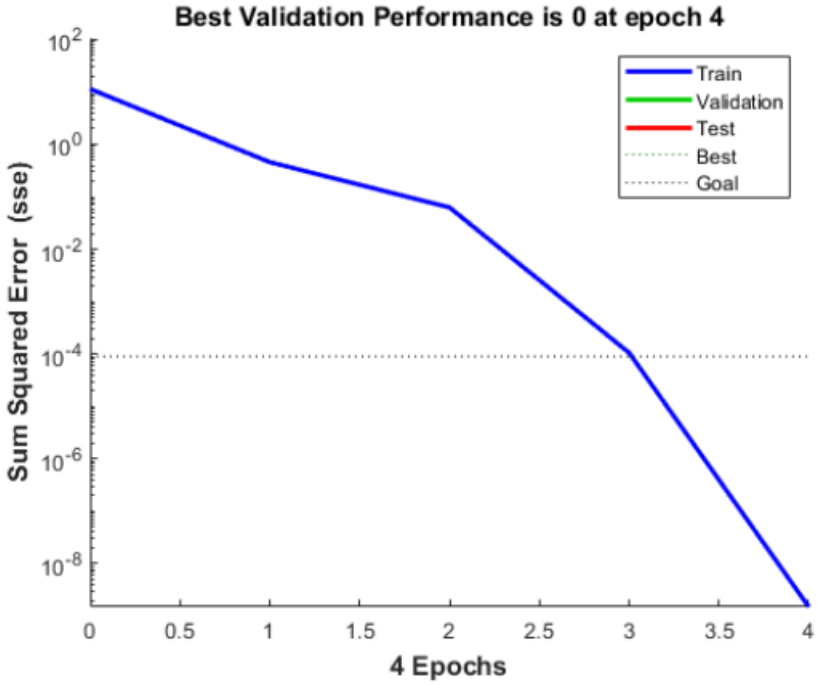


Рисунок 2.19 - Демонстрація навчання нейронної мережі з використанням алгоритму Левенберга-Маркварда

Для проведення порівняльного аналізу розглянутих алгоритмів навчання штучних нейронних мереж використовували три критерії [17]: середня абсолютна похибка навчання, кількість епох навчання та кількість нейронів у прихованому шарі.

За результатами цього дослідження найменша похибка навчання була виявлена при навчанні нейронної мережі методом Левенберга-Маркварда. Для подальшого прогнозування даних буде використовуватися саме цей метод.

3 РОЗРОБКА НЕЙРОННОЇ МЕРЕЖІ ПРЯМОГО ПОШИРЕННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

3.1 Математична модель нейронної мережі

В межах даного розділу наведемо приклад прогнозування часових рядів на основі нейронної мережі прямого поширення.

Для прогнозування інформаційних даних з використанням штучних нейронних мереж у розділі 2 роботи було визначено архітектуру та основні параметри нейронної мережі.

В якості вихідних даних для прогнозування було використано кількість отриманих повідомлень, що надійшли на обслуговуючий пристрій.

Нормування вхідних даних являє собою перехід до однакового опису для всіх характеристик, тобто, до введення нової умовної одиниці, що допускають формальні зіставлення значень. Інакше кажучи, нормування даних необхідне для подання вхідних даних у певному інтервалі значень [18].

Багато нейронних мереж працюють із даними, представленими на інтервалі [0,1]. Але, варто врахувати, що якщо дані зосереджені в невеликій області одиничного інтервалу, то інформативність таких даних мала. Якщо ж дані розподілені на одиничному інтервалі за нормальним законом, то інформативний зміст даних буде великим [18].

Однією з найпростіших функцій для нормування даних є лінійне перетворення (3.1-3.5). Цей алгоритм передбачає, що дані змінюються в деякому діапазоні [min, max] і забезпечується нормування змінної на весь діапазон розкиду її значень. Функції мають вигляд:

$$Y^H = \frac{Y_i - Y_{min}}{Y_{max} - Y_{min}}; \quad (3.1)$$

$$Y^H = \frac{Y_i - \bar{Y}}{Y_{max} - Y_{min}}; \quad (3.2)$$

$$Y^H = \frac{Y_i}{\bar{Y}}; \quad (3.3)$$

$$Y^H = \frac{Y_i}{Y_{max}}; \quad (3.4)$$

$$Y^H = \frac{Y_i}{Y_{min}}, \quad (3.5)$$

де Y_i – середнє значення вибірки;

Y_{max} – максимальне значення вибірки;

Y_{min} – мінімальне значення вибірки.

Математична модель нейронної мережі:

$$f(x) = (\sum_{i_N} \omega_{i_n j_{nN} \ldots \omega_{i_2 j_{22}}} \cdot F(\omega_{i_1 j_{11}} \cdot x_{i_1 j_{11} - b_{j_1 1})} - b_{j22} - \cdots - b_{jnN}), \quad (3.6)$$

де $w_i(t)$- вагові коефіцієнти на вході нейронів;

$x_i(t)$ - значення на вході нейронів;

b - значення зміщення.

В якості функції активації прихованого шару попередньо будемо використовувати логістичну сигмоїдальну функцію, оскільки ця функція диференційована на всій осі абсцис і має властивість посилювати слабкі сигнали краще, ніж сильні, тим самим запобігаючи насиченню від сильних сигналів. Математично цю функцію можна виразити так:

$$\sigma(x) = \frac{1}{1+\exp(-tx)}, \quad (3.7)$$

де t — це параметр функції, що визначає її крутизну.

Зовнішній вигляд логістичної функції наведений на рисунку 3.1.

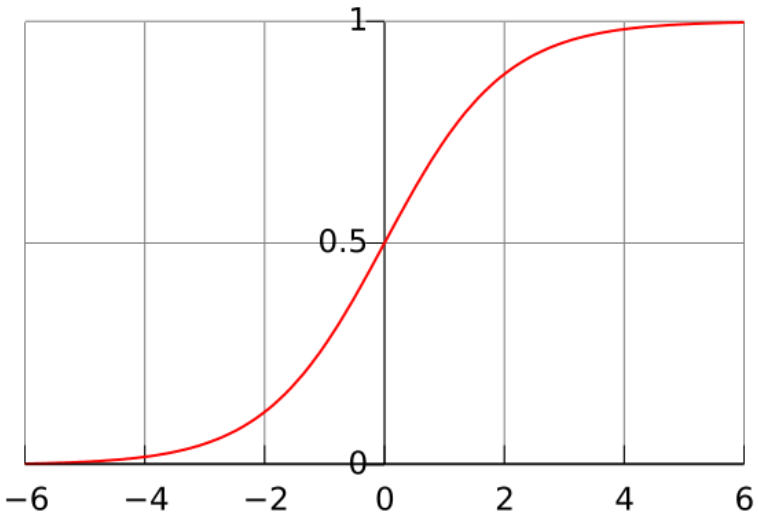


Рисунок 3.1 – Зовнішній вигляд логістичної функції

В якості функції вихідного шару активації попередньо використовується лінійна функція, у зв'язку з тим, що лінійна ділянка функції активації дає змогу виконувати дії з безперервними сигналами в діапазоні (-∞, ∞).

Зовнішній вигляд лінійної функції наведений на рисунку 3.2.

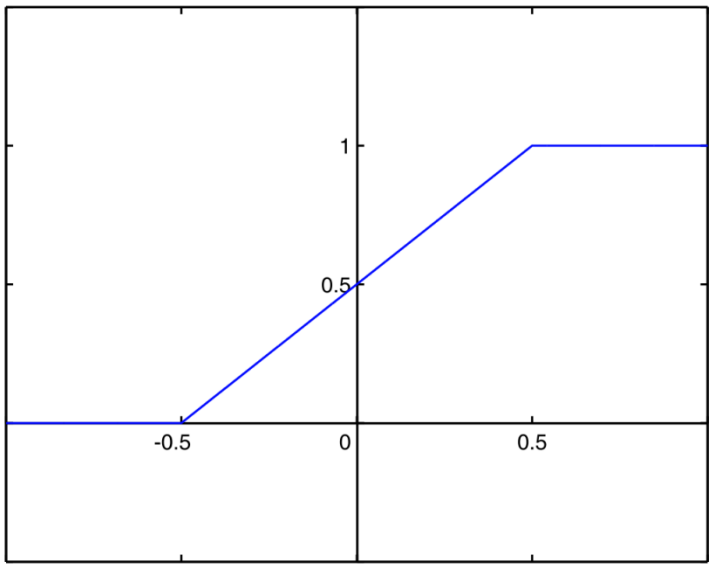


Рисунок 3.2 – Зовнішній вигляд лінійної функції

3.2 Структура нейронної мережі прямого поширення для прогнозування часових рядів

Перед тим, як навести структуру нейронної мережі прямого поширення, необхідно провести нормування вихідних даних.

На рисунках 3.1 – 3.3 наведено графіки розподілу даних після нормування з використанням функцій (3.1) – (3.3).

При $Y_{min} = 0$, вирази 3.1 і 3.4 тотожно дорівнюють, а вираз 3.5 не може бути використаний.

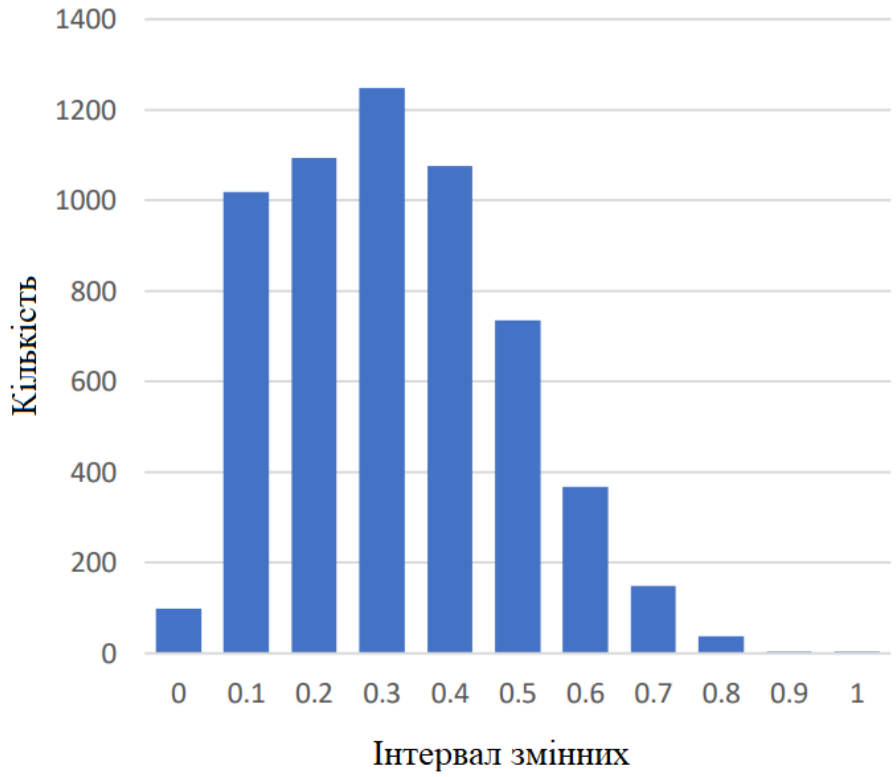


Рисунок 3.1 - Нормування кількості отриманих повідомлень за функцією (3.1)

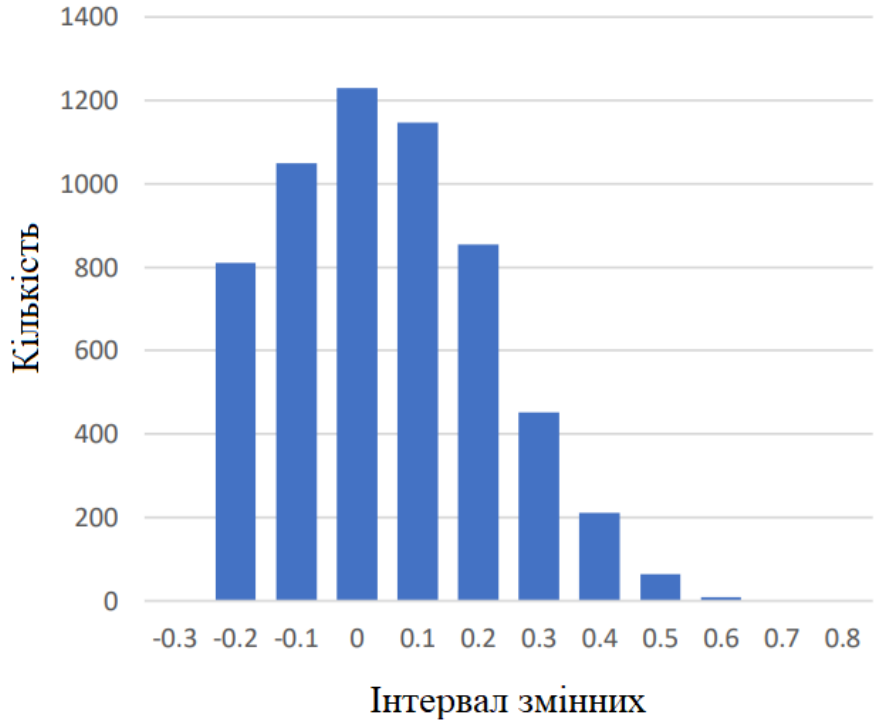


Рисунок 3.2 - Нормування кількості отриманих повідомлень за функцією (3.2)

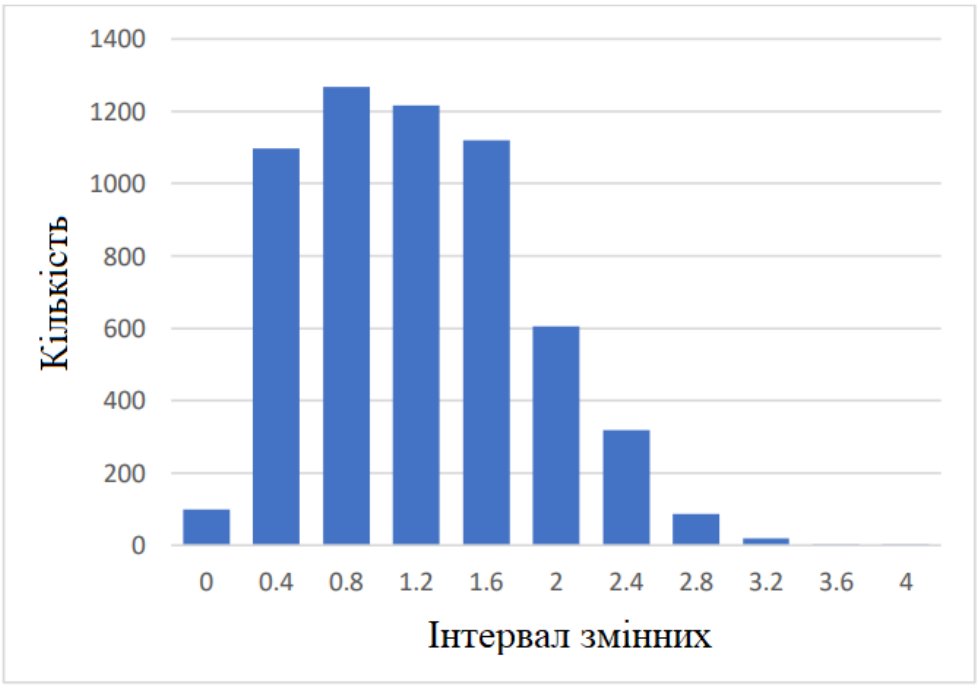


Рисунок 3.3 - Нормування кількості отриманих повідомлень за функцією (3.3)

Методи лінійного нормування не завжди застосовні, оскільки під час лінійного нормування в даних можуть виникати відносно рідкісні викиди, що перевищують типовий розкид вибірки, саме ці викиди визначають масштаб нормування. Таким чином, основна маса значень нормованої змінної зосередиться поблизу одного значення.

Також популярний у використанні спосіб лінійного нормування, званий центруванням. Цей метод надійніший, порівняно з іншими методами лінійного нормування, оскільки орієнтований на типові значення, такі як математичне очікування, середнє значення і дисперсія:

$$Y^H = \frac{(Y_i - M(Y_i))}{\sigma(Y_i)}, \tag{3.8}$$

де $M(Y_i)$ - математичне сподівання вибірки Y;

$\sigma(Y_i)$ - оцінка середнього квадратичного відхилення.

У разі нормування даних з використанням функції 3.8, нормовані величини не належать гарантовано одиничному інтервалу, тобто, максимальний розкид значень заздалегідь не відомий. Після нормування даних кожне значення являє собою відхилення від середнього значення. На рисунку 3.4 показано розподіл нормованих даних при використанні виразу 3.8, дані не належать до одиничного інтервалу.

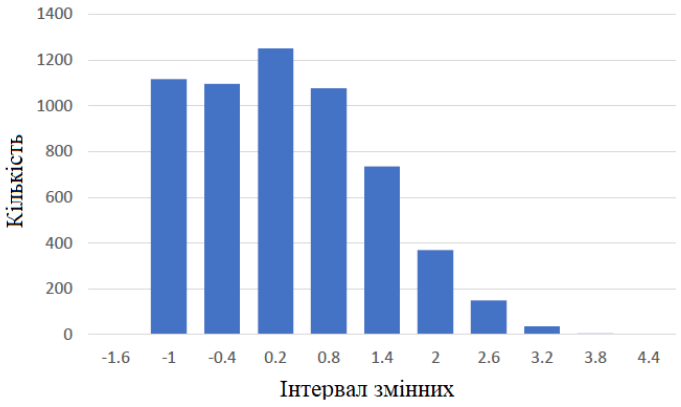


Рисунок 3.4 – Нормування кількості отриманих повідомлень

Найчастіше лінійне перетворення значень не забезпечує нормування основної маси даних і обмеження діапазону можливих значень цих даних. Виходом із цієї ситуації може бути використання для попереднього оброблення даних функції активації:

$$Y^H = f(u), \tag{3.9}$$

де $f(u)$ – функція активації нейронів.

На рисунку 3.5 наведено графіки розподілу вхідних даних після нормування з використанням функції 3.9.

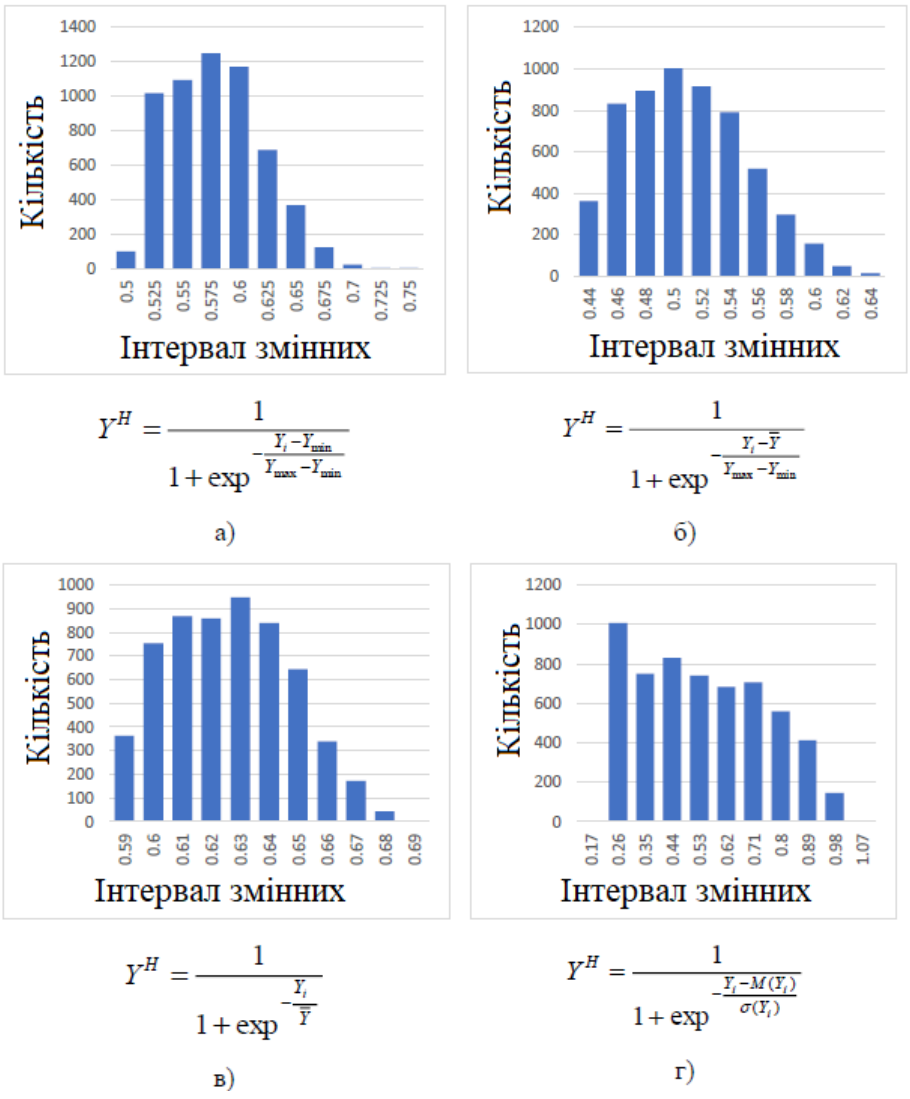


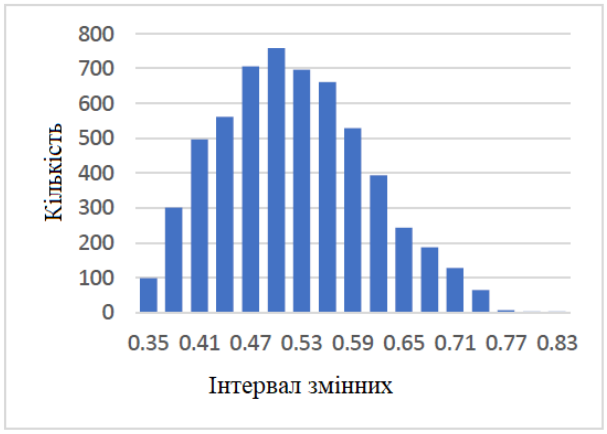
Рисунок 3.5 – Нормування кількості отриманих повідомлень з застосування виразу (3.9)

Також можливе застосування нелінійного перетворення:

$$Y^H = \frac{a}{1+\exp\left(-\frac{bY_i}{Y_{max}}\right)} - c\ , \qquad (3.10)$$

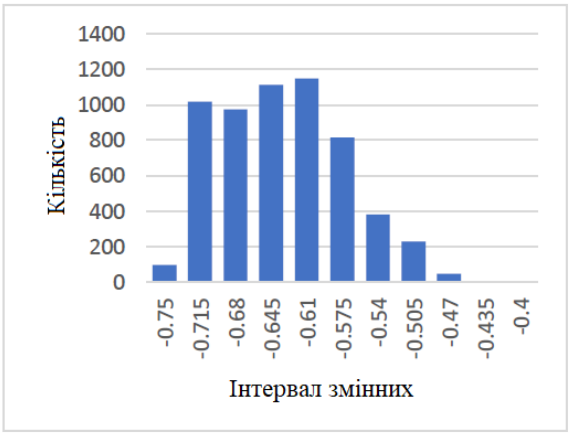
де a, b, c - константи, що визначаються емпірично.

На рисунку 3.6 наведено графіки розподілу вхідних даних після нормування з використанням функції 3.6(а) – (a= 1.5, b= 1.5, c= 0.4) та 3.6(б) – (a=1.5, b=1, c=1.5).



$$Y^H = \frac{1.5}{1+\exp\left(-1.5\times\frac{Y_i}{Y_{max}}\right)} - 0.4$$

а)



$$Y^H = \frac{1.5}{1+\exp\left(-\frac{Y_i}{Y_{max}}\right)} - 1.5$$

б)

Рисунок 3.6 – Нормування кількості отриманих повідомлень з застосування виразу (3.10)

Належність розподілів даних до нормального закону визначається з використанням критерію узгодженості Пірсона (χ^2). Критерій узгодженості Пірсона дає можливість здійснювати перевірку емпіричного та теоретичного розподілів однієї ознаки.

Критерій узгодженості Пірсона застосовується в таких випадках:

- для зіставлення теоретичного розподілу з емпіричним (нормальним, рівномірним тощо);

- для зіставлення двох емпіричних розподілів однієї ознаки.

Ідея цього методу полягає у визначенні ступеня розбіжності відповідних частот (отриманої практичним методом і розрахованої теоретичним методом). Чим більшою є ця розбіжність, тим більшим буде значення критерію узгодженості Пірсона.

В якості критерію перевірки гіпотези приймають величину:

$$\chi^2 = \sum \frac{(n_i - n_i')^2}{n_i}, \tag{3.8}$$

де n_i - це емпіричні частоти, отримані з вибірки;

n_i' - теоретичні частоти, отримані теоретичним шляхом.

Доведено, що за нескінченно великого значення обсягу вибірки закон розподілу випадкової величини прагне до закону χ^2з k ступенями свободи.

Число ступенів свободи визначають за формулою:

$$k = S - 1 - r, \tag{3.9}$$

де S - це число груп інтервалів;

r - число параметрів.

Для рівномірного і нормального розподілів значення r приймається рівним 2 (математичне очікування і середньоквадратичне відхилення), для показникового розподілу r =1.

Перевірка гіпотези про належність розподілів до нормального закону визначається за допомогою виразу 3.10. При виконанні нерівності 3.10 немає підстав спростовувати гіпотезу про те, що дані розподілені за нормальним законом:

$$\chi^2_{\text{спост}} < \chi^2_{\text{крит}}. \quad (3.10)$$

У таблиці 3.1 подано значення $\chi^2_{\text{спост}}$ і $\chi^2_{\text{крит}}$ для різних функції нормування вихідних даних.

Таблиця 3.1 – Значення $\chi^2_{\text{спост}}$ і $\chi^2_{\text{крит}}$ для різних функції нормування

Вид нормування	$\chi^2_{\text{спост}}$	$\chi^2_{\text{крит}}$
Лінійне перетворення (3.1)	17,79	33,92
Лінійне перетворення (3.2)	32,73	16,92
Лінійне перетворення (3.3)	23,72	12,59
Центрування	30,48	19,67
Нелінійне перетворення (рис.3.5, а)	30,78	18,31
Нелінійне перетворення (рис.3.5, б)	20,65	18,3
Нелінійне перетворення (рис.3.5, в)	17,62	18,31
Нелінійне перетворення (рис.3.5, г)	33,15	24,99
Нелінійне перетворення (рис.3.6, а)	8,34	23,69
Нелінійне перетворення (рис.3.6, б)	22,48	23,69

Як видно з наведеного вище, у разі лінійного та нелінійних перетворень розподіли даних належать до нормального закону, тому що виконується нерівність 3.10. Отже, нормовані дані, отримані з використанням виразів, наведених вище, можна використовувати для прогнозування.

Для прогнозування обирають дані, перетворені з використанням виразу 3.1, оскільки різниця значень критичного і спостережуваного критеріїв Пірсона максимальна.

Вибір початкових вагових коефіцієнтів ґрунтується на мінімізації кількості епох навчання нейронної мережі та нейронів прихованого шару.

Значення початкових вагових коефіцієнтів варіюються в проміжку (-1;1). Для наочності на рисунку 3.7 наведено значення кількості епох навчання (мінімальна кількість – за початковими ваговими коефіцієнтами 0.1 і 0.2).

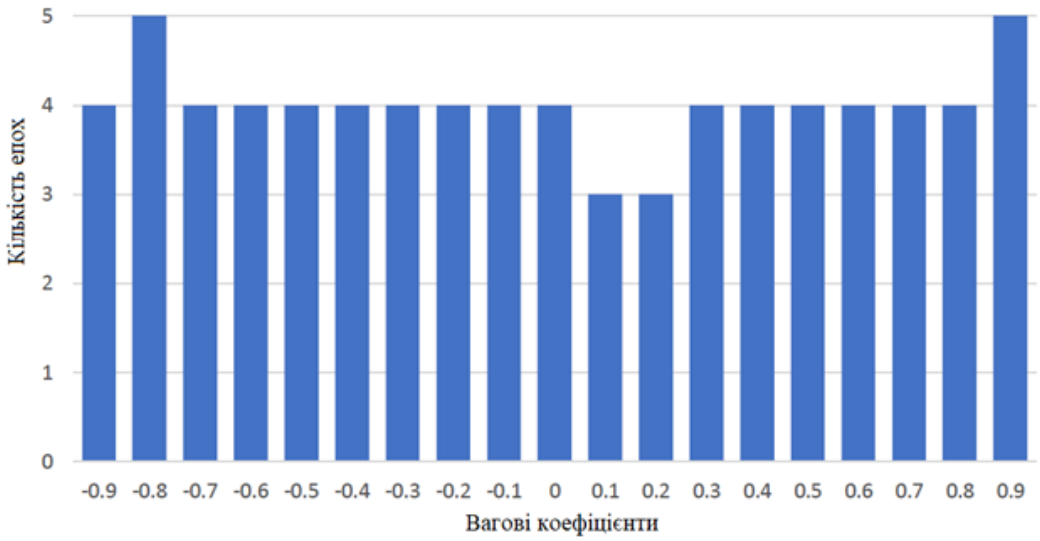


Рисунок 3.7 – Мінімальні значення кількості епох навчання при різних значеннях вагових коефіцієнтів

На рисунку 3.8 наведено залежності кількості епох навчання нейронної мережі за значеннями початкових вагових коефіцієнтів 0.1 і 0.2.

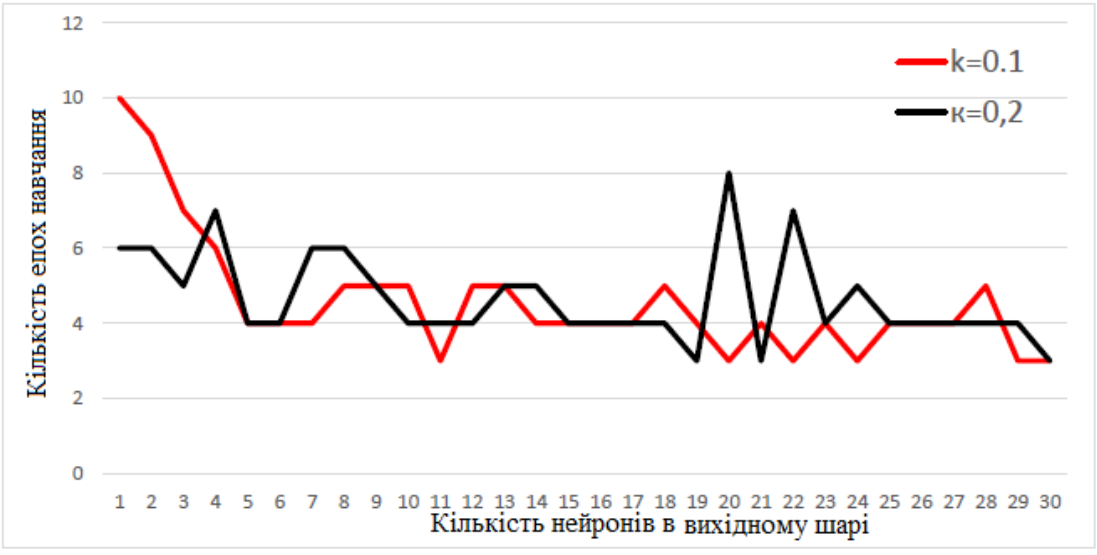


Рисунок 3.8 – Залежність кількості епох навчання при значенні вагових коефіцієнтів 0,1 і 0,2

З рисунка 3.8 видно, що за $k=0.1$ спостерігається шість мінімумів, за $k=0.2$ кількість мінімумів дорівнює трьом.

При значенні $k=0.1$ мінімальні значення кількості епох навчання були досягнуті при кількості нейронів у вихідному шарі: 11, 20, 22, 24, 29, 30, при значенні $k=0{,}2$: 19, 21, 30. У цій ситуації вибір значення початкових вагових коефіцієнтів ґрунтується на виборі кількості нейронів у вихідному шарі, за якого значення абсолютної похибки навчання мінімальне.

Значення похибки навчання мережі при мінімальній кількості епох навчання представлене у вигляді таблиці 3.2.

Таблиця 3.2 - Значення похибки навчання мережі при мінімальній кількості епох навчання

Значення вагових коефіцієнтів	Кількість шарів у вихідному шарі	Значення похибки навчання, %
k =0,1	11	0,0018
	20	0,0027
	22	0,0204
	24	0,1874
	29	0,0803
	30	0,1065
k =0,2	19	0,1626
	21	0,2199
	30	0,2937

Виходячи з результатів, представлених у таблиці 3.2, найменше значення похибки навчання спостерігається за значення k =0.1 і кількості нейронів у вихідному шарі, що дорівнює 11.

Прихований шар нейронної мережі складається з 24 нейронів (за кількістю входів (за значеннями часового ряду, години).

Структура нейронної мережі прямого поширення представлена на рисунку 3.3.

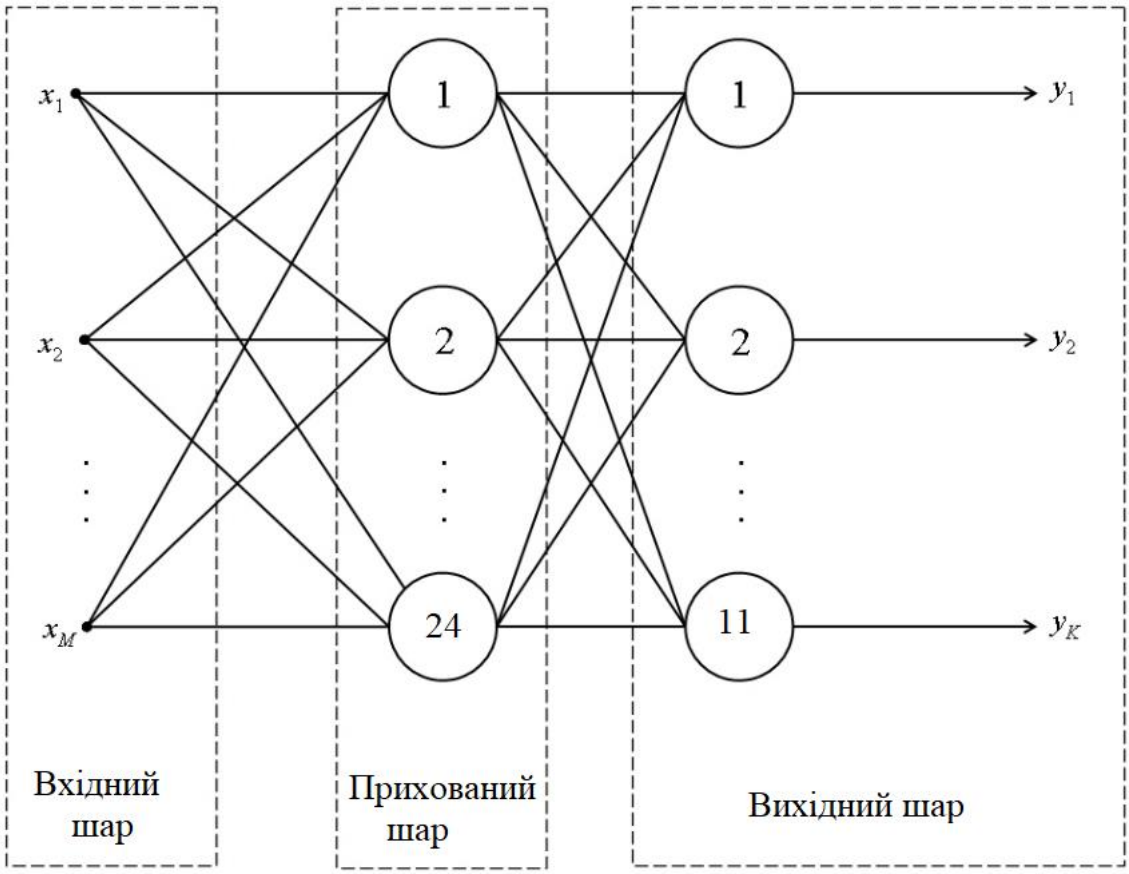


Рисунок 3.9 – Структура нейронної мережі прямого поширення

Результати реалізації моделі наведено далі.

3.3 Приклад реалізації моделі

Для реалізації нейронної мережі з метою прогнозування використовувалася універсальна інтегрована система STATISTICA Neural Networks.

Під час запуску інструмента NeuralNetworks у програмі Statistica. пропонується вибрати вид розв'язуваної задачі. Вибираємо пункт "Тимчасові ряди (регресія)", як показано на рисунку 3.10.

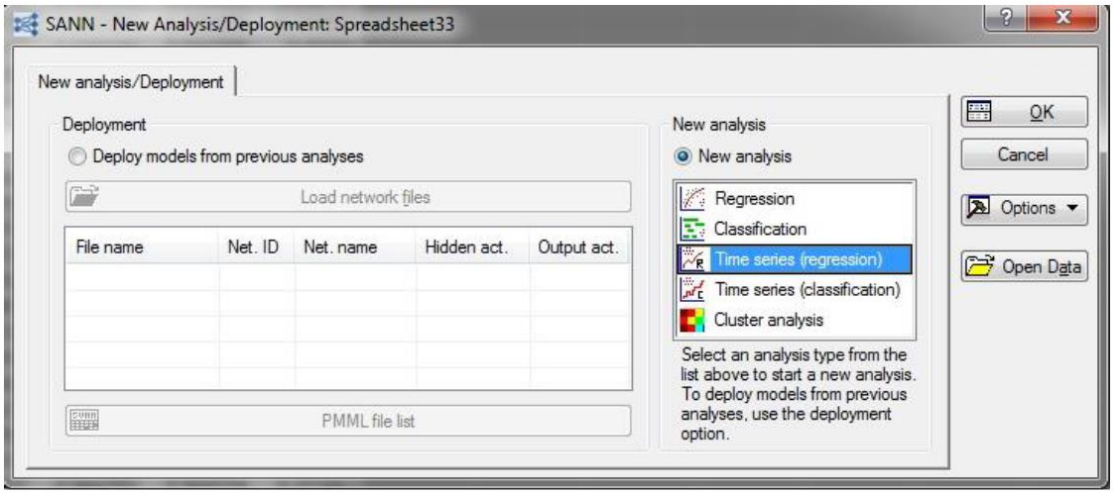


Рисунок 3.10 – Вибір типу завдання

Розподіляємо дані на входи і виходи та обираємо тип нейронної мережі. Вибір функції активації, способу навчання та інших параметрів модельованої нейронної мережі виконується у вікні, яке представлене на рисунку 3.11.

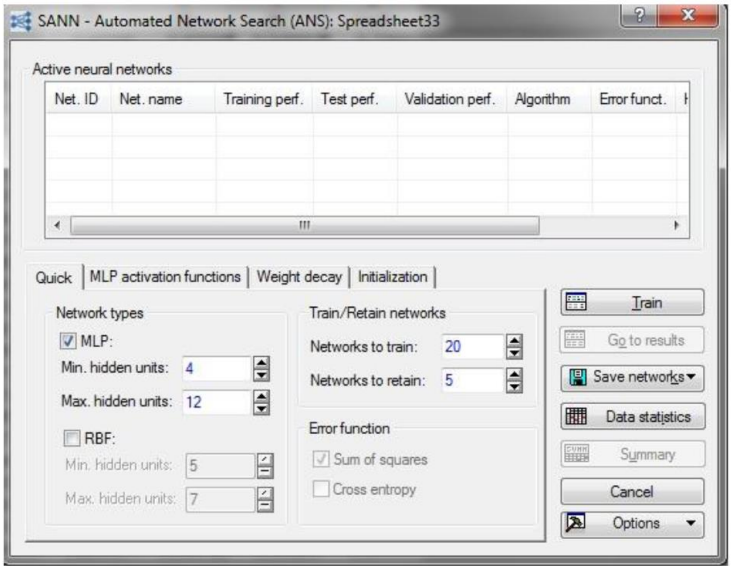


Рисунок 3.11 - Вибір параметрів модельованої нейронної мережі

Результати прогнозування даних з застосуванням нейронної мережі прямого поширення представлені на рисунку 3.12.

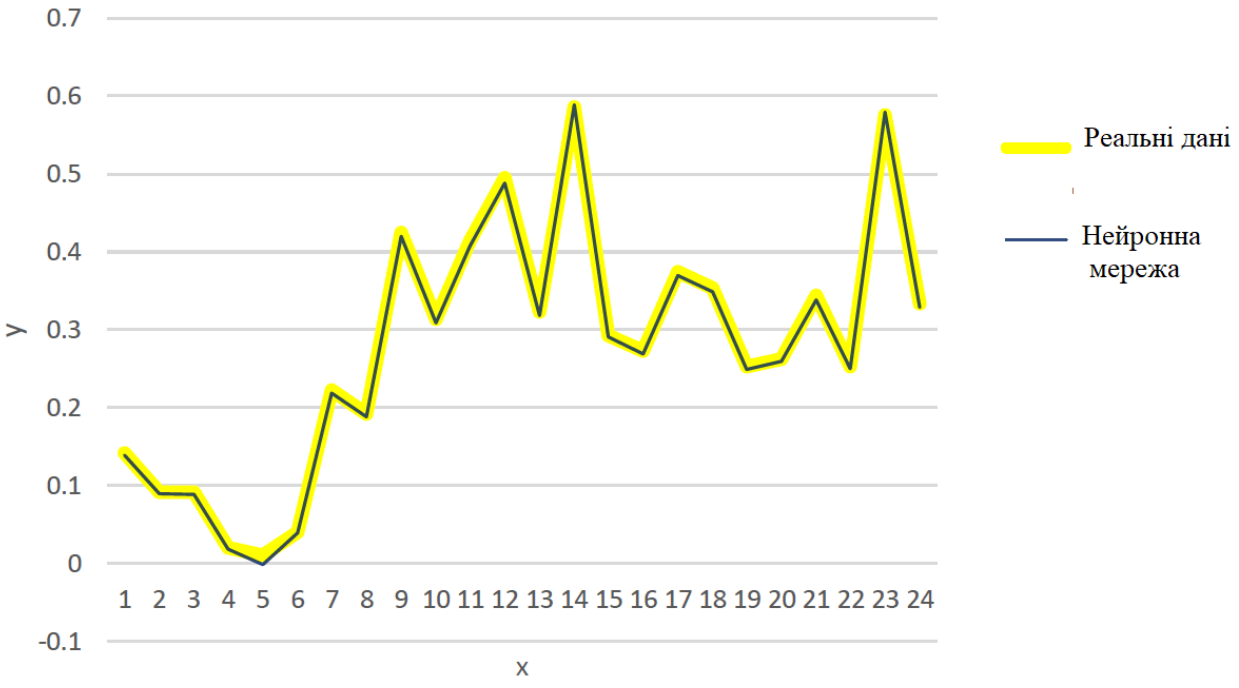


Рисунок 3.12 – Результати прогнозування даних з застосуванням нейронної мережі прямого поширення

Під час прогнозування даних для оцінки якості прогнозу використовується коефіцієнт детермінації:

$$R^2 = \frac{\sum(y_i^p - y)^2}{\sum(y_i - y)^2}, \quad (3.11)$$

де y – середнє значення елементів змінної y;

y_i^p – розраховане значення і-го елемента змінної y.

В нашому випадку $R^2 = 0{,}993$.

За значенням коефіцієнта детермінації можна зробити висновок, що прогноз є високоточним.

Таким чином, прогнозування з використанням штучних нейронних мереж є найбільш достовірним. Проте варто відзначити, що при цьому витрачається велика кількість ресурсів

ВИСНОВКИ

В результаті написання роботи розглянуто класифікацію методів прогнозування часових рядів, описано найпоширеніші моделі прогнозування, а також способи їх реалізації. Проведено аналіз описаних моделей прогнозування часових рядів, унаслідок чого визначено переваги та недоліки кожної з моделей.

Викладено етапи побудови нейронних мереж, правила вибору архітектури мережі та обробки даних, методи навчання, а також функції активації нейронів для прогнозованої мережі.

Встановлено, що використання в прогнозуванні штучних нейронних мереж цілком обґрунтоване високим значенням коефіцієнта детермінації, порівняно з іншими поширеними методами прогнозування.

На підставі отриманих даних для розв'язання задачі прогнозування було побудовано штучну нейронну мережу прямого поширення з використанням методу навчання Левенберга-Маркварда з одинадцятьма нейронами у вихідному шарі та значенням початкових вагових коефіцієнтів 0.1. Значення коефіцієнта детермінації розробленої мережі близьке до одиниці.

У зв'язку з можливістю виявляти в процесі навчання складні залежності між вхідними та вихідними даними, штучні нейронні мережі є необхідним інструментом для розв'язання задач прогнозування часових рядів.

В роботі були розглянуті наступні завдання:

1. Виконано огляд основних методів і моделей прогнозування.

2. Розглянуто математичні моделі машинного навчання для прогнозування часових рядів.

3. Розроблено математичну модель прогнозування та наведено приклад реалізації моделі для часових рядів на основі нейронної мережі прямого поширення.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних: Навч. посібник. Київ: КНЕУ, 2007. 376 с.

2. Данильченко О.М., Данильченко А.О. Аналіз даних: навч. посібник. Житомир: ЖДТУ, 2009. 405 с.

3. Черняк О.І. Інтелектуальний аналіз даних: підручник. Київ: Знання, 2014. 599с.

4. Присенко Г.В., Равікович Є.І. Прогнозування соціально-економічних процесів. Київ: КНЕУ, 2005. 256 с.

5. Берзлев, О. Ю. Методика передпрогнозного фрактального аналізу часових рядів. *Управління розвитком складних систем*. 2013. № 16. С. 76–81.

6. І. Г. Лук'яненко, В. М. Жук. Аналіз часових рядів. Побудова ARIMA, ARCH/GARCH моделей з використанням пакета E.Views 6.0: практичний посібник для роботи в комп'ютерному класі. Київ: КНЕУ, 2013. 188 с.

7. Юрченко М.Є. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика». Чернігів: ЧНТУ, 2018. 88 с.

8. Берзлев, О. Ю. Сучасний стан інформаційних систем прогнозування часових рядів. *Управління розвитком складних систем*. 2013. №13. С. 78–82.

9. Goldin, D. Q. On Similarity Queries for Time-Series Data: Constraint Specification and Implementation. *Proceedings of the 1-st International Conference on the Principles and Practice of Constraint Programming*. 1995. P. 137–153.

10. Снитюк, В. Є. Прогнозування. Моделі. Методи. Алгоритми: навч. посібник. Київ : Маклаут, 2008. 364 с.

11. Субботін С. О. Нейронні мережі : теорія та практика: навч. посіб. Житомир : Вид. О. О. Євенок, 2020. 184 с.

12. Кононюк А.Ю. Нейроні мережі і генетичні алгоритми. Київ: Корнійчук, 2008. 446 с.

13. Ethem Alpaydin. Introduction To Machine Learning, 2009. 584 p.

14. Christopher M Bishop. Pattern recognition. Machine Learning, 2006. 128 p.

15. Andrew Moore. Statistical Data Mining Tutorials. URL: http://www.autonlab.org/tutorials/ (дата звернення: 18.11.2022).

16. Michael A. Nielsen, "Neural Networks and Deep Learning". Determination Press, 2015. 587 p.

17. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. MIT Press, 2016. 446 p.

18. Sciencing: The Advantages & Disadvantages of a Regression Model. URL: https://sciencing.com/ (дата звернення: 19.11.2022).

19. Руденко О. Г. Искусственные нейронные сети: архитектуры, обучение, применения / О. Г. Руденко, Е. В. Бодянский. – Харьков: ТЕЛЕТЕХ, 2004. – 370 с.