**Лабораторна робота №2**

Тема: Апроксимація функцій з використанням нейромереж (регресія). Робота з синтетичними наборами даних.

Зв’язок між деякими взаємозалежними змінними величинами та можна описати математичним виразом у вигляді деякої невідомої функції . Для визначення характеру цього взаємозв’язку використовують апроксимацію.

В загальній постановці задачі апроксимація полягає в отриманні такої функції яка мінімізує помилку між передбаченими значеннями та реальними значеннями .

Регресія є одним із методів апроксимації функції на основі спостережень. Традиційні методи регресії передбачають знаходження аналітичної форми функції за допомогою відомих класів функцій, наприклад, прямої, параболи, поліноміальної, логарифмічної кривої – в цьому випадку завдання полягає у визначенні коефіцієнтів цих ліній (рис.1).

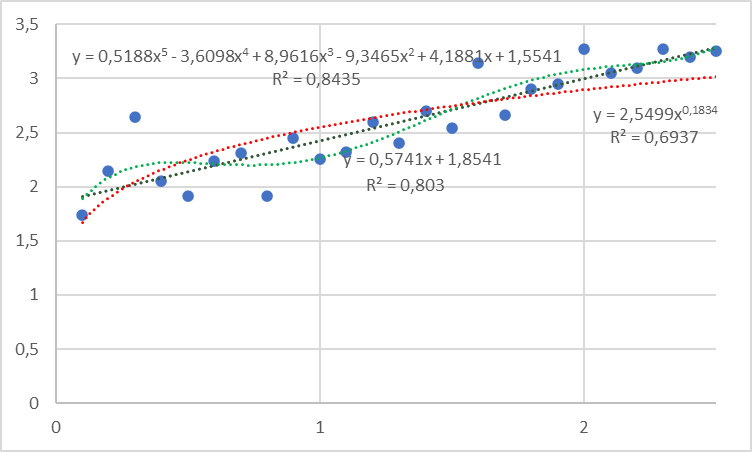


Рисунок 1 – Приклади регресії різних видів

Використання нейромережі для задачі регресії не передбачає розрахунку значень таких коефіцієнтів і не дозволяє отримати аналітичну форму функції, що описує залежність . Натомість за допомогою нейромережі ми отримуємо лише значення функції.

**Завдання 1 (2 бали):**

1. Відповідно до індивідуального варіанту завдання1 необхідно сформувати синтетичні набори даних для подальшого використання в задачі апроксимації функції однієї змінної. Синтетичні дані **обов’язково повинні містити випадковий шум**, що забезпечить їх реалістичність. Згенеруйте вибірку зі 100 таких даних та візуалізуйте її разом з цільовою функцією. Діапазон аргументів функції визначте самостійно. Для отримання максимальної оцінки за завдання вибірку формуйте програмно.
2. Розділіть згенеровану вибірку на навчальну та тестову в пропорції 80-20.
3. Оберіть та використайте готову модель нейронної мережі для вирішення задачі регресії, опишіть її характеристики.
4. Навчіть нейромережу на тренувальних даних та перевірте результати на тестовій вибірці.
5. Визначте якість регресії з використанням коефіцієнта детермінації :

,

де – середнє значення реальних даних.

Якщо , це означає, що модель ідеально передбачає всі значення.

Якщо , це означає, що модель не дає кращого результату, ніж просто взяти середнє значення.

Якщо , це означає, що модель працює гірше за випадкове передбачення.

**Завдання 2 (3 бали):**

1. Для варіанту індивідуального завдання2 визначте діапазони ознак та згенеруйте набір зі 200 пар синтетичних даних.
2. Приведіть ознаки до одного масштабу, наприклад, з використанням нормалізації чи стандартизації.
3. Визначте (придумайте) формулу для обчислення цільової функції. Цільова функція обов’язково повинна містити шум. Характер шуму визначте самостійно.
4. Сформуйте вихідні значення на основі формули з п.8.
5. Розділіть отриману вибірку на навчальну та тестову в пропорції 80-20.
6. Оберіть та використайте готову модель нейронної мережі для вирішення задачі регресії, опишіть її характеристики.
7. Навчіть нейромережу на тренувальних даних та перевірте результати на тестовій вибірці.
8. Визначте якість регресії з використанням коефіцієнта детермінації .
9. Виконайте п.11-13 для інших пропорцій розподілу навчальної та тестової вибірки та інших видів шумів, узагальніть результати у вигляді таблиці (проведіть не менше 5 різних експериментів).
10. Проведіть аналіз отриманих результатів.

**Зміст звіту:**

1. Титульний аркуш.
2. Варіант індивідуального завдання1.
3. Архітектура нейромережі та опис її параметрів.
4. Візуалізація точок вхідних даних та графіку цільової функції.
5. Результати моделювання завдання 1(оцінка якості моделі).
6. Екранні форми з результатами роботи програми 1.
7. Лістинг програми 1.
8. Варіант індивідуального завдання2.
9. Опис ознак, спосіб їх масштабування, опис цільової функції.
10. Архітектура нейромережі та опис її параметрів.
11. Результати моделювання завдання 2(таблиця з оцінками якості моделі для різних експериментів).
12. Екранні форми з результатами роботи програми 2.
13. Лістинг програми 2.

Таблиця 1 – Варіанти для завдання 1

| **№** | Цільова функція |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Таблиця 2 – Варіанти для завдання 2

|  |  |
| --- | --- |
| **№** | **Варіант завдання** |
|  | Передбачити температуру об'єкта (T) на основі часу нагрівання (t) та потужності нагрівача (P). |
|  | Передбачити рівень вологості (%) на основі температури (°C) та швидкості вітру (м/с). |
|  | Передбачити швидкість випаровування рідини (г/год) на основі температури (°C) та площі відкритої поверхні (см²). |
|  | Передбачити рівень забруднення повітря (PM2.5, мкг/м³) на основі трафіку транспорту (авто/год) та температури (°C). |
|  | Передбачити рівень шуму (дБ) на основі інтенсивності трафіку (авто/хв) та відстані до дороги (м). |
|  | Передбачити вартість квартири (грн) на основі площі (м²) та відстані до центру (км). |
|  | Передбачити тепловтрати будинку (Вт/м²) на основі товщини утеплювача (см) та зовнішньої температури (°C). |
|  | Передбачити рівень освітленості в кімнаті (люкс) на основі розміру вікна (м²) та кількості ламп. |
|  | Передбачити вологість всередині будівлі (%) на основі температури повітря (°C) та кількості людей у приміщенні. |
|  | Передбачити час висихання фарби (год) на основі вологості (%) та температури (°C). |
|  | Передбачити рівень зношеності шин (%) на основі пробігу (км) та середньої швидкості (км/год). |
|  | Передбачити інтенсивність руху (авто/год) на основі часу доби (год) та дня тижня. |
|  | Передбачити прибуток компанії (грн) на основі витрат на рекламу (грн) та кількості клієнтів |
|  | Передбачити середній бал студента на основі кількості годин навчання на тиждень та кількості відвіданих лекцій |
|  | Передбачити час проходження тесту (хв) на основі складності тесту (бал) та рівня підготовки студента (годин навчання). |
|  | Передбачити рівень стресу людини (бал) на основі кількості робочих годин у тиждень та часу сну (год/день). |
|  | Передбачити тривалість перегляду відео (хв) на основі кількості переглядів та середньої оцінки користувачів. |
|  | Передбачити рівень цукру в крові (ммоль/л) на основі кількості з'їдених вуглеводів (г) та часу після прийому їжі (хв). |
|  | Передбачити тривалість життя людини (роки) на основі індексу маси тіла (ІМТ) та кількості годин фізичної активності на тиждень. |
|  | Передбачити рівень продажів книги (шт/місяць) на основі середньої оцінки читачів (бал) та витрат на рекламу (грн). |