

Лабораторна робота №7

ІІЗ-21-5 Пархомчук Іван

Мета: набути навичок працювати з даними і опонувати роботу у Python з використанням теореми Байєса.

Завдання 1. Ретельно опрацювати теоретичні відомості:

- Теорема Байєса

Теорема Байєса дозволяє визначити ймовірність гіпотези на основі спостережуваних даних. Формула виглядає так:

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

Де:

$P(H|E)$ — ймовірність гіпотези ННН, якщо спостерігаються дані ЕЕЕ (апостеріорна ймовірність);

$P(E|H)$ — ймовірність отримання даних ЕЕЕ за умови, що ННН істинна (правдоподібність);

$P(H)$ — апріорна ймовірність гіпотези ННН;

$P(E)$ — ймовірність спостереження ЕЕЕ.

Теорема широко використовується в машинному навчанні для класифікації та оцінки ризиків.

- Типи наївного байєсівського класифікатора

Наївний байєсівський класифікатор має кілька різновидів, які застосовуються залежно від типу даних:

Гаусівський (Gaussian Naive Bayes):

Використовується для безперервних даних.

Припускає, що дані розподілені нормально (гаусівський розподіл).

Мультиноміальний (Multinomial Naive Bayes):

Підходить для текстової класифікації (наприклад, аналіз тональності тексту).

Використовується, коли атрибути є дискретними (наприклад, кількість разів, коли слово зустрічається в документі).

Бернуллівський (Bernoulli Naive Bayes):

Використовується для бінарних даних.

Застосовується, коли потрібно враховувати лише наявність або відсутність ознаки.

- Де використовується Наївний Байєс

Текстова класифікація: Фільтрація спаму, Аналіз тональності тексту, Класифікація документів.

Медицина: Діагностика захворювань на основі симптомів.

Фінанси: Оцінка кредитоспроможності, Виявлення шахрайства.

Рекомендаційні системи: Прогнозування переваг користувачів.

Обробка зображень: Розпізнавання об'єктів і класифікація зображень.

Завдання 2. Ретельно розібрати приклад: прогнозування з використанням теореми Байєса.

Розрахунок правдоподібностей для кожної умови:

$$P(\text{Outlook} = \text{Rain} \mid \text{Yes}) = \frac{3}{9}$$

$$P(\text{Humidity} = \text{High} \mid \text{Yes}) = \frac{3}{9}$$

$$P(\text{Wind} = \text{Weak} \mid \text{Yes}) = \frac{6}{9}$$

$$P(\text{Yes}) = \frac{9}{14}$$

Тоді:

$$P(\text{Yes}) = P(\text{Outlook} = \text{Rain} | \text{Yes}) \times P(\text{Humidity} = \text{High} | \text{Yes}) \times P(\text{Wind} = \text{Weak} | \text{Yes}) \times P(\text{Yes})$$

Розрахунок для "No":

$$P(\text{Outlook} = \text{Rain} | \text{No}) = \frac{2}{5}$$

$$P(\text{Humidity} = \text{High} | \text{No}) = \frac{4}{5}$$

$$P(\text{Wind} = \text{Weak} | \text{No}) = \frac{2}{5}$$

$$P(\text{No}) = \frac{5}{14}$$

Тоді:

$$P(\text{No}) = P(\text{Outlook} = \text{Rain} | \text{No}) \times P(\text{Humidity} = \text{High} | \text{No}) \times P(\text{Wind} = \text{Weak} | \text{No}) \times P(\text{No})$$

Нормалізація результатів:

Після обчислення значень для "Yes" та "No", нормалізуємо їх:

$$P(\text{Yes} | \text{дані}) = \frac{P(\text{Yes})}{P(\text{Yes}) + P(\text{No})}$$

$$P(\text{No} | \text{дані}) = \frac{P(\text{No})}{P(\text{Yes}) + P(\text{No})}$$

Висновок: набув навичок працювати з даними із використанням теореми Байєса.