

Міністерство освіти і науки України
Національний авіаційний університет
Навчально-науковий інститут комп'ютерних інформаційних технологій
Кафедра комп'ютеризованих систем управління

Курсова робота
з дисципліни «Системне програмне забезпечення»

Пояснювальна записка
Тема: реалізація наївного баєсового класифікатора
на мові програмування Python

Виконав:
студент групи СП-325
Клокун В. Д.

Київ — 2018

**Завдання на виконання курсової роботи
студента групи СП-325 Клокуна Владислава Денисовича**

1. Тема курсової роботи: реалізація наївного баєсового класифікатора на мові програмування Python для класифікації спостережень, що містять неперервні дані.
2. Термін виконання курсової роботи:
з « ____ » _____ 2018 р. по « ____ » _____ 2018 р.
3. Вхідні дані до роботи: набір даних для класифікації.
4. Етапи виконання курсової роботи:
 - Огляд теоретичних відомостей про наївний баєсов класифікатор.
 - Реалізація та тестування наївного баєсового класифікатора.
5. Перелік обов'язкових додатків і графічного матеріалу:
 - FIXME.

Завдання отримав: « ____ » _____ 2018 р.

Підпис студента: _____ (Клокун В. Д.)

Зміст

1. Теоретична частина	4
1.1. Короткі теоретичні відомості	4
2. Практична частина	6
Висновки	7

1. Теоретична частина

1.1. Короткі теоретичні відомості

Припустимо, що в ході деякого експерименту проводились спостереження, під час проведення яких збирались неперервні (недискретні) дані про результат події. Також були визначені категорії (або класи), до яких ці дані можуть належати. Поставлена задача класифікувати дані спостережень. *Класифікація* — це задача визначення, до якої з категорій належить певне спостереження [1]. *Класифікатор* — це алгоритм, який виконує класифікацію [1].

Наївний баєсів класифікатор — це ймовірнісний класифікатор, який використовує теорему Баєса для класифікації спостережень. Такі класифікатори отримують на вхід спостереження, оцінюють його і роблять припущення про клас, до якого воно належить. Вхідні дані, тобто спостереження, представляються у вигляді вектора відомих значень випадкових змінних, які називаються *ознаками*. Результатом роботи класифікатора є певне значення цільової змінної або змінних, які зазвичай називаються класовими, і позначають клас, до якого належить спостереження.

Принцип класифікації полягає в обчисленні умовних імовірностей (визначення 1) того, що вхідні дані належать до кожного з класів (події, які нас цікавлять), за умови, що ознаки мають певні значення (події, які ми спостерігаємо). Після обчислення кожної з умовних імовірностей знаходиться найбільша та робиться висновок про належність спостереження до кожного з класів. Саме тому наївний баєсів класифікатор називають ймовірнісним.

Визначення 1 (Умовна ймовірність). Нехай A і B — події. Позначимо ймовірність настання кожної з них незалежно одна від одної як $P(A)$ і $P(B)$ відповідно. Тоді умовною імовірністю $P(A | B)$ називається ймовірність настання події A за умови, що подія B настала.

Наприклад, нехай подія A — дане спостереження належить до певного класу, подія B — ознаки спостереження мають певні значення. Тоді щоб знайти ймовірність, що дане спостереження з певним значенням ознак належить до певного класу, необхідно обчислити умовну ймовірність $P(A | B)$. Для обчислення цієї імовірності необхідно використати теорему Баєса (теорема 1).

Теорема 1 (Баєса). Нехай $P(A | B)$ — умовна ймовірність настання події A

за умови, що подія B настала; $P(B | A)$ — умовна ймовірність настання події B за умови, що подія A настала і $P(B)$ — ймовірність настання події B , причому $P(B) \neq 0$. Тоді умовна ймовірність $P(A | B)$ обчислюється так:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}.$$

Оскільки при обчисленні умовних ймовірностей використовується теорема Баєса, такий ймовірнісний класифікатор називається баєсовим.

1.2. Класифікація неперервних даних за допомогою найвного баєсового класифікатора

2. Практична частина

Висновки

Література

1. Statistical classification. — URL: https://en.wikipedia.org/wiki/Statistical_classification (дата зверн. 20.11.2018).