

ПРОСТОРОВИЙ АНАЛІЗ ТА СТАТИСТИЧНЕ МОДЕЛЮВАННЯ В ГЕОЛОГІЇ

Практичне робота 5 (2023)

Завдання 1 : Дано набір даних і метою є класифікація нової точки (5, 7) за допомогою k-nearest neighbor класифікацію.

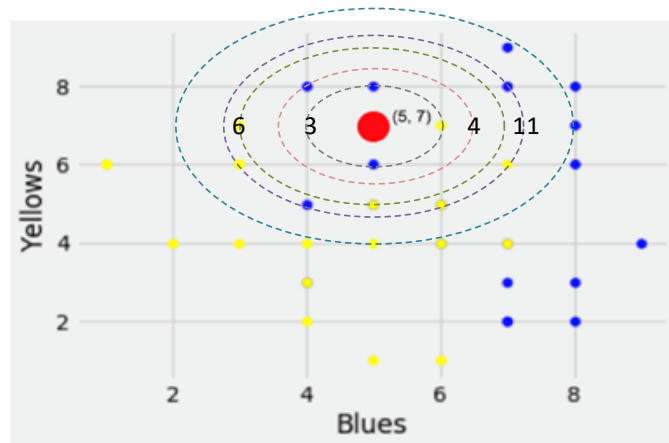
pointsBlue

X	Y
8	2
7	2
6	5
7	2
5	6
9	4
... (14 rows omitted)	

pointsYellow

X	Y
6	5
4	3
1	6
1	6
3	4
5	5
... (14 rows omitted)	

Ось діаграма розсіювання даних. Спробуємо класифікувати точку (5, 7).



Який колір точки (5, 7), якщо Ви використовуєте класифікатор з **3 найближчими сусідами**?

2 сині, 1 жовта

Blue

Yellow

Який колір точки (5, 7), якщо Ви використовуєте класифікатор з **5 найближчими сусідами**?

3 сині, 2 жовті

Blue

Yellow

Який колір точки (5, 7), якщо Ви використовуєте класифікатор з **7 найближчими сусідами**?

5 сині, 6 жовті – вибір залежить від того, яку точку додати сьомою раніше

Blue

Yellow

Завдання 2 : Оберіть Вашу відповідь для наступних тверджень:

У класифікації методом k-найближчих сусідів збільшення значення k завжди підвищить точність класифікації для тестової вибірки?

Так

Ні

У класифікації методом k-найближчих сусідів збільшення значення k ніколи не підвищує точність класифікації для тестової вибірки?

*Збільшення k зробить класифікацію **більш загальною** (biased), а зменшення призведе до перенавчання і росту помилок (variance), тому вибір правильного k – це пошук компромісу.*

Так

Ні

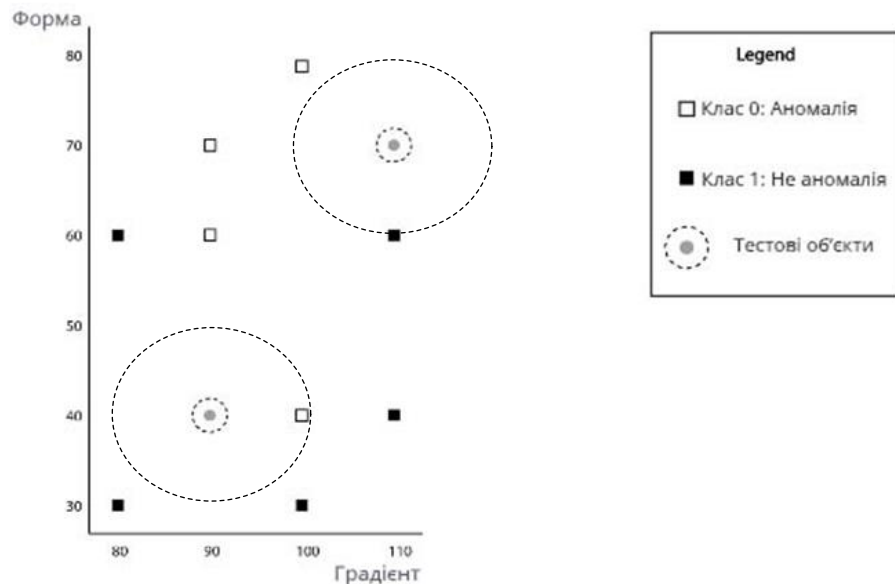
У класифікації методом k-найближчих сусідів слід вибирати тестову вибірку, яка включає деякі точки даних із навчальної вибірки та деякі точки даних, яких немає в навчальній вибірці.

Тестова вибірка повинна складатися виключно з точок даних, яких нема в навчальній

Так

Ні

Завдання 3 : Ми хочемо спрогнозувати, чи є об'єкт аномалією (клас 0) чи ні (клас 1), на основі двох атрибутів: (а) **форми** та (б) **градієнту**. У нас є навчальна вибірка із 9 об'єктів і тестову вибірку із 2 об'єктів, як показано на діаграмі нижче. Немає двох об'єктів однакової форми і градієнтами.



Перший об'єкт із тестової вибірки має градієнт 90 ($1/\text{см}^2$) і форму 40 (безрозмірна величина). Який клас передбачить класифікатор найближчого сусіда з **1 найближчим сусідом** для цього об'єкту?

0-клас (аномалія)

1-клас (не аномалія)

Який клас передбачить класифікатор з **3 найближчими сусідами** для цього об'єкту?

0-клас (аномалія)

1-клас (не аномалія)

Який клас передбачить класифікатор з **5 найближчими сусідами** для цього об'єкту?

0-клас (аномалія)

1-клас (не аномалія)

Завдання 4 : Припустимо, у нас є ще **два об'єкти**: один формою 41 (безрозмірна величина) і градієнтом 100 ($1/\text{см}^2$), а інший формою 52 (безрозмірна величина) і градієнтом 100 ($1/\text{см}^2$). Використовуючи визначення відстані, надане в лекції, знайдіть відстань між цими двома об'єктами для використання методу k-найближчих сусідів, якщо обрами форму та градієнт як єдині дві ознаки.

Який клас передбачить класифікатор з **3 найближчими сусідами** для об'єкту формою 41 (безрозмірна величина) і градієнтом 100 ($1/\text{см}^2$)?

0-клас (аномалія)

1-клас (не аномалія)

Який клас передбачить класифікатор з 3 найближчими сусідами для об'єкту формою 52 (безрозмірна величина) і градієнтом 100 (1/см²)?

рішення залежить від того, яку точку додати третьою раніше

0-клас (аномалія)

1-клас (не аномалія)

Завдання 5:

Виконайте інструкції наведені в файлі [pr5_5.ipynb](#) з директорії pr5_5 (архіву pr5_5.zip). Приєднайте до вашого проекту JUPYTER NOTEBOOK систему контролю версій Git та розмістіть його на GitHub надавши посилання нижче:

[Вставити посилання на Вашу Практичну роботу 5 \(2023\) завдання 5 5 на GitHub](#)