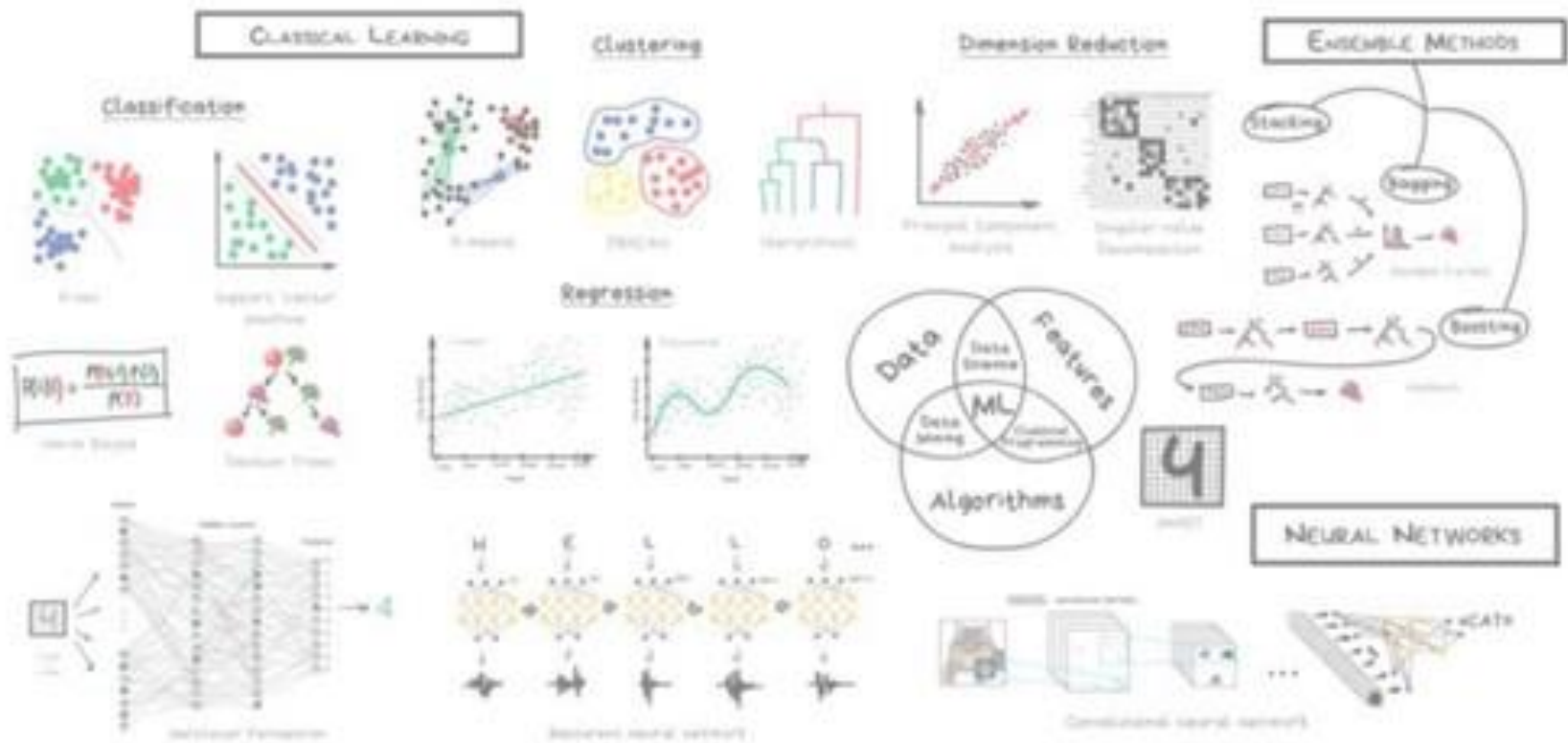


Машинне навчання

Класифікація методом kNN



Лабораторна робота №2

Метод класифікації kNN

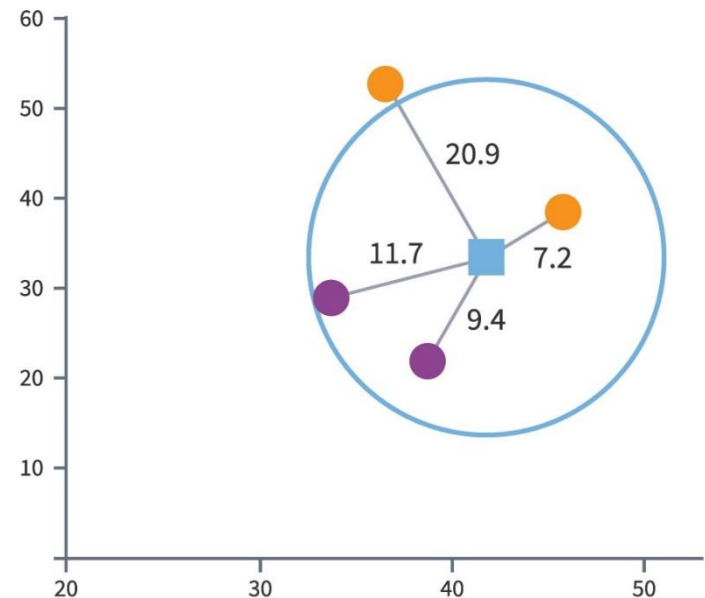
Постановка задачі

Розглянемо задачу класифікації, в якій об'єктам з $X = \mathbb{R}^n$ (об'єкти описуються n числовими признаками): $X = \{x_1 \dots x_n\}$ відповідає один з кількох класів $Y = \{0, \dots M\}$. Нехай задана вибірка пар "об'єкт-відповідь": $(x_i, y_i), i = 1 \dots N$. Необхідно класифікувати новий об'єкт.

Приклад алгоритму

f_1 Вік	f_2 Дохід	Y
46	41	No
36	54	No
34	29	Yes
38	23	Yes

Задача полягає в тому, щоб провести класифікацію нового об'єкта для якого $f_1 = 42$ и $f_2 = 34$



Метод класифікації kNN

Етапи розв'язання

1. Зафіксувати значення параметра k
2. Для кожного об'єкта з тестової вибірки:
Розрахувати відстань від об'єкта з набором ознак a_i до кожного об'єкта з навчальної вибірки з ознаками f_i за формулою

$$D_j = \sqrt{\sum_{i=1}^l (f_i - a_i)^2}$$

Розрахувати коефіцієнт близькості

$$Q_j = \sum_{i=1}^{n_j} \frac{1}{D^2(f, a_{ij})}$$

Визначити до якого класу з навчальної вибірки належить об'єкт

3. Порахувати кількість помилок класифікації шляхом порівняння результатів класифікації з відомими значеннями класів для кожного об'єкта з тестової вибірки
4. Провести оптимізацію параметра k
5. Порівняти результати з kNN з sklearn
6. Оформити результати у вигляді звіту

Метод класифікації kNN

Постановка задачі на 60 балів

1. З використанням вбудованого алгоритму **kNN** з бібліотеки **sklearn** провести класифікацію об'єктів. Визначити кількість помилок.
2. Перевірити результати роботи на тестовій вибірці.
3. Оформити результати у вигляді звіту.