

Задача кластеризации. Методы (Часть 1)

Задание 1

Используйте метод ближайших соседей (k -Nearest Neighbors, k -NN) для задачи классификации на предоставленном наборе данных. Определите качество классификации при $k=3$. Разбейте данные на обучающую и тестовую выборки, и рассчитайте точность на тестовой выборке.

- 1)** Выполните кластеризацию с использованием метода ближайших соседей.
- 2)** Оцените точность на тестовой выборке.

Задание 2

Изучите, как влияет число соседей k на качество классификации. Используйте значения k от 1 до 10. Постройте график зависимости точности классификации от значения k на тестовой выборке.

- 1)** Реализуйте цикл для k от 1 до 10.
- 2)** Постройте график зависимости точности от k .

Задание 3

Используя кросс-валидацию, оцените точность метода ближайших соседей для разных значений k . Выполните кросс-валидацию с 5 фолдами и определите, какое k обеспечивает наибольшую точность.

- 1)** Выполните кросс-валидацию с использованием метода ближайших соседей.
- 2)** Рассчитайте коэффициент силуэта для каждого k .
- 3)** Определите оптимальное значение k .

Критерии оценивания

K1 Выполнено задание 1 2 балл

K2 Выполнено задание 2 2 балл

K3 Выполнено задание 3 1 балл

Максимальное количество баллов 5 баллов

Минимальное количество баллов
чтобы преподаватель смог зачесть вашу работу 2 балла