

Екзаменаційний білет 8

1. Алгоритм к-найближчих сусідів, к- зважених найближчих сусідів, адаптація методу під задачу регресії.

Алгоритм **к-найближчих сусідів** здійснює класифікацію розглядаючи k найближчих сусідів до вказаного об'єкту. Для цього алгоритм вважає що всі об'єкти є точками у певному n -вимірному просторі, та визначає сусідів в термінах певної відстані. При цьому враховується k найближчих сусідів.

Основна ідея: k найближчих сусідів голосують за клас

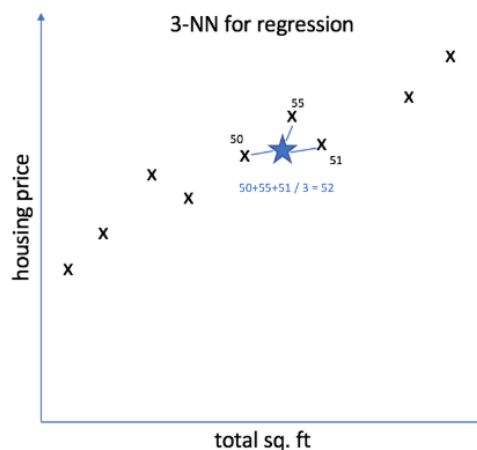
Дистанція може рахуватися за різними правилами, наприклад:

$$\text{Евклідова: } \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Можливо враховувати відстань, тоді це алгоритм k -зважених найближчих сусідів (ближчий сусід має більший вплив):

$$\operatorname{argmax} \sum_{i=1}^K [y^i] \omega_i, \quad \omega_i = \frac{1}{d(x_i, x_{\text{test}})}$$

Адаптація методу **під задачу регресії** полягає у пошуку найближчих точок, та отримання результату як середнє арифметичне їх значень.



2. Візуалізація даних методом t-SNE

Метод SNE (стохастичного вбудовування сусідів) – нелінійна техніка зменшення розмірності, яка застосовується для візуалізації даних у 2d або 3d

просторі таким чином, щоб схожі об'єкти моделювалися як близькі один до одного точки, а різні об'єкти – далекими точками з великою ймовірністю.

t-SNE (t-розподілене SNE) робить задачу простішою та вирішує проблему скупченості. Для цього, використовується розподіл Стюдента (t-розподіл) замість нормального при моделюванні у маловимірному просторі: