

Метрическая классификация

Широбокова Маргарита

Data Scientist, Team Lead,
LegalDoc.AI

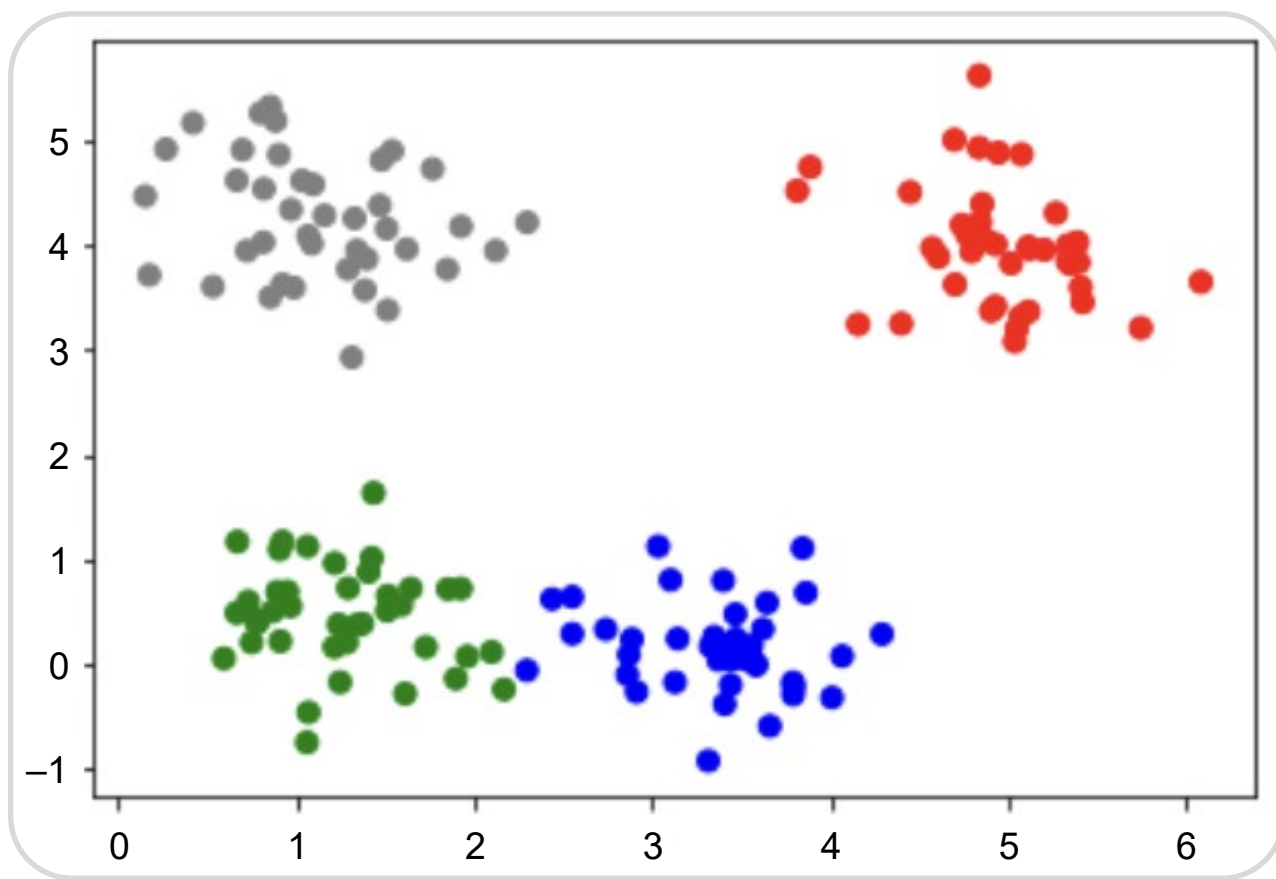
Skillbox

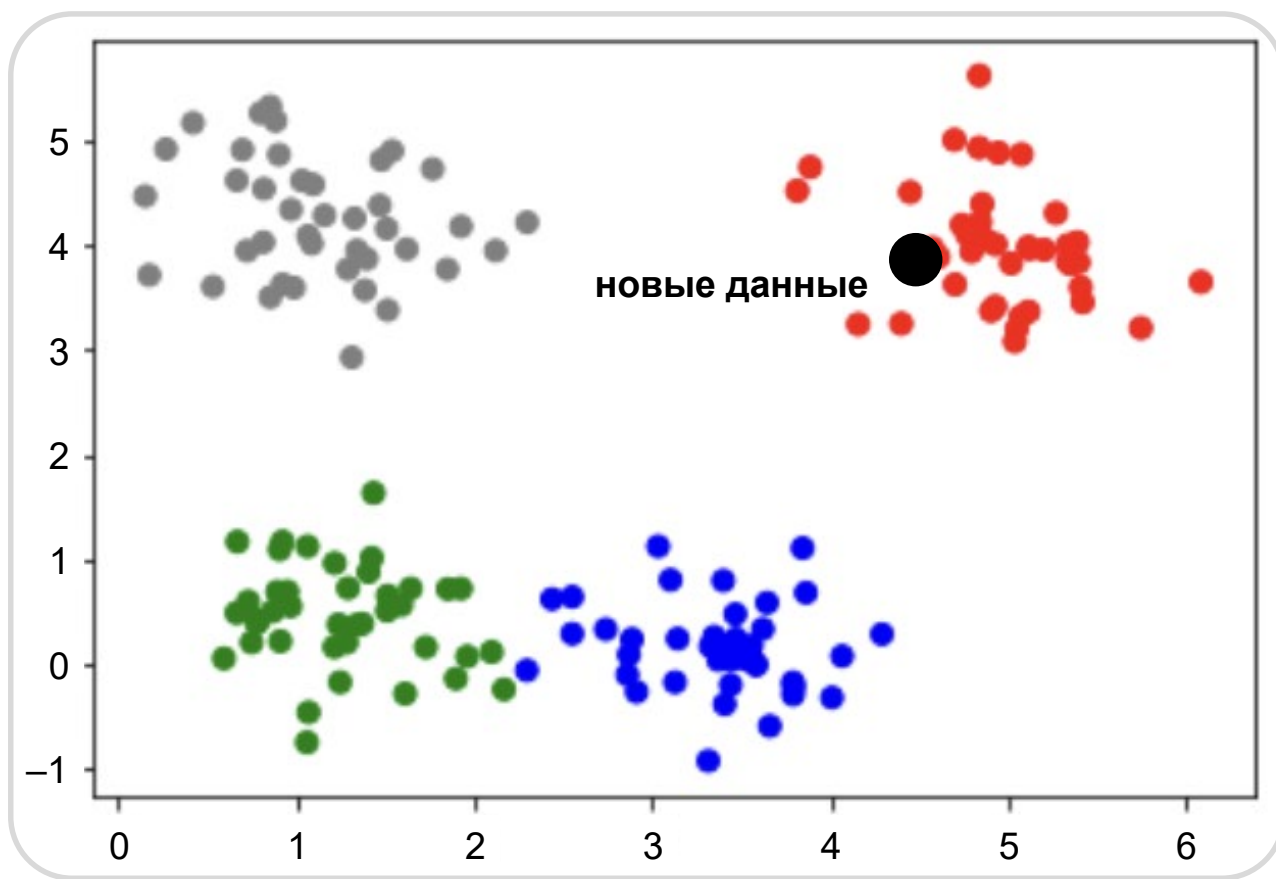
Метрическая классификация

Метрики и метрические методы. Метод ближайших соседей

Skillbox

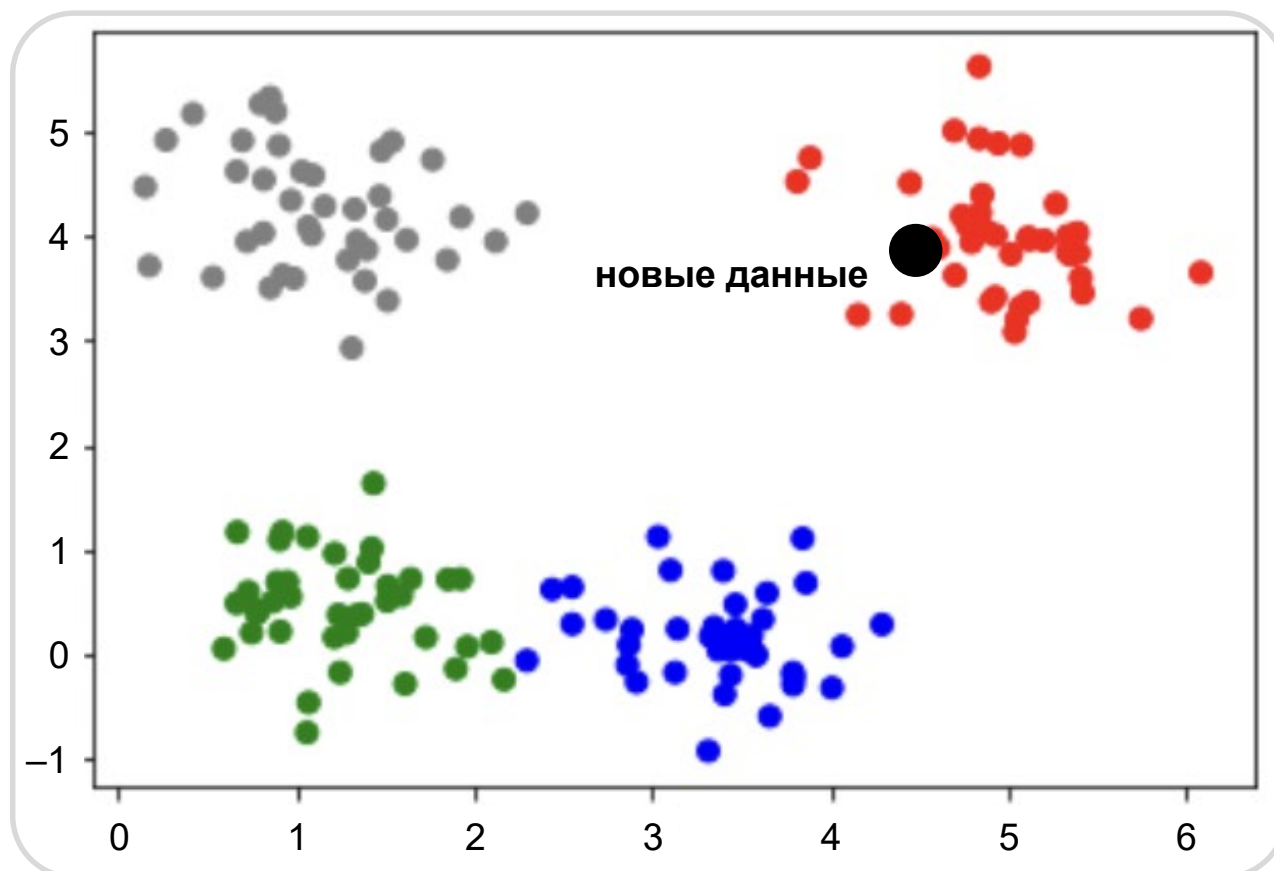
образовательная платформа





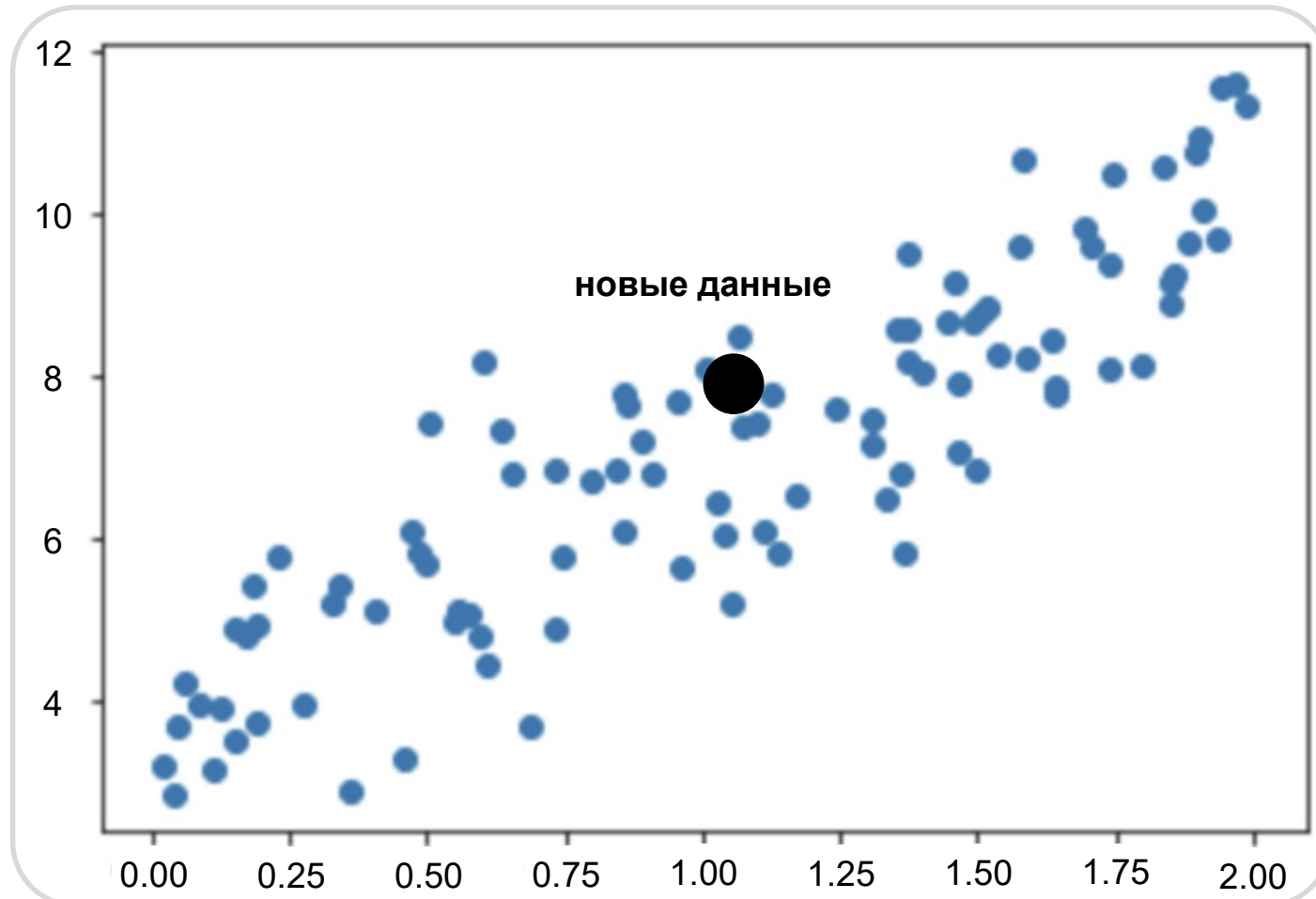
Гипотеза компактности (для классификации)

Гипотеза компактности — в задачах классификации предположение о том, что схожие объекты гораздо чаще лежат в одном классе, чем в разных.



Гипотеза непрерывности (для регрессии)

Гипотеза непрерывности — это предположение о том, что близким объектам соответствуют близкие ответы.



Ленивое обучение

Не применяются заранее обученные модели,
все имеющиеся данные используются во время
классификации.



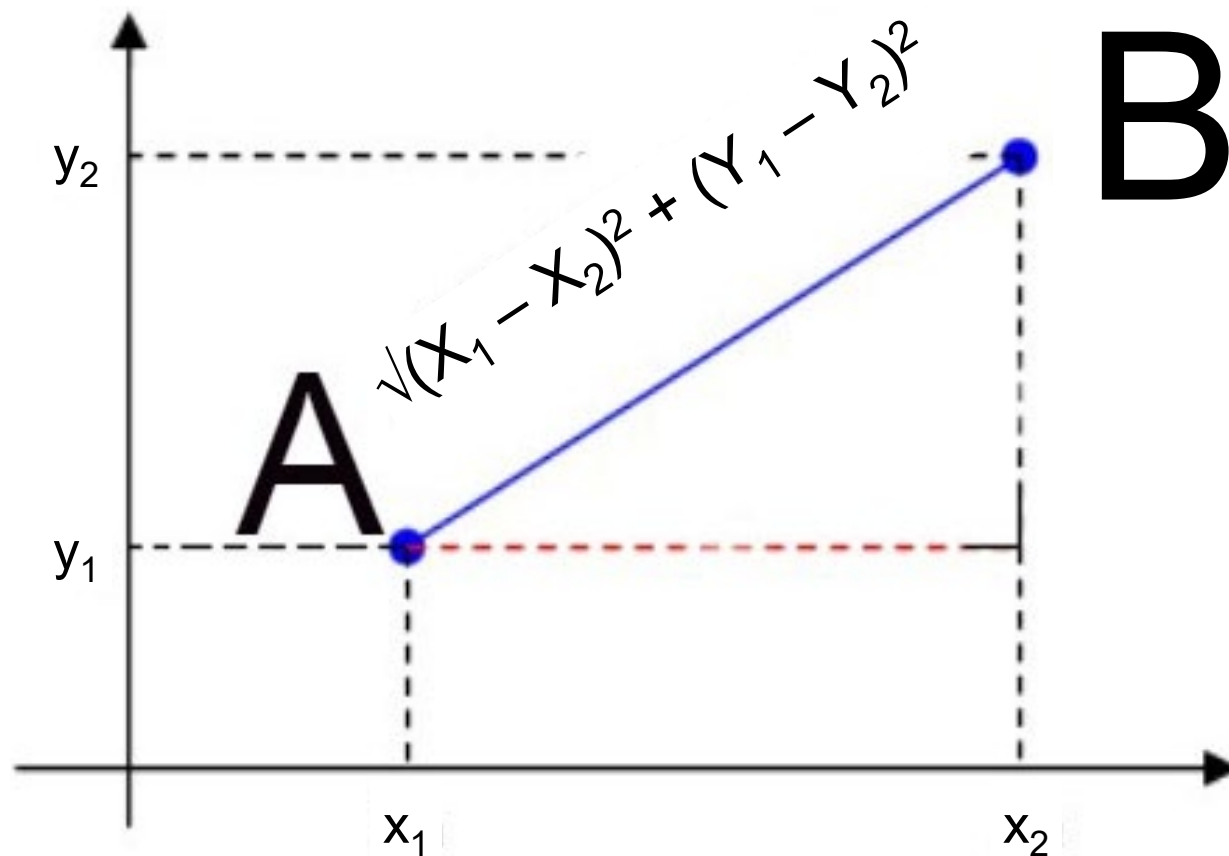
Пусть множество X — объекты, а множество Y — ответы;

то есть $XI = (x_i, y_i)$ — обучающая выборка

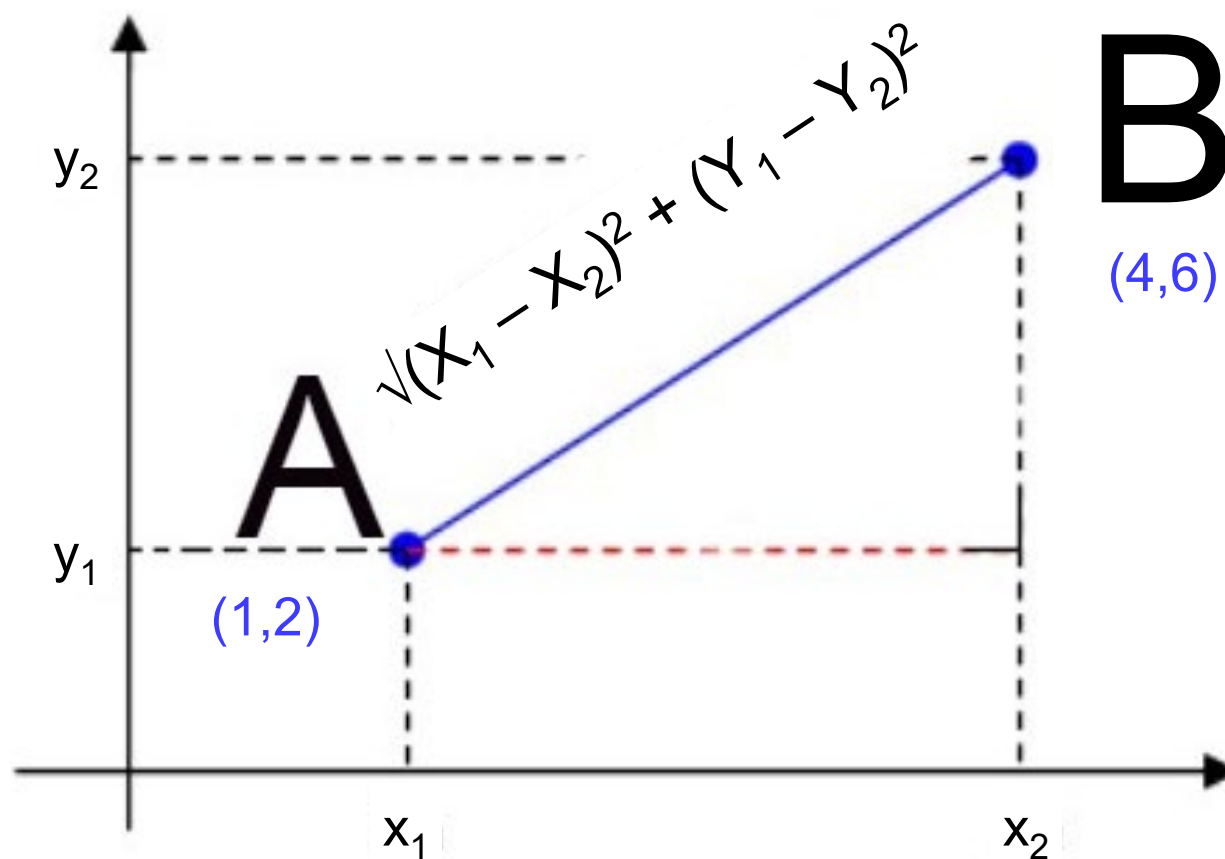
Тогда функция расстояния $\rho : X \times X \rightarrow [0, \infty)$

Евклидово расстояние

Евклидово расстояние — это геометрическое расстояние между двумя точками с координатами А и В.



Евклидово расстояние



$$p = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2} = (4 - 1)^2 + (6 - 2)^2 = 3^2 + 4^2 = 9 + 16 = \sqrt{25} = 5$$

Евклидово расстояние

Расстояние между объектами, описанными двумя признаками

$$\rho(A, B) = \sqrt{(x_{A1} - x_{B1})^2 + (x_{A2} - x_{B2})^2}$$

Расстояние между объектами, описанными тремя признаками

$$\rho(A, B) = \sqrt{(x_{A1} - x_{B1})^2 + (x_{A2} - x_{B2})^2 + (x_{A3} - x_{B3})^2}$$

Расстояние между объектами, описанными k признаками

$$\rho(A, B) = \sqrt{(x_{A1} - x_{B1})^2 + (x_{A2} - x_{B2})^2 + (x_{A3} - x_{B3})^2 + \dots + (x_{Ak} - x_{Bk})^2} =$$
$$\sqrt{\sum_{i=1}^k (x_{Ai} - x_{Bi})^2}$$

где x_{A1}, x_{B1} — координаты объектов A и B в k-мерном пространстве