

Машинное обучение.

Домашнее задание №3

Задача 1. Приведите пример целевого вектора y_1, \dots, y_ℓ и ответов алгоритма $b(x_1), \dots, b(x_\ell)$, для которых AUC-ROC будет равен 0.5.

Задача 2. Пусть дан классификатор $b(x)$, который возвращает оценки принадлежности объектов классам: чем больше ответ классификатора, тем более он уверен в том, что данный объект относится к классу «+1». Отсортируем все объекты по неубыванию ответа классификатора b : $x_{(1)}, \dots, x_{(\ell)}$. Обозначим истинные ответы на этих объектах через $y_{(1)}, \dots, y_{(\ell)}$.

Покажите, что AUC-ROC для данной выборки будет равен вероятности того, что случайно выбранный положительный объект окажется в отсортированном списке позже случайно выбранного отрицательного объекта.

Задача 3. Пусть даны выборка X , состоящая из 8 объектов, и классификатор $b(x)$, предсказывающий оценку принадлежности объекта положительному классу. Предсказания $b(x)$ и реальные метки объектов приведены ниже:

$$\begin{aligned}b(x_1) &= 0.1, & y_1 &= +1, \\b(x_2) &= 0.8, & y_2 &= +1, \\b(x_3) &= 0.2, & y_3 &= -1, \\b(x_4) &= 0.25, & y_4 &= -1, \\b(x_5) &= 0.9, & y_5 &= +1, \\b(x_6) &= 0.3, & y_6 &= +1, \\b(x_7) &= 0.6, & y_7 &= -1, \\b(x_8) &= 0.95, & y_8 &= +1.\end{aligned}$$

Постройте ROC-кривую и вычислите AUC-ROC для множества классификаторов $a(x; t)$, порожденных $b(x)$, на выборке X .