

Машинное обучение.

Домашнее задание №6

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. Допустим, мы решаем задачу регрессии с помощью линейной модели, настраиваемой на среднеквадратичный функционал. Как соотносятся выставление весов объектов в данном функционале и oversampling/undersampling?