

# Машинное обучение, ФКН ВШЭ

## Метрики классификации

**Задача 1.** Бандерлог из Лога<sup>1</sup> ведёт свой блог, любит считать логарифмы и оценивать логистические регрессии. С помощью нового классификатора  $b(x)$ , предсказывающего оценку принадлежности объекта положительному классу, Бандерлог решил задачу классификации на 8 объектах. Предсказания  $b(x)$  и реальные метки объектов приведены ниже:

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

1. Постройте ROC-кривую и вычислите AUC-ROC для множества классификаторов  $a(x; t)$ , порожденных  $b(x)$ , на выборке  $X$ .
2. Постройте PR-кривую и найдите площадь под ней для того же множества классификаторов.
3. Как по-английски будет «бревно»?

**Задача 2.** Пусть дан классификатор  $b(x)$ , который возвращает оценку принадлежности объекта  $x$  положительному классу. Отсортируем все объекты по неубыванию ответа классификатора:  $b(x_{(1)}) \leq \dots \leq b(x_{(\ell)})$ . Обозначим истинные ответы на этих объектах через  $y_{(1)}, \dots, y_{(\ell)}$ .

1. Покажите, что AUC-ROC для данной выборки будет равен вероятности того, что случайно выбранный положительный объект окажется в отсортированном списке не раньше случайно выбранного отрицательного объекта.
2. Покажите, что число дефектных пар в выборке  $y_{(1)}, \dots, y_{(\ell)}$  будет совпадать с числом итераций, которые нужно сделать для того, чтобы отсортировать этот массив с помощью сортировки пузырьком.

---

<sup>1</sup>деревня в Кадуйском районе Вологодской области

**Задача 3.** Пусть дана некоторая выборка  $X$  и классификатор  $b(x)$ , возвращающий в качестве оценки принадлежности объекта  $x$  положительному классу 0 или 1 (а не некоторое вещественное число, как предполагалось на семинарах).

1. Постройте ROC-кривую для классификатора  $b(x)$  на выборке  $X$ .
2. Покажите, что AUC-ROC классификатора  $b(x)$  на выборке  $X$  может быть выражен через долю правильных ответов и полноту классификатора  $a(x; t)$ , получающегося при выборе некоторого порога  $t \in (0; 1)$ . Помимо указанных величин в формулу могут входить только величины  $\ell_-$ ,  $\ell_+$ ,  $\ell$  (количество отрицательных, положительных и общее количество объектов в выборке  $X$  соответственно).
3. Покажите, что в случае сбалансированной выборки ( $\ell_- = \ell_+$ ) AUC-ROC классификатора  $b(x)$  на выборке  $X$  совпадает с долей правильных ответов классификатора при выборе некоторого порога  $t \in (0; 1)$ .

**Задача 4.** В анализе данных для сравнения среднего значения некоторой величины у объектов двух выборок часто используется критерий Манна–Уитни–Уилкоксона<sup>2</sup>, основанный на вычислении  $U$ -статистики.

Пусть у нас имеется выборка  $X$  и классификатор  $b(x)$ , возвращающий оценку принадлежности объекта  $x$  положительному классу. Тогда вычисление  $U$ -статистики для подвыборки  $X$ , состоящей из объектов положительного класса, производится следующим образом: объекты обеих выборок сортируются по неубыванию значения  $b(x)$ , после чего каждому объекту в полученном упорядоченном ряду  $x_{(1)}, \dots, x_{(\ell)}$  присваивается ранг — номер позиции  $r_{(i)}$  в ряду (начиная с 1, при этом для объектов с одинаковыми значением  $b(x)$  в качестве ранга присваивается среднее значение ранга для таких объектов). Тогда  $U$ -статистика для объектов положительного класса равна:

$$U_+ = \sum_{\substack{i=1 \\ y_{(i)}=+1}}^{\ell} r_{(i)} - \frac{\ell_+(\ell_+ + 1)}{2}.$$

Покажите, что для значения AUC-ROC классификатора  $b(x)$  на выборке  $X$  и  $U$ -статистики верно следующее соотношение:

$$\text{AUC} = \frac{U_+}{\ell_- \ell_+}.$$

**Задача 5.** Женья хочет построить в этом новом мире атмосферу доброты и взаимоподдержки. Для этого он собирается обучить классификатор токсичных сообщений, чтобы тот их автоматически банил. В выборке Жени половина сообщений токсичные. В обучающей 9000 сообщений, в тестовой 1000 сообщений.

1. В качестве порога для бана Женья взял 0.5. Точность классификатора (precision) получилась 0.9. Полнота (recall) получилась 0.7. Как выглядит матрица ошибок (confusion matrix)?

<sup>2</sup>[https://en.wikipedia.org/wiki/Mann-Whitney\\_U\\_test](https://en.wikipedia.org/wiki/Mann-Whitney_U_test)

- 
2. На самом деле в чате токсик встречается только в 20% случаев. Женя натравил классификатор на 1000 свежих сообщений. Сколько из них классификатор назовёт токсичными? Сколько раз он ошибётся в этом? Как примерно будет выглядеть матрица ошибок?
  3. Айнура требует от Жени метрики качества его классификатора. Женя размстил 100 случайных сообщений, которые классификатор забанил и 100 случайных сообщений, которые классификатор не стал банить. По этой выборке он оценил precision, recall, accuracy и FPR. Он утверждает, что эти метрики отражают качество работы классификатора на потоке свежих сообщений.

Айнура не согласна с Женей и считает, что его оценки смещены из-за того, что они посчитаны на сбалансированной выборке, а токсика всего лишь 20%. Правда ли это?
  4. Чтобы все метрики качества соответствовали потоку, можно было бы сделать случайную выборку из него. Однако Айнуре лень. Помогите ей вывести формулы, которые скорректируют оценки Жени.