Машинное обучение, ФКН ВШЭ
Семинар №5

AUC-ROC

На лекции мы познакомились с такой важной метрикой качества бинарной классификации, как площадь под ROC-кривой (AUC-ROC). Напомним её определение. Рассмотрим задачу бинарной классификации с метками классов $\mathbb{Y} = \{-1, +1\}$, и пусть задан некоторый алгоритм $b(x)$, позволяющий вычислять оценку принадлежности объекта x положительному классу. AUC-ROC позволяет оценивать качество классификации для семейства алгоритмов следующего вида:

$$a(x;t) = \begin{cases} -1, b(x) \leqslant t \\ +1, b(x) > t \end{cases}$$

т.е. алгоритмов, присваивающих метки объектам в соответствии с оценками $b(x)$, отсекая их по некоторому порогу t. Каждый алгоритм (получающийся при фиксации значения порога t) представляется точкой на плоскости (FPR, TPR), где

$$\begin{aligned} \mathrm{FPR} &= \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} = \frac{\mathrm{FP}}{\ell_-} \\ \mathrm{TPR} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = \frac{\mathrm{TP}}{\ell_+} \end{aligned}$$

ℓ_-, ℓ_+- количество объектов отрицательного и положительного классов соответственно. AUC-ROC, в свою очередь, является площадью под получившейся кривой.

Изучим подробнее некоторые важные свойства данной метрики.

Критерий AUC-ROC имеет большое число интерпретаций - например, он равен вероятности того, что случайно выбранный положительный объект окажется позже случайно выбранного отрицательного объекта в ранжированном списке, порожденном $b(x)$. Разберем подробнее немного другую формулировку.

Задача 1.1. В ранжировании часто используется функционал «доля дефектных пар». Его можно определить и для задачи бинарной классификации.

Пусть дан классификатор $b(x)$, который возвращает оценки принадлежности объектов классу $+1$, и пусть все значения $b\left(x_i\right), i = \overline{1, \ell}$, для некоторой выборки $X = \{(x_i, y_i)\}_{i=1}^{\ell}$ различны. Отсортируем все объекты по возрастанию ответа классификатора: $b\left(x_{(1)}\right) < \cdots < b\left(x_{(\ell)}\right)$. Обозначим истинные ответы на этих объектах

через $y_{(1)}, \ldots, y_{(\ell)}$. Тогда доля дефектных пар записывается как

$$DP(b, X) = \frac{2}{\ell(\ell-1)} \sum_{i<j}^{\ell} \left[y_{(i)} > y_{(j)}\right].$$

Как данный функционал связан с AUC-ROC?

Решение. Для начала разберем процедуру построения ROC-кривой. Сперва все объекты сортируются по неубыванию оценки $b(x)$, тем самым формируя список $x_{(1)}, \dots, x_{(\ell)}$. Заметим, что для построения ROC-кривой достаточно рассмотреть $(\ell + 1)$ различных значений порога t, соответствующих всем различным способам классификации выборки, порожденным алгоритмом $b(x)$, — например, в качестве таких порогов можно рассмотреть следующий набор:

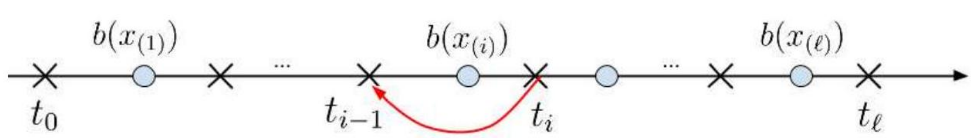
$$t_\ell = b(x_{(\ell)}) + 1,$$
$$t_i = \frac{b(x_{(i)}) + b(x_{(i+1)})}{2},\ i = \overline{1, \ell-1},$$
$$t_0 = b(x_{(1)}) - 1.$$

Зафиксируем значение порога $t = t_\ell = b(x_{(\ell)}) + 1$, в этом случае имеем

$$\text{FPR} = \frac{\text{FP}}{\ell_-} = \frac{0}{\ell_-} = 0,$$

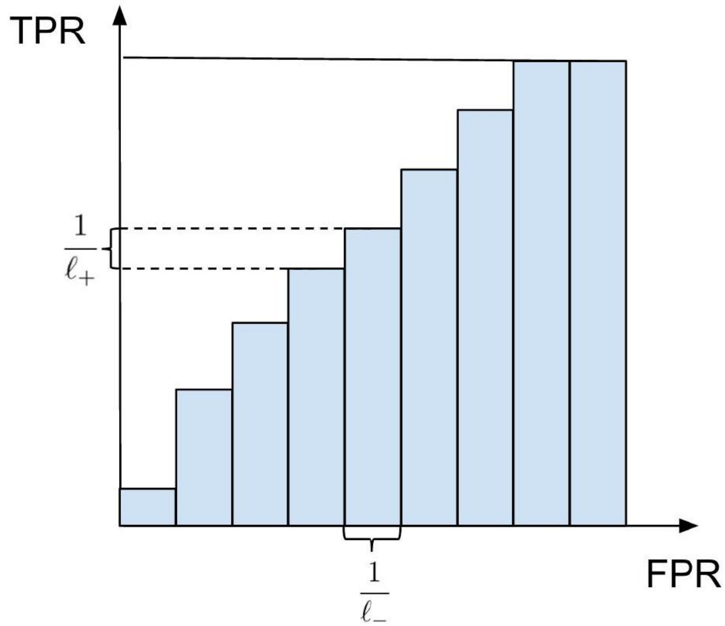
$$\text{TPR} = \frac{\text{TP}}{\ell_+} = \frac{0}{\ell_+} = 0.$$

Таким образом, алгоритму $a(x; t_\ell)$ соответствует точка $(0; 0)$ на плоскости, откуда начинается построение ROC-кривой. Будем перебирать пороги в порядке невозрастания их значения, начиная с t_ℓ. Пусть мы хотим уменьшить значение порога с t_i до t_{i-1}. При этом классификация объекта $x_{(i)}$ (и только его) изменится с отрицательной на положительную. Рассмотрим 2 случая.



1. $y_{(i)} = +1$. В этом случае классификатор начнет верно классифицировать объект, на котором ранее допускал ошибку, при этом FPR не изменится, а TPR повысится на $\frac{1}{\ell_+}$.

2. $y_{(i)} = -1$. В этом случае классификатор начнет ошибаться на объекте, который ранее классифицировал верно, при этом TPR не изменится, а FPR повысится на $\frac{1}{\ell_-}$.



Теперь рассмотрим, как при этом изменяется AUC-ROC. Заметим, что область под ROC-кривой состоит из непересекающихся прямоугольников, каждый из которых снизу ограничен осью FPR, а сверху — одним из горизонтальных отрезков, соответствующих второму из рассмотренных случаев. Поэтому каждый раз, когда имеет место второй случай, к текущей накопленной площади под кривой (которая изначально в точке (0; 0) равна 0) добавляется площадь прямоугольника, горизонтальные стороны которого равны $\frac{1}{\ell_-}$, а вертикальные равны $\frac{1}{\ell_+}\sum_{j=i+1}^{\ell}[y_{(j)} = +1]$ (доля уже рассмотренных положительных объектов среди всех положительных), поэтому в этом случае текущее значение AUC-ROC увеличивается на $\frac{1}{\ell_-\ell_+}\sum_{j=i+1}^{\ell}[y_{(j)} = +1]$. Итого, финальное значение AUC-ROC можно посчитать следующим образом:

$$
\begin{aligned}
\text{AUC} &= \frac{1}{\ell_+\ell_-}\sum_{i=1}^{\ell}[y_{(i)} = -1]\sum_{j=i+1}^{\ell}[y_{(j)} = +1] = \\
&= \frac{1}{\ell_+\ell_-}\sum_{i=1}^{\ell}\sum_{j=i+1}^{\ell}[y_{(i)} < y_{(j)}] = \\
&= \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}(1 - [y_{(i)} = y_{(j)}] - [y_{(i)} > y_{(j)}]) = \\
&= \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}(1 - [y_{(i)} = y_{(j)}]) - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)} > y_{(j)}] = \\
&= \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}([y_{(i)} \neq y_{(j)}]) - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)} > y_{(j)}] = \\
&= \frac{\ell_+\ell_-}{\ell_+\ell_-} - \frac{1}{\ell_+\ell_-}\sum_{i<j}^{\ell}[y_{(i)} > y_{(j)}] = 1 - \frac{1}{\ell_+\ell_-}\sum_{i<j}[y_{(i)} > y_{(j)}].
\end{aligned}
$$

Отсюда получаем, что AUC-ROC и доля дефектных пар связаны следующим соотношением:

$$\mathrm{DP}(b, X) = \frac{2\ell_-\ell_+}{\ell(\ell-1)}(1 - \mathrm{AUC}(b, X))).$$

Заметим, что в случае, когда несколько объектов выборки имеют равные значения $b(x)$, при уменьшении значения порога с $t_i > b(x)$ до $t_{i-1} < b(x)$, где x — один из таких объектов, изменение значений FPR и TPR происходит одновременно, поэтому соответствующий участок ROC-кривой будет наклонным, а не горизонтальным или вертикальным.

Задача 1.2. Пусть даны выборка X, состоящая из 5 объектов, и классификатор $b(x)$, предсказывающий оценку принадлежности объекта положительному классу. Предсказания $b(x)$ и реальные метки объектов приведены ниже:

$$\begin{aligned}
b\left(x_1\right) &= 0.2, \quad & y_1 &= -1,\\
b\left(x_2\right) &= 0.4, \quad & y_2 &= +1,\\
b\left(x_3\right) &= 0.1, \quad & y_3 &= -1,\\
b\left(x_4\right) &= 0.7, \quad & y_4 &= +1,\\
b\left(x_5\right) &= 0.05, \quad & y_5 &= +1.
\end{aligned}$$

Вычислите AUC - ROC и $PR - ROC$ для b(x) на выборке X.

Решение. В соответствии с процессом построения ROC-кривой, описанным в предыдущей задаче, отсортируем оценки $b(x_i)$ в порядке их неубывания: $(b(x_{(i)}))_{i=1}^{\ell} = (0.05, 0.1, 0.2, 0.4, 0.7)$. Также составим последовательность реальных меток объектов из этого упорядоченного списка: $(y_{(i)})_{i=1}^{\ell} = (+1, -1, -1, +1, +1)$.

Построим ROC-кривую (см. рис. 1), откуда AUC-ROC $= \frac{2}{3}$.

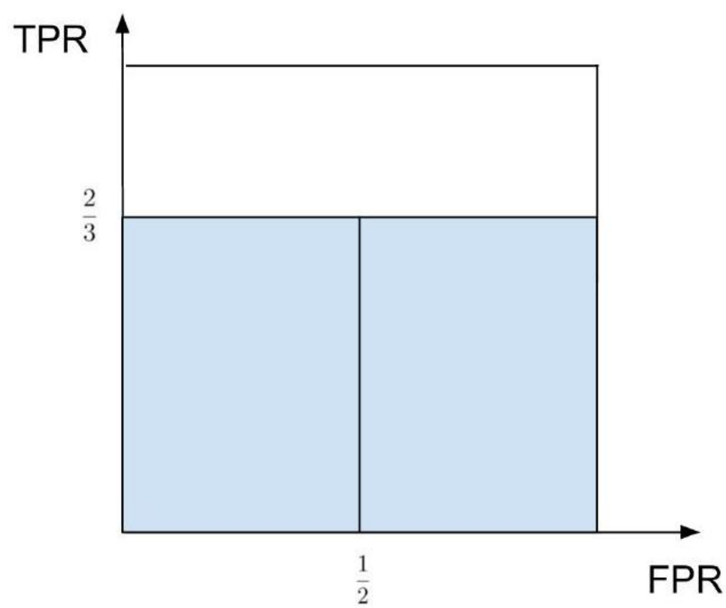


Рис. 1. Иллюстрация к задаче 1.2.

Заметим, что при вычислении AUC-ROC на некоторой выборке X для итогового классификатора $a(x; t)$ важны не конкретные значения $b(x_i)$, $i = \overline{1, \ell}$, а порядок расположения объектов в отсортированном по неубыванию списке $b(x_{(1)}), \ldots, b(x_{(\ell)})$, порожденным алгоритмом $b(x)$. Таким образом, для фиксированной выборки X алгоритм $b(x)$ задаёт перестановку на её объектах, которая в дальнейшем используется при расчёте AUC-ROC.

Задача 1.3. Пусть $b(x)$ - некоторый классификатор, предсказывающий оценку принадлежности объекта x положительному классу, и при этом AUC-ROC множества классификаторов $a(x;t)$, порожденных $b(x)$, на некоторой выборке X принимает значение, меньшее 0.5. Как можно скорректировать прогнозы классификаторов $a(x;t)$, чтобы они были более осмысленными по сравнению с прогнозами классификатора, выдающего случайные ответы?

Задача 1.4. На ответах алгоритма b(x), отнормированных на интервал от 0 до 1, объекты отрицательного класса распределены с плотностью p(b) = 2 2b, а объекты положительного класса распределены с плотностью p(b) = 2b. Выпишите формулу для ROC-кривой и посчитайте площадь под ней.

Задача 1.5. У банка всего 4000 клиентов. Маркетингового бюджета нового пред- ложения банка хватит на то, чтобы обзвонить 800 клиентов. По историческим дан- ным аналитики банка выяснили, что лишь 6 % клиентов действительно начинают пользоваться новым предложением после маркетингового звонка. У компании уже есть два классификатора A и B, для которых положительный класс – это клиенты, которые отреагируют на маркетинговый звонок, а отрицательный – клиенты, на которых он не повлияет. Известно, что для A FPR = 0.1, TPR = 0.2, а для B FPR = 0.25, TPR = 0.6. Постройте на их основе классификатор, который выберет ровно 800 клиентов для совершения маркетинговых звонков.

Задача 1.6. Зафиксируем число объектов положительного l_+ и отрицательного l_-классов. Докажите, что ROC-кривая классификатора A не ниже ROC-кривой классификатора B в любой точке тогда и только тогда, когда PR-кривая классификатора A не ниже PR-кривой классификатора B в любой точке.

Задача 1.7. Обратимая ли PR-кривая?

Рассмотрим бинарный классификатор, который каждой выборке из N объектов сопоставляет оценки $b_i \in [0,1]$, $i = 1, \ldots, N$. Пусть истинные метки обозначены как $y_i \in \{0,1\}$. Так же дополнительно известна π - доля положительных примеров в выборке.

Пусть PR-кривая классификатора задана в аналитическом виде, то есть известна точная зависимость $Precision$ =f($Recall$).

1. Докажите, что зная функцию $f(Recall)$ и значение π, можно восстановить ROC-кривую (то есть получить зависимость TPR от FPR).

2. Найдите явное выражение для FPR через $Precision$, $Recall$ и π.

3. Покажите, что обратное преобразование (ROC $\to$ PR) невозможно без знания π, приведя контрпример (две выборки с разными π, но одинаковой ROC-кривой и различной PR-кривой).

Указание. Используя соотношения:

$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN}, \quad \pi = \frac{TP+FN}{N}.$$

Можно выразить через эти параметры FP, FN и получить

$$FPR = \frac{FP}{FP + TN},$$

учитывая, что $FP + TN = (1 - \pi)N$.

Ответ.

$$FPR = \frac{\pi\, Recall\,(1 - Precision)}{(1-\pi)\, Precision}, \qquad TPR = Recall.$$

Комментарий. При известной PR-кривой $Precision(Recall)$ и доле положительных π, ROC-кривая однозначно восстанавливается. Однако без знания π обратное преобразование невозможно, так как ROC-кривая инвариантна к изменению соотношения классов, а PR-кривая — нет.

Прямая оптимизация AUC-ROC

При обучении модели в бинарной классификации чаще всего решается задача минимизации верхней оценки функционала ошибки:

$$Q(a, X) = \frac{1}{\ell}\sum_{i=1}^{\ell}\left[a\left(x_i\right) \neq y_i\right] \leqslant \frac{1}{\ell}\sum_{i=1}^{\ell}\tilde{L}\left(M_i\right) \to \min_w$$

Однако иногда возникает необходимость оптимизировать более сложные метрики - в частности, AUC-ROC. Напрямую оптимизировать подобные метрики не представляется возможным из-за их дискретной структуры, однако мы можем использовать трюк с верхней оценкой функционала ошибки и в этом случае. В задаче 1.1 мы показали, что AUC-ROC связан с долей дефектных пар в выборке, поэтому максимизация AUC-ROC равносильна минимизации доли дефектных пар.

$$DP(b, X) = \frac{2}{\ell(\ell-1)}\sum_{i<j}^{\ell}\left[y_i < y_j\right]\left[b\left(x_i\right) > b\left(x_j\right)\right] =$$

$$\frac{2}{\ell(\ell-1)}\sum_{i<j}^{\ell}\left[y_i < y_j\right]\left[b\left(x_j\right) - b\left(x_i\right) < 0\right] \leqslant \frac{2}{\ell(\ell-1)}\sum_{i<j}^{\ell}\left[y_i < y_j\right]\tilde{L}\left(b\left(x_j\right) - b\left(x_i\right)\right) \to \min_b$$

Если верхняя оценка $\tilde{L}$ дифференцируема по параметрам модели, то можно оптимизировать такой функционал при помощи градиентных методов.