

Лекция 4

Оценки качества в задачах регрессии и классификации

Макаренко В.А., Габдуллин Р.А.

МГУ им. М.В. Ломоносова

31 января 2023

- ℓ – размер обучающей выборки.
- y_i – ответ на i -м объекте.
- \hat{y}_i – предсказание на i -м объекте.
- $Y = \{y_1, y_2, \dots, y_\ell\}$.
- $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_\ell\}$.

Метрики в задаче классификации

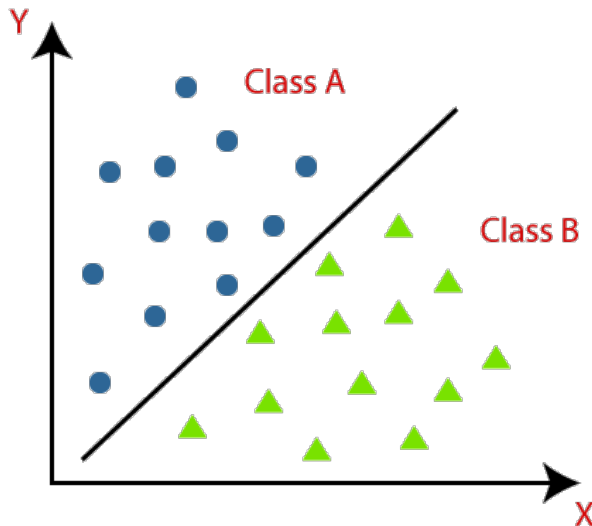


Рис.: Источник: javatpoint.com

Confusion matrix

- True Positive:

$$\text{TP}(Y, \hat{Y}) = \sum_{i=1}^{\ell} [y_i = 1, \hat{y}_i = 1].$$

- False Positive:

$$\text{FP}(Y, \hat{Y}) = \sum_{i=1}^{\ell} [y_i = 0, \hat{y}_i = 1].$$

- False Negative:

$$\text{FN}(Y, \hat{Y}) = \sum_{i=1}^{\ell} [y_i = 1, \hat{y}_i = 0].$$

- True Negative:

$$\text{TN}(Y, \hat{Y}) = \sum_{i=1}^{\ell} [y_i = 0, \hat{y}_i = 0].$$

Confusion matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Рис.: Источник: towardsdatascience.com

Confusion matrix

	actual	predicted
0	0	0
1	1	1
2	1	0
3	0	0
4	1	0
5	0	1
6	0	0
7	0	0
8	1	1
9	1	0
10	1	1

Чему равны TP, FP, FN, TN?

Доля верных ответов (accuracy)

Доля верных ответов (accuracy):

$$\text{Accuracy}(Y, \hat{Y}) = \frac{1}{\ell} \cdot \sum_{i=1}^{\ell} [y_i = \hat{y}_i] = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}.$$

- Метрика применима и для множественной классификации.
- Метрика плохо интерпретируема в случае дисбаланса классов.

Чувствительность и специфичность

Чувствительность (True Positive Rate):

$$\text{TPR}(Y, \hat{Y}) = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

Специфичность (True Negative Rate):

$$\text{TNR}(Y, \hat{Y}) = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Точность (precision) и полнота (recall)

Точность классификации (положительного класса):

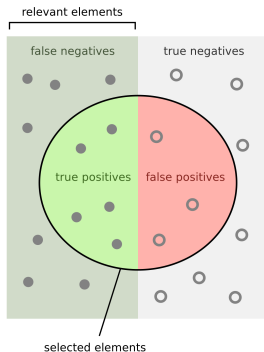
$$\text{Precision}(Y, \hat{Y}) = \frac{TP}{TP + FP}.$$

Полнота классификации (положительного класса):

$$\text{Recall}(Y, \hat{Y}) = \frac{TP}{TP + FN}.$$

- Неинформативны по отдельности (как максимизировать Precision? Recall?)
- Легко обобщается для большего числа классов.
- Часто фиксируют желаемую точность и максимизируют полноту.

Точность (precision) и полнота (recall)



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Рис.: Источник: [Википедия](#)

F_1 -мера (F_1 -score)

Среднее гармоническое Precision и Recall:

$$F_1(Y, \hat{Y}) = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

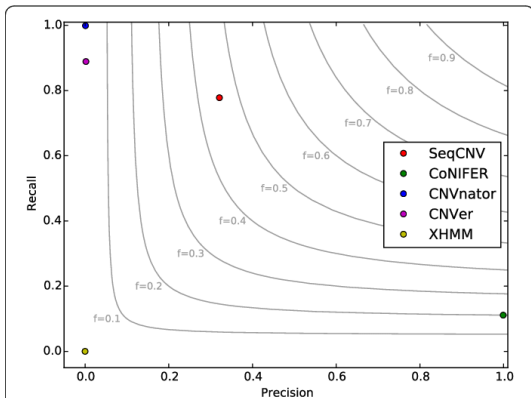


Рис.: Источник: [researchgate.net](https://www.researchgate.net)

Accuracy, precision, recall, F_1

actual		0	1
predicted	0	4	3
	1	1	3

Вычислить accuracy, precision, recall, F_1 .

ROC-кривая

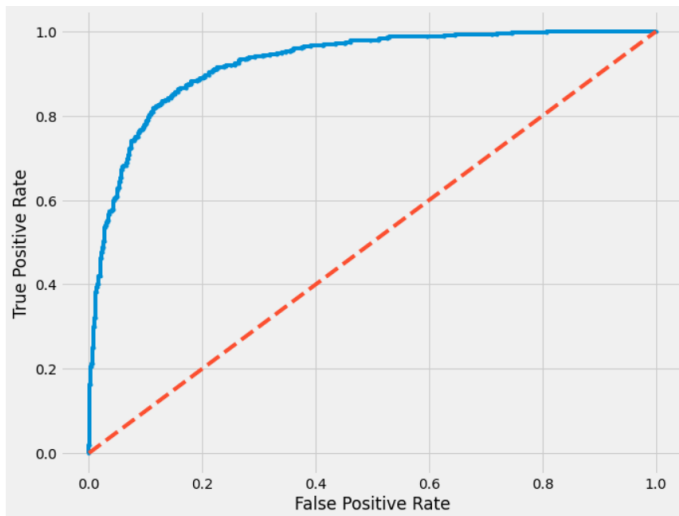


Рис.: Источник: medium.com

Пусть в выборке m объектов положительного класса и n объектов отрицательного класса. Построение ROC-кривой:

- Разбиваем единичный квадрат вертикальными и горизонтальными прямыми на m вертикальных и n горизонтальных блоков.
- Упорядочиваем объекты по убыванию «уверенности» позитивного класса.
- Начинаем рисовать кривую с точки $(0,0)$.
- Если встречаем объект позитивного класса, то делаем шаг вверх.
- Если встречаем объект негативного класса, то делаем шаг вправо.
- Если несколько объектов имеют одну и ту же «уверенность», то делаем a шагов вверх и b шагов вправо, где a и b – соответственно количество объектов положительного и отрицательного классов.
- В итоге оказываемся в точке $(1, 1)$.

ROC-кривая

id	оценка	класс
4	0.6	1
1	0.5	0
6	0.3	1
3	0.2	0
5	0.2	1
2	0.1	0
7	0.0	0

Рис.: Источник: dyakonov.org

Как будет выглядеть ROC-кривая? Чему равна площадь под ней?

ROC AUC – площадь под ROC-кривой.

- Количество блоков: mn .
- Количество пар <объект положительного класса, объект отрицательного класса>: mn .
- Если блок под ROC-кривой, то объект положительного класса оказался выше (по уверенности) объекта отрицательного класса.
- Если блок над ROC-кривой, то объект отрицательного класса оказался выше (по уверенности) объекта положительного класса.
- ROC AUC – доля пар <объект положительного класса, объект отрицательного класса>, которые алгоритм упорядочил правильно.

Метрики в задаче регрессии

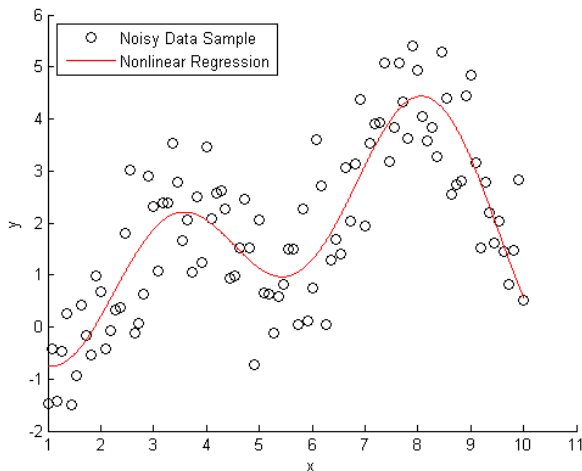


Рис.: Источник: datascience.stackexchange.com

Среднеквадратическая ошибка (MSE)

Среднеквадратическая ошибка (mean square error):

$$\text{MSE}(Y, \hat{Y}) = \frac{1}{\ell} \cdot \sum_{i=1}^{\ell} (y_i - \hat{y}_i)^2.$$

Корень из среднеквадратической ошибки (root mean square error):

$$\text{RMSE}(Y, \hat{Y}) = \sqrt{\text{MSE}(Y, \hat{Y})}.$$

- Возникает в методе наименьших квадратов (ошибки нормально распределены)
- Метрика неустойчива к выбросам.

Средняя абсолютная ошибка (MAE)

Средняя абсолютная ошибка (mean absolute error):

$$\text{MAE}(Y, \hat{Y}) = \frac{1}{\ell} \cdot \sum_{i=1}^{\ell} |y_i - \hat{y}_i|.$$

- Возникает в методе наименьших модулей (ошибки имеют распределение Лапласа).
- Метрика устойчива к выбросам.

Средняя абсолютная процентная ошибка (MAPE)

Средняя абсолютная процентная ошибка (mean absolute percent error):

$$\text{MAPE}(Y, \hat{Y}) = \frac{1}{\ell} \cdot \sum_{i=1}^{\ell} \left| \frac{y_i - \hat{y}_i}{y_i} \right|.$$

Эмпирический риск в задаче квантильной регрессии:

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \rho_{\tau}(y_i - \hat{y}_i),$$

где

$$\rho_{\tau}(u) = \begin{cases} \tau u, & u \geq 0, \\ (\tau - 1)u, & u < 0. \end{cases}$$

- По-разному штрафует за «недопрогноз» и «перепрогноз».

Коэффициент детерминации (R^2)

Коэффициент R^2 :

$$R^2(Y, \hat{Y}) = 1 - \frac{\sum_{i=1}^{\ell} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{\ell} (y_i - \bar{y})^2},$$

где $\bar{y} = \frac{1}{\ell} \cdot \sum_{i=1}^{\ell} y_i$.

- Показывает долю объясненной дисперсии.

- Метрики в задаче классификации
 - Confusion matrix
 - Accuracy
 - Precision, Recall
 - F-мера
 - ROC-кривая
 - ROC AUC
- Метрики в задаче регрессии
 - MSE
 - MAE
 - MAPE
 - Квантильная метрика
 - R^2