



Занятие 5

Метрики качества классификации

Елена Кантонистова

elena.kantonistova@yandex.ru

ВШЭ, 2021

МЕТРИКИ КАЧЕСТВА КЛАССИФИКАЦИИ

- Accuracy – доля правильных ответов:

$$accuracy(a, X) = \frac{1}{n} \sum_{i=1}^n [a(x_i) = y_i]$$

ACCURACY И ДИСБАЛАНС КЛАССОВ

- Пример:
 - Класс -1: 950 объектов
 - Класс +1: 50 объектов
- $a(x) = -1$
- Доля правильных ответов: 0.95
- Если получили большой accuracy — посмотрите на баланс классов

МЕТРИКИ КАЧЕСТВА КЛАССИФИКАЦИИ

- Accuracy – доля правильных ответов:

$$\text{accuracy}(a, X) = \frac{1}{n} \sum_{i=1}^n [a(x_i) = y_i]$$

Недостатки:

- *при сильно несбалансированной выборке не отражает качество работы алгоритма*
- *не учитывает тип ошибки*

ПРИМЕР: КРЕДИТНЫЙ СКОРИНГ

- **Пример:** Кредитный скоринг
- **Модель 1: одобряет 100 кредитов**
 - 80 кредитов вернули
 - 20 кредитов не вернули
- **Модель 2: одобряет 50 кредитов**
 - 48 кредитов вернули
 - 2 кредита не вернули
- **Какая лучше?**

На выборке,
где 100 вернули,
100 не вернули

МАТРИЦА ОШИБОК

Матрица ошибок (confusion matrix):

		Actual Value	
		positives	negatives
Predicted Value	positives	TP True Positive	FP False Positive
	negatives	FN False Negative	TN True Negative

МЕТРИКИ КАЧЕСТВА КЛАССИФИКАЦИИ: PRECISION, RECALL

- **Precision (точность):**

$$\textit{Precision}(a, X) = \frac{TP}{TP + FP}$$

Показывает, насколько можно доверять классификатору при $a(x) = +1$.

PRECISION: ПРИМЕР

Модель $a_1(x)$:

$$\text{precision}(a_1, X) = 0.8$$

	$y = 1$ Могут вернуть	$y = -1$ Не могут вернуть
$a(x) = 1$ Получили кредит	80	20
$a(x) = -1$ Не получили кредит	20	80

Модель $a_2(x)$:

$$\text{precision}(a_2, X) = 0.96$$

	$y = 1$ Могут вернуть	$y = -1$ Не могут вернуть
$a(x) = 1$ Получили кредит	48	2
$a(x) = -1$ Не получили кредит	52	98

МЕТРИКИ КАЧЕСТВА КЛАССИФИКАЦИИ: PRECISION, RECALL

- Precision (точность):

$$\textit{Precision}(a, X) = \frac{TP}{TP + FP}$$

Показывает, насколько можно доверять классификатору при $a(x) = +1$.

- Recall (полнота):

$$\textit{Recall}(a, X) = \frac{TP}{TP + FN}$$

Показывает, как много объектов положительного класса находит классификатор.

RECALL: ПРИМЕР

Модель $a_1(x)$:

$$\text{recall}(a_1, X) = 0.8$$

	$y = 1$ Могут вернуть	$y = -1$ Не могут вернуть
$a(x) = 1$ Получили кредит	80	20
$a(x) = -1$ Не получили кредит	20	80

Модель $a_2(x)$:

$$\text{recall}(a_2, X) = 0.48$$

	$y = 1$ Могут вернуть	$y = -1$ Не могут вернуть
$a(x) = 1$ Получили кредит	48	2
$a(x) = -1$ Не получили кредит	52	98

F-МЕРА

F-мера – это метрика качества, учитывающая и точность, и полноту

$$F(a, X) = \frac{2 \cdot \textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}}$$

