

Метрики качества классификации

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ

Вопросы занятия

1. Метрики качества моделей

В конце занятия научимся:

- *измерять качество* решений в задачах классификации.

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

МАТРИЦА ОШИБОК

confusion matrix	$y = 1$	$y = 0$
$a = 1$	True Positive	False Positive
$a = 0$	False Negative	True Negative

На тестовой выборке имеем:

- * y - вектор истинных значений
- * a - вектор предсказаний классификатора

Будем раскладывать все пары (предсказание, истина) по ячейкам матрицы ошибок

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

ACCURACY

confusion matrix	$y = 1$	$y = 0$
$a = 1$	0	0
$a = 0$	2	998

Ассурасу, Доля верных ответов
(в просторечии точность, но не путать с точностью из ML!)

Простая метрика, но абсолютно непоказательна в задачах с несбалансированными классами

Пример:

определение качества теста на рак. Тест постоянно предсказывает отсутствие рака.
Доля верных ответов: 99.8%

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

PRECISION

confusion matrix	$y = 1$	$y = 0$
$a = 1$	170	100
$a = 0$	30	700

Precision, Точность

отсутствие ложных срабатываний

Пример:

правильное распознавание намерения пользователя: лучше переспросить пользователя, чем сделать не то, что нужно

$$\text{Точность} = 170 / (170 + 100) = 0.629$$

$$\text{precision} = \frac{TP}{TP + FP}$$

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

RECALL

confusion matrix	$y = 1$	$y = 0$
$a = 1$	170	100
$a = 0$	30	700

Recall, Полнота

отсутствие ложных пропусков

Пример:

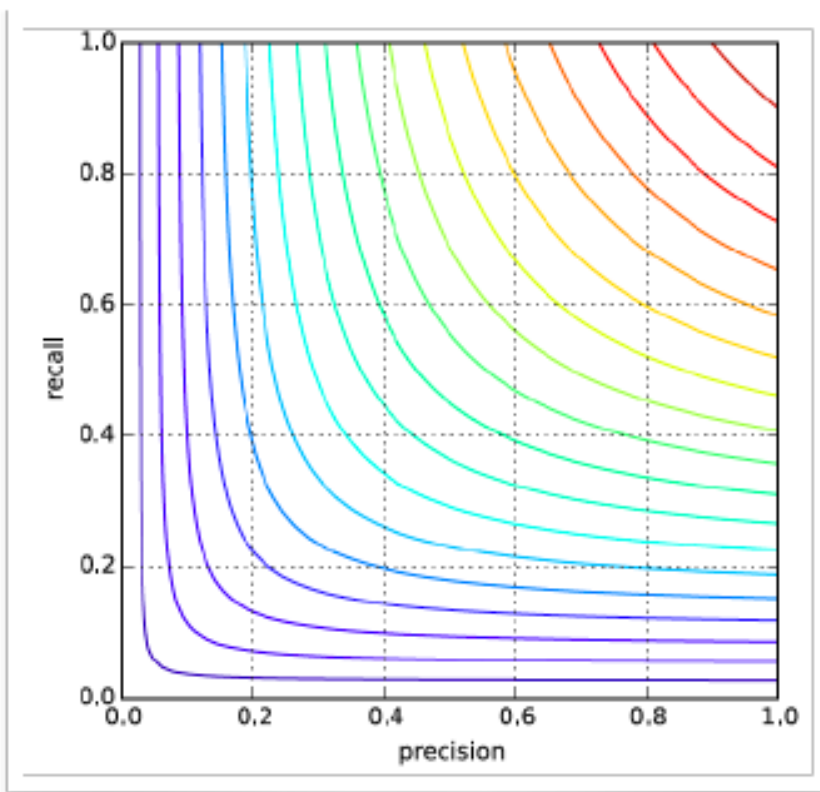
определение мошеннических действий в банке:
лучше найти лишнее и проверить, чем не найти

Точность = $170 / (170 + 30) = 0.85$

$$recall = \frac{TP}{TP + FN}$$

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

F1 - МЕРА



F1-мера

комбинация точности и полноты в одну метрику

Пример:

правильное распознавание намерения пользователя. Насколько мы уверены в том, что правильно поняли? Надо ли уточнить?

$$F1 = 2 * 0.629 * 0.85 / (0.629 + 0.85) = 0.723$$

$$F = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

Классификаторы обычно работают в 2 этапа:

- оценка вероятности принадлежности к классам $a(x)$;
- выбор порога отсечения, при котором идёт распределение в тот или иной класс.

Это 2 отдельные задачи, после получения оценки вероятности можно отсортировать объекты и в различные периоды времени использовать разные пороги.

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ

Задача: $a(x) \in \{1, \dots, K\}$

сводится к K задачам отделения класса N от остальных

Как усреднить качество K задач?

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ

macro-averaging:

- вычислим итоговую метрику для каждой задачи;
- усредним по задачам.

все классы делают равный вклад

weighted-averaging:

- вычислим confusion matrix для каждой задачи;
- усредним по задачам;
- вычислим итоговую метрику.

классы делают вклад, пропорциональный размеру

ОЦЕНКА КАЧЕСТВА КЛАССИФИКАЦИИ

РЕАЛИЗАЦИЯ В SKLEARN

sklearn.metrics

Схожий интерфейс: функции от y , y_{pred}

<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics>

ЧТО МЫ СЕГОДНЯ УЗНАЛИ

Метрик качества много, они разные по смыслу,
для своих задач надо выбирать подходящую

Метрики качества классификации

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ