Постановка задачи

Дано: задача классификации.

$X^\ell = \{x_1, \ldots, x_\ell\}$ — выборка;

$y_i = y(x_i) \in \{0, 1\}, \;\; i = 1, \ldots, \ell$ — известные бинарные ответы.

$a\colon X \to Y$ — алгоритм, решающая функция, приближающая y на всём множестве объектов X.

Вопрос:
Как измерить качество $a(x)$ на выборке X^ℓ?

Accuracy

Доля правильных ответов на выборке (accuracy):

$$\frac{1}{\ell}\sum_{i=1}^{\ell}[a(x_i) = y_i]$$

- Соответствует интуитивным представлениям о качестве классификации
- Имеет проблемы с интерпретацией на несбалансированных выборках.

Пример (медицинская диагностика):

- 950 объектов класса 0,
- 50 объектов класса 1,
- $a(x) = 0$ для всех x.

Доля правильных ответов $a(x)$: 95%!

Решение: смотреть на базовую долю правильных ответов

$$\text{BaseRate} = \arg\max_{y_0 \in \{0,1\}} \frac{1}{\ell} \sum_{i=1}^{\ell} [y_0 = y_i]$$

В примере: BaseRate = 95%.

Матрица ошибок

Ошибки бывают разные:

	$y = 1$	$y = 0$
$a(x) = 1$	True Positive (TP)	False Positive (FP)
$a(x) = 0$	False Negative (FN)	True Negative (TN)

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

Матрица ошибок

Пример: задача медицинской диагностики ($y = 1$ — больные, $y = 0$ — здоровые).

	$y = 1$	$y = 0$
$a(x) = 1$	20	50
$a(x) = 0$	5	1000

Доля правильных ответов: 94.9%

	$y = 1$	$y = 0$
$a(x) = 1$	0	0
$a(x) = 0$	25	1050

Доля правильных ответов константного классификатора: 97.6%

У разных типов ошибки может быть разная *цена*.

Точность и полнота

Точность (precision) — насколько можно доверять классификатору:

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

	$y = 1$	$y = 0$
$a(x) = 1$	20	50
$a(x) = 0$	5	1000

Точность классификатора: 28.6%

Точность константного классификатора: 0%

Точность и полнота

Полнота (recall) — как много объектов класса 1 находит классификатор:

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

	$y = 1$	$y = 0$
$a(x) = 1$	20	50
$a(x) = 0$	5	1000

Полнота классификатора: 80%

Полнота константного классификатора: 0%

Точность и полнота

- Точность и полнота характеризуют разные стороны качества классификатора
- Чем выше точность, тем меньше ложных срабатываний
- Чем выше полнота, тем меньше ложных пропусков
- Приоритет в сторону точности или полноты выбирается в зависимости от задачи

Точность и полнота

Пример 1: определение мошеннических действий на банковских счетах.
Важнее **полнота**: лучше проверить лишний раз, чем пропустить вредоносные действия.

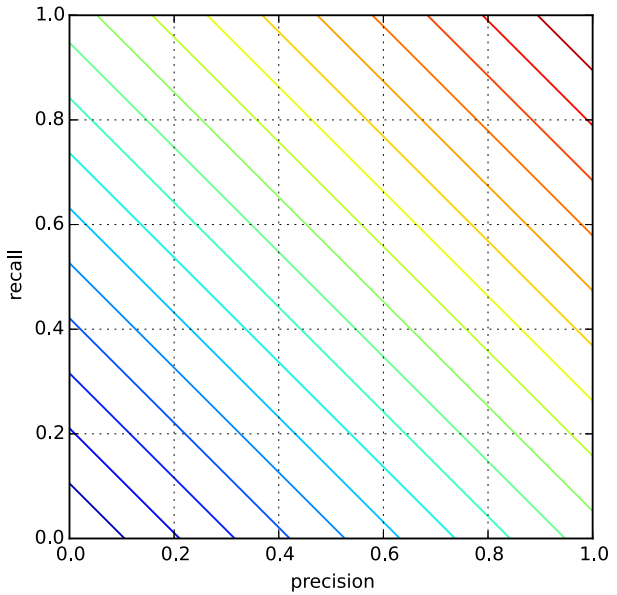
Пример 2: поиск вражеских самолетов для автоматического уничтожения ракетой
Важнее **точность**: нельзя допустить стрельбы по своему самолету.

Усреднение точности и полноты

Арифметическое среднее:

$$A = \frac{1}{2}(\text{precision} + \text{recall})$$

- Если precision $= 0.05$, recall $= 1$, то $A = 0.525$.
- Если precision $= 0.525$, recall $= 0.525$, то $A = 0.525$.
- Первый классификатор — константный, не имеет смысла.
- Второй классификатор показывает неплохое качество.



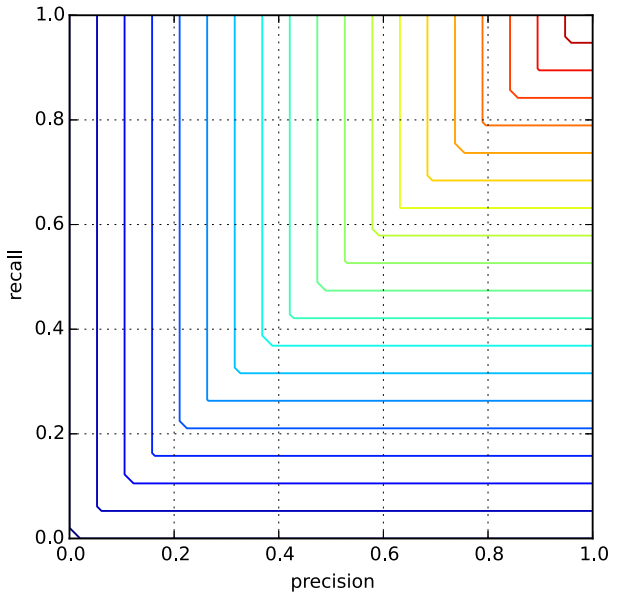
Усреднение точности и полноты

Минимум:

$$M = \min(\text{precision}, \text{recall})$$

- Если $\text{precision} = 0.05$, $\text{recall} = 1$, то $M = 0.05$.
- Если $\text{precision} = 0.525$, $\text{recall} = 0.525$, то $M = 0.525$.

- Если $\text{precision} = 0.2$, $\text{recall} = 1$, то $M = 0.2$.
- Если $\text{precision} = 0.2$, $\text{recall} = 0.3$, то $M = 0.2$.

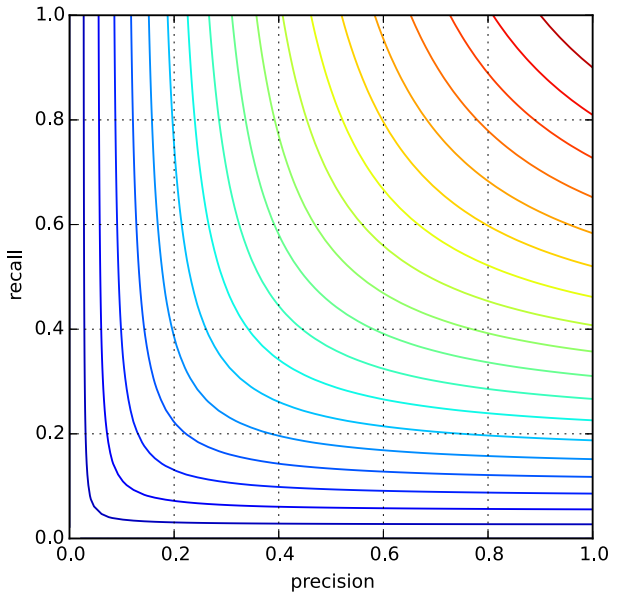


Усреднение точности и полноты

Гармоническое среднее, или F-мера:

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

- Если precision $= 0.05$, recall $= 1$, то $F = 0.1$.
- Если precision $= 0.525$, recall $= 0.525$, то $F = 0.525$.
- Если precision $= 0.2$, recall $= 1$, то $F = 0.33$.
- Если precision $= 0.2$, recall $= 0.3$, то $F = 0.24$.



Заключение

- Простая мера качества классификации — доля верных ответов
- Не учитывает цены ошибок
- Точность и полнота позволяют различать ложные срабатывания и ложные пропуски
- F-мера — способ усреднения точности и полноты