

Дано: $\{x_1, \dots, x_\ell\} \subset X$ — *выборка*;
 $y_i = y(x_i) \in \{1, \dots, K\}$, $i = 1, \dots, \ell$ — известные ответы.

Найти: $a: X \rightarrow Y$ — алгоритм, решающую функцию, приближающую y на всём множестве объектов X .

Вопросы:

- 1 Как свести задачу к бинарной классификации?
- 2 Как измерить качество решения?

Идея: построить K классификаторов, отделяющих каждый класс от остальных.

Получим K задач бинарной классификации:

- Объекты: $X^k = X^\ell$;
- Ответы: $y_i^k = [y_i = k]$;
- Оценка принадлежности: $b_k(x) \in \mathbb{R}$.

Итоговый алгоритм:

$$a(x) = \arg \max_{k=1, \dots, K} b_k(x).$$

Идея: построить классификаторы для каждой пары классов.

Получим $K(K - 1)$ задач бинарной классификации:

- Объекты: $X^{km} = \{x \in X^\ell \mid y(x) = k \text{ или } y(x) = m\}$;
- Ответы: $y_i^{km} = [y_i = k]$;
- Оценка принадлежности: $b_{km}(x) \in \mathbb{R}$;
- Симметрия: $b_{km}(x) = -b_{mk}(x)$.

Итоговый алгоритм:

$$a(x) = \arg \max_{k=1, \dots, K} \sum_{m=1}^K b_{km}(x).$$

One-vs-all:

- Линейное число классификаторов, но каждый обучается на полной выборке.
- Может возникнуть проблема с несбалансированными выборками.

All-vs-all:

- Квадратичное число классификаторов, но каждый обучается на небольшой подвыборке.

Доля правильных ответов (accuracy):

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

Матрица ошибок:

	$y = 1$	\dots	$y = K$
$a(x) = 1$	q_{11}	\dots	q_{1K}
\dots	\dots	\dots	\dots
$a(x) = K$	q_{K1}	\dots	q_{KK}

где

$$q_{ij} = \sum_{m=1}^{\ell} [a(x_m) = i][y_m = j].$$

Как обобщить точность, полноту, AUC?

Рассмотрим K задач отделения одного из классов от остальных.

- Микро-усреднение (micro-averaging):
 - Найдем TP, FP, FN, TN для каждой из задач;
 - Усредним их по всем задачам;
 - Вычислим итоговую метрику.

Вклад каждого класса зависит от его размера.

- Макро-усреднение (macro-averaging):
 - Вычислим итоговую метрику для каждой из задач;
 - Усредним по всем классам.

Все классы вносят равный вклад.

	TP	FP	FN	TN
$y = 1$	900	120	100	930
$y = 2$	850	70	150	980
$y = 3$	10	100	40	1900

Чему равна точность (precision)?

Микро-усреднение:

TP	FP	FN	TN
586.7	96.7	96.7	1270

Точность: 86%

Макро-усреднение:

Класс 1	Класс 2	Класс 3
88%	92%	9%

Точность: 63%

- Многоклассовую классификацию можно свести к серии бинарных задач
- Два подхода: one-vs-all и all-vs-all
- Вычисление качества также производится через сведение к бинарным задачам
- Микро-усреднение учитывает наиболее крупные классы
- Макро-усреднение учитывает все классы одинаково, без учета их размеров