Московский государственный университет имени М.В. Ломоносова
физический факультет
NeuroFuzzy Lab.

Машинное обучение как подбор оптимальных значений параметров метода распознавания образов на основе примеров

Зубюк Андрей Владимирович
zubjuk@physics.msu.ru

http://NeuroFuzzy.Phys.MSU.ru

МФК «Искусственный интеллект в естествознании»

Какой метод лучше? Проблема с подсчётом примеров

Фундаментальная проблема: при предложенном выше подходе к измерению качества алгоритма **качество будет разным для разных наборов примеров**!

Решение проблемы: **разные наборы примеров — это выборки из одного и того же (возможно, неизвестного) распределения вероятностей**. Тогда вероятности ошибок существуют «сами по себе» — без обучающих примеров — и неизменны. А доли ошибок, сосчитанные на основе обучающих примеров, — это случайные величины, примерно равные вероятностям при большом количестве примеров.

Аналогия с бросанием монеты. Вероятности выпадения орла и решки равны 1/2 (50%). Но если 100 раз бросить монету, то орёл может выпасть 50 раз ($50/100 = 1/2$), а может 49, 51, 48 и т. д. (строго говоря — любое количество раз от 0 до 100 включительно).
Т. е. вероятность выпадения орла — «фундаментальная» константа, а доля случаев, когда выпал орёл, — случайная величина, примерно равная 1/2 при большом количестве бросаний.

Стохастический эксперимент

Предметом изучения теории вероятностей является математическая модель стохастического эксперимента.

Определение

Стохастический эксперимент — принципиально недетерминированный эксперимент, который может быть многократно повторен при одних и тех же условиях, но исходы которого не могут быть предугаданы.

Примеры стохастических экспериментов:

- Бросание монеты.
- Бросание игральной кости.
- Перемешивание колоды и раздача карт.

Случайные события

Определение

Событие — любой верифицируемый исход стохастического эксперимента.

Примеры событий:

- При бросании монеты:
  $A_1 =$ «выпадение орла».
  $A_2 =$ «выпадение решки».
  $A_3 =$ «выпадение орла или решки», наступает при любом исходе эксперимента.
- Бросание игральной кости:
  $A_1 =$ «выпала 1».
  $A_2 =$ «выпало значения, превосходящее 3».
  $\ldots$

Вероятность

Пусть Э — некоторый стохастический эксперимент. Проведём его N раз и обозначим N_A число испытаний, в которых реализовалось событие A.

Определение

Вероятностью[a] $\mathrm{P}(A)$ события A будем называть величину

$$\mathrm{P}(A) \stackrel{\text{def}}{=} \lim_{N\to\infty} \frac{N_A}{N} \in [0, 1].$$

[a] В современной науке принято иное определение вероятности (А. Н. Колмогоров), однако в настоящем курсе мы для простоты используем «старое» частотное определение.

$\mathrm{P}(A)$ принято рассматривать как функцию, аргументом которой является событие A. Функция P является математической моделью стохастического эксперимента Э.

Вероятность — примеры

- Бросание монеты:
  — *Орёл и решка выпадают одинаково часто.*
  $\mathrm{P}(\text{«выпадение орла»}) = \mathrm{P}(\text{«выпадение решки»}) = 1/2.$
- Бросание игральной кости:
  — *Все грани выпадают одинаково часто.*
  $\mathrm{P}(\text{«выпала 1»}) = \mathrm{P}(\text{«выпала 2»}) = \ldots = 1/6,$
  $\mathrm{P}(\text{«выпало значения, превосходящее 3»}) = 3/6 = 1/2.$

Случайная величина

Пусть Э — некоторый стохастический эксперимент.

Определение

Величину $\xi \in (-\infty, \infty)$ будем называть случайной, если её значение однозначно[a] определяется исходом эксперимента Э.

[a] То есть если эксперимент Э проведён два раза, и все те события, которые реализовались (не реализовались) в первый раз, реализовались (не реализовались) также и во второй раз, то значения ξ в этих экспериментах должны совпадать.

Примеры:

- Бросание монеты: $\xi = 1$, если выпал орёл, и $\xi = 0$, если выпала решка.
- Бросание игральной кости: ξ — номер выпавшей грани.

Плотность распределения случайной величины

Далее будем рассматривать только абсолютно непрерывные случайные величины, то есть такие величины, для которых $\mathrm{P}(\text{«значение } \xi \text{ равно } a\text{»}) = 0$ для любого числа a.

Определение

Плотностью распределения случайной величины ξ назовём такую функцию p, что

$$\int\limits_a^b \mathrm{p}(x)\mathrm{d}x = \mathrm{P}(\text{«значение } \xi \text{ принадлежит интервалу } [a, b]\text{»})$$

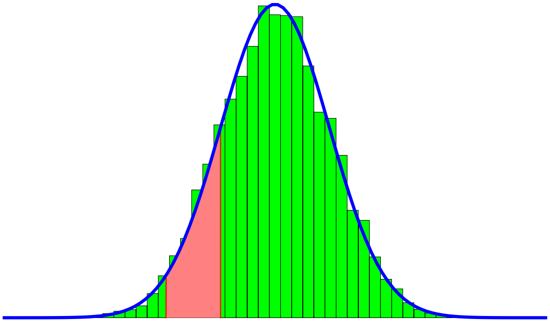
для любых чисел $a < b$.

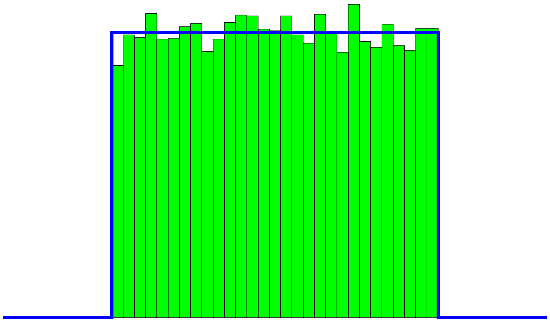
Графическая интерпретация плотности распределения

Из определений вероятности P и плотности распределения p следует, что для любого числа a и малого Δx

$$\mathrm{p}(a) \approx \frac{1}{\Delta x} \lim_{N\to\infty} N_A/N \ ,$$

где $A =$ «значение ξ попало в интервал $[a, a+\Delta x]$».





Математическое ожидание и дисперсия случайной величины

Математическое ожидание случайной величины ξ определим как

$$\mathbf{E}\xi \stackrel{\text{def}}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i = \int\limits_{-\infty}^{\infty} x\, \mathrm{p}(x)\, \mathrm{d}x.$$

Дисперсию ξ определим как

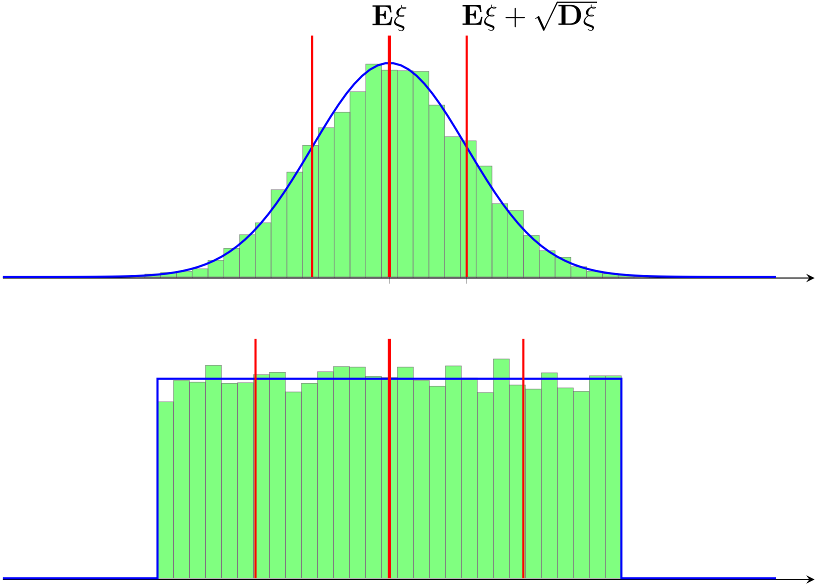
$$\mathbf{D}\xi \stackrel{\text{def}}{=} \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (x_i - \mathbf{E}\xi)^2 = \int\limits_{-\infty}^{\infty} (x - \mathbf{E}\xi)^2\, \mathrm{p}(x)\, \mathrm{d}x.$$

Здесь x_i — значение, которое приняла случайная величина ξ в i-ом испытании в серии из N испытаний.

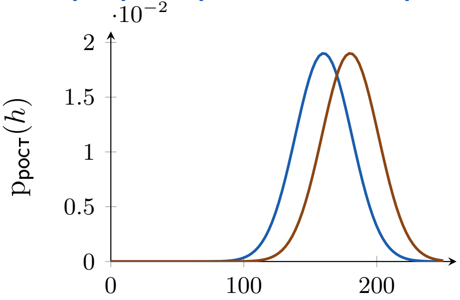
Принятая терминология

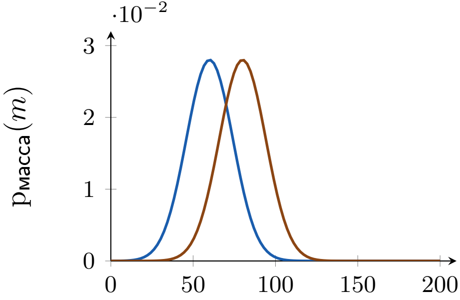
Математическое ожидание ξ иначе называют средним значением ξ.
Дисперсия ξ — это по определению средний квадрат отклонения ξ от своего среднего значения. Поэтому корень из дисперсии $\sqrt{\mathbf{D}\xi}$ принято называть среднеквадратичным отклонением ξ от среднего значения.

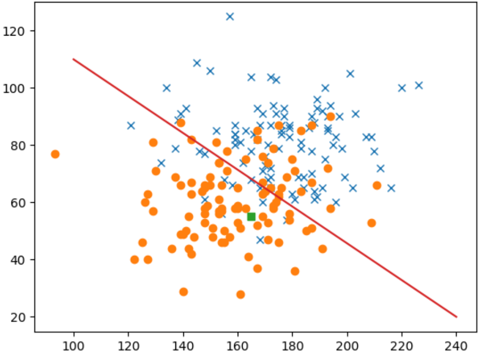
Графическая интерпретация $\mathrm{E}\xi$ и $\mathrm{D}\xi$



Набор примеров как выборка из распределения вероятностей

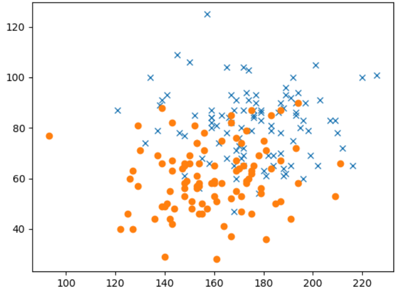


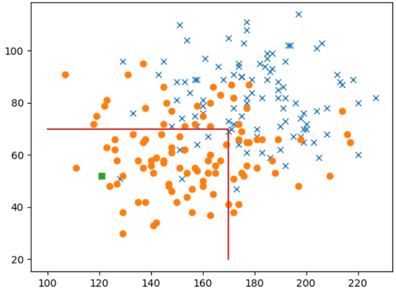




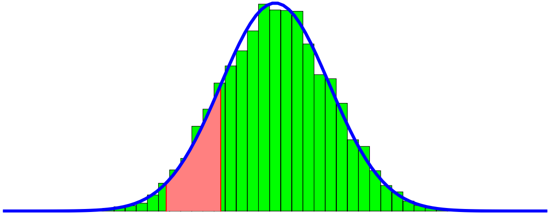
$$P(\text{«женщина»}) = 1/2,$$
$$P(\text{«мужчина»}) = 1/2.$$

Разные наборы примеров как выборки из одного распределения





Принцип устойчивости частот: гистограммы различаются, но близки к одной неизменной плотности распределения вероятностей.



Задача машинного обучения

Гиперплоскость: $(w, x) + b = 0$. Параметры гиперплоскости: $w,\ b$.

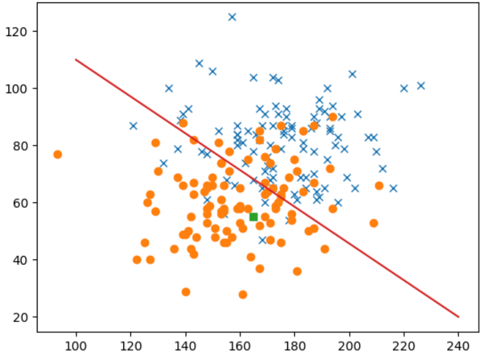
Задача **оптимального распознавания образов** — подобрать параметры $w,\ b$ так, чтобы минимизировать вероятность ошибки классификации:

$$\mathrm{P}(\text{«ошибка классификации»}) =$$

$$= \lim_{N\to\infty} \frac{N_{\text{«ошибка классификации»}}}{N} \sim \min_{w,b},$$

машинного обучения — частоту ошибок классификации (Вапник, Червоненкис):

$$\frac{N_{\text{«ошибка классификации»}}}{N} \sim \min_{w,b}$$



$w = \begin{pmatrix} 9 & 14 \end{pmatrix} \quad b = -2440$