Функция правдоподобия и ее роль

Сегодня

- Обсудим что такое функция правдоподобия
- Какую роль она играет в машинном обучении
- Выведем одну из основных функций потерь

Функция правдоподобия

Метод максимального правдоподобия - способ построения оценки неизвестного параметра. Он состоит в том, что в качестве «наиболее правдоподобного» значения параметра берут значение , максимизирующее вероятность получить при n опытах данную выборку $\vec{x} = (x_1, x_2 \ldots, x_n)$. Это значение параметра зависит от выборки. Саму эту вероятность оценивают с помощью **функции правдоподобия**.

- Сама функция правдободобия выглядит примерно так:

$$L(\vec{x}|\theta) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \ldots \cdot f_\theta(x_n)$$

- Часто можно видеть логарифм функции правдоподобия:

$$lnL(\vec{x}|\theta) = ln(f_\theta(x_1) \cdot f_\theta(x_2) \cdot \ldots \cdot f_\theta(x_n)) = ln(f_\theta(x_1)) + ln(f_\theta(x_2)) + \ldots ln(\cdot f_\theta(x_n))$$

Немного вычислений

> Каждый день, проходя мимо пруда, Олег то видел уток, то не видел. За 30 дней утки плавали 20. Каково, скорее всего, значение вероятности встретить уток завтра?

x - будут/не будут утки

$$x \sim Be(\theta)$$

$$L(\vec{x}|\theta) = \binom{30}{20} \cdot \theta^{20} \cdot (1-\theta)^{10}$$

$$lnL(\vec{x}|\theta) = ln\binom{30}{20} + ln\theta^{20} + ln(1-\theta)^{10}$$

- Наша задача, найти параметр θ, при котором наша функция L(следовательно и $ln(L)$) будет максимальной.

$$ln'(\theta) = \frac{20}{\theta} - \frac{10}{1-\theta}$$

$$\frac{20}{\theta} - \frac{10}{1-\theta} = 0$$

$$20 \cdot (1-\theta) - 10 \cdot \theta = 0$$

$$-30 \cdot \theta = -20$$

$\theta = \frac{20}{30} = \frac{2}{3}$ - Это и будет той вероятностью, при которой у нас наибольшая функция правдоподобия, иначе говоря, вероятность встретить утку, скорее всего $\frac{2}{3}$.

Применение в машинном обучении

Площадь, м2(x_1)	Цена, млн (y)
50	1
33	0
45	0
…	…
76	0

- Попробуем выразить вероятности

Вероятности

- Как можно выразить вероятность на каждом объекте?
- Каждый объект, это распределение Бернулли. ($y \sim Be(\theta)$)
- Этот θ мы не знаем, но хотим найти

А как мы помним из статистики:

$$P(\theta|y = k) = \theta^k \cdot (1 - \theta)^{(1-k)}$$

$$\ln(\theta|P(y = k)) = k \cdot \theta + (1 - k) \cdot (1 - \theta)$$

- В нашем задаче в роли k, выступает значение столбца y, а в роли θ выступает предсказание в вероятностном виде алгоритма.
- Эту функцию как правило максимизируют по параметру θ. Как сделать это задачей минимизации?

Вероятности

- Так же, нам нужно настроить алгоритм машинного обучения таким образом, чтобы наши предсказания были в угоду всем объектам. На выходе получаем BCELoss

$$L = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot ln(\hat{y}_i) + (1 - y_i) \cdot ln(1 - \hat{y}_i)$$

- При машинном оубчении именно минимизация этой функции статистически самая обоснованная