

Обучение логистической регрессии

Елена Кантонистова

Skillbox

Логистическая регрессия

$$a(x) = \frac{1}{1 + e^{-(w,x)}}$$

Чего-то не хватает...

Логистическая регрессия

- Формула для предсказания $a(x) = \hat{y} = (w, x)$
- Функция потерь: $MSE = \frac{1}{l} \sum_{i=1}^l (\hat{y}_i - y_i)^2$
- При обучении минимизируем MSE

Классификатор

Классификатор (а значит, его название) определяется функцией потерь, которую вы минимизируете при обучении!

Логистическая регрессия

- Формула для предсказания:

$$a(x) = \hat{y} = \frac{1}{1 + e^{-(w,x)}}$$

- Функция потерь — логистическая (log-loss):

$$Q(w) = -\frac{1}{l} \sum_{i=1}^l (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

Логистическая регрессия

- Формула для предсказания:

$$a(x) = \frac{1}{1 + e^{-(w,x)}}$$

- Функция потерь — логистическая (log-loss):

$$Q(w) = -\frac{1}{l} \sum_{i=1}^l (y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i))$$

Комментарий: минимизация log-loss обеспечивает предсказание математической вероятности (а не произвольного числа из отрезка $[0;1]$).

Почему не подходит MSE

Так как предсказанные моделью ответы лежат на отрезке $[0:1]$, то MSE одинаково мало штрафует.

- Как абсолютно верные предсказания, например:

$$(y - \hat{y})^2 = (0 - 0)^2 = 0$$

- Так и абсолютно неверные предсказания, например:

$$(y - \hat{y})^2 = (1 - 0)^2 = 1$$

Почему лучше log-loss

- Для абсолютно верного предсказания, например, для $y = \hat{y} = 1$ имеем ошибку

$$-1 \cdot \log(1) = 0$$

- Для абсолютно неверного предсказания, например, для $y = 1, \hat{y} = \varepsilon \rightarrow 0$ имеем ошибку

$$-1 \cdot \log(\varepsilon) \rightarrow \infty$$