Лекция 4
Линейная классификация.

Кантонистова Е.О.

ВШЭ, 2022

ОБУЧЕНИЕ ЛИНЕЙНОЙ РЕГРЕССИИ (НАПОМИНАНИЕ)

Обучающая выборка:

пусть $\boldsymbol{x}$ – объект ($x_1, x_2, \ldots, x_l$ - его признаки), а y – ответ на объекте (произвольное число), n – количество объектов.

Модель линейной регрессии:

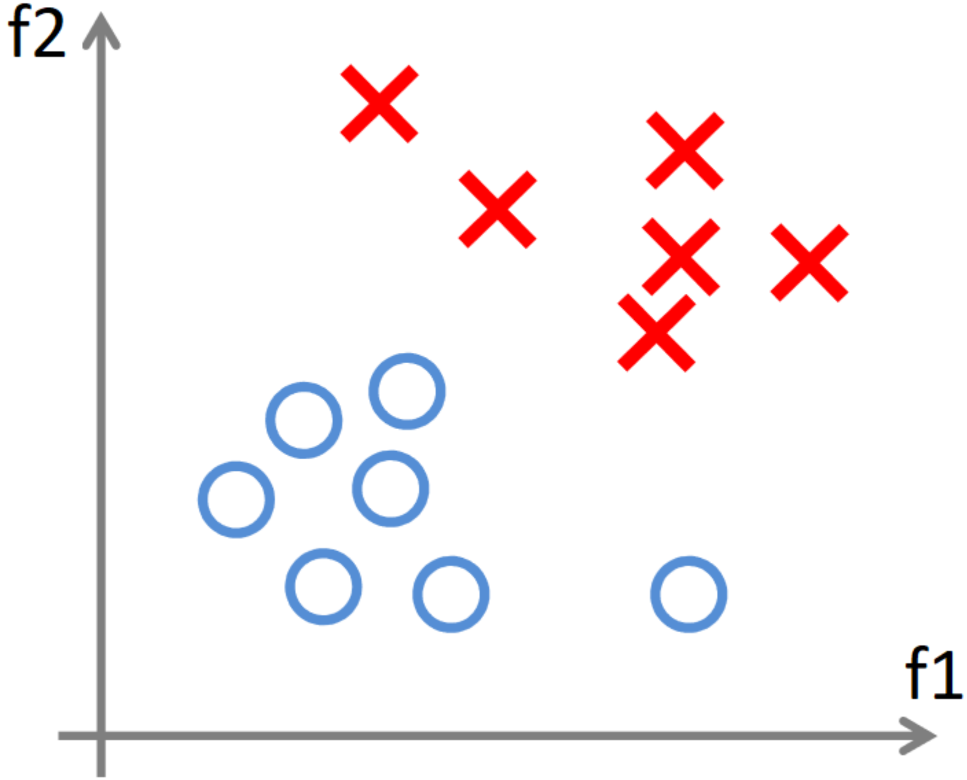
$$a(\boldsymbol{x}, w) = \sum_{i=1}^{l} w_j x_j$$

- Метод обучения – метод наименьших квадратов (***минимизируем разность между предсказанием и правильным ответом***):

$$Q(w) = \sum_{i=1}^{n} (a(\boldsymbol{x}_i, w) - y_i)^2 \to \min_w$$

БИНАРНАЯ КЛАССИФИКАЦИЯ

$y_1, y_2, \ldots, y_n$ - ответы (***+1 или -1***).



БИНАРНАЯ КЛАССИФИКАЦИЯ

Модель линейного классификатора:

$$\boldsymbol{a(x,w) = sign\left(\sum_{j=1}^{l} w_j x_j\right)}$$

БИНАРНАЯ КЛАССИФИКАЦИЯ

Модель линейного классификатора:

$$\boldsymbol{a(x,w) = \color{red}{sign}\left(\sum_{j=1}^{l} w_j x_j\right)}$$

- если $\sum_{j=1}^{l} w_j x_j > 0$, то $sign(\sum_{j=1}^{l} w_j x_j) = +1$, то есть объект отнесён к положительному классу
- если $\sum_{j=1}^{l} w_j x_j < 0$, то $sign(\sum_{j=1}^{l} w_j x_j) = -1$, то есть объект отнесён к отрицательному классу
- значит, $\sum_{j=1}^{l} w_j x_j = 0$ – *уравнение разделяющей границы* между классами. *Это уравнение плоскости* (или прямой в двумерном случае), поэтому *классификатор является линейным*.

ОБУЧЕНИЕ КЛАССИФИКАТОРА

- Обучение - минимизация доли ошибок классификатора:

$$Q(a, X) = \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{a}(\boldsymbol{x_i}) \neq \boldsymbol{y_i}] \rightarrow min \ (*),$$

где $[a(\boldsymbol{x_i}) \neq y_i] = 1$, если предсказание на объекте неверное, то есть $a(\boldsymbol{x_i}) \neq y_i$, и 0 иначе.

- Обозначим $M_i = y_i \cdot (w, \boldsymbol{x_i})$ - ***отступ*** *на i-м объекте*.
- Решение задачи (*) эквивалентно решению задачи

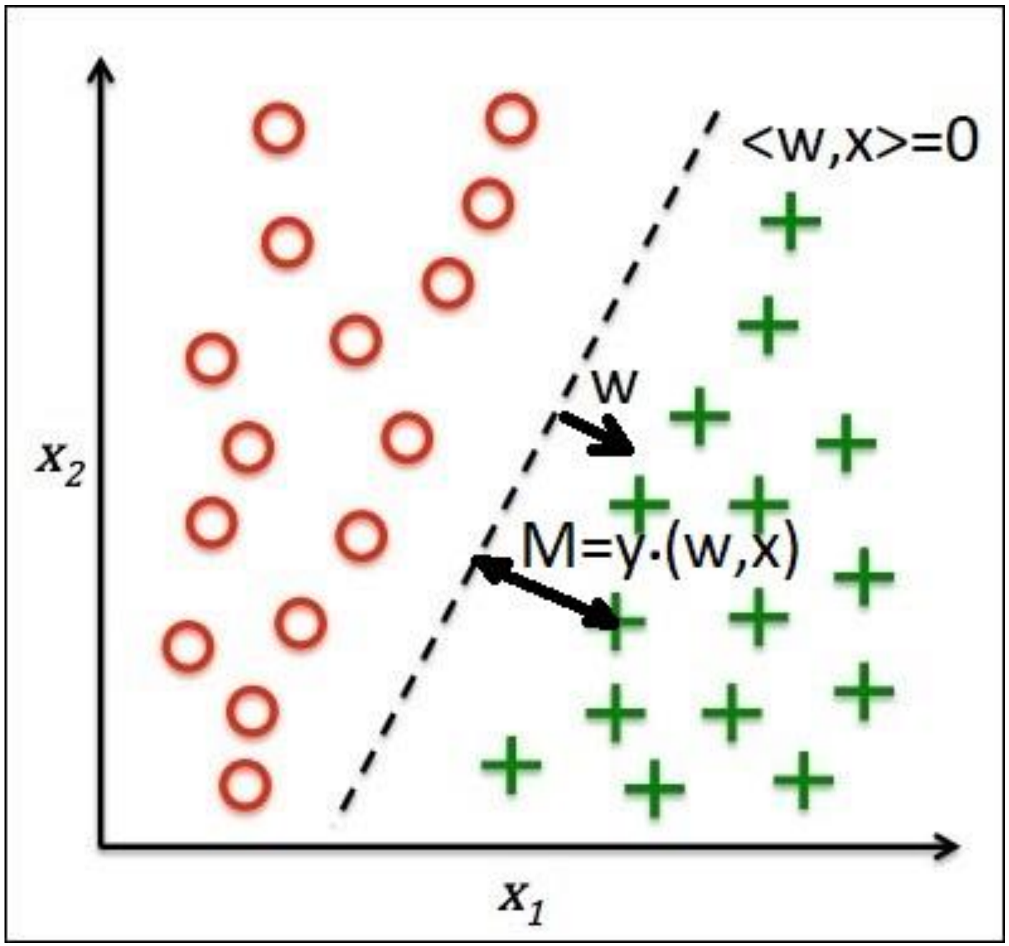
$$Q(a, X) = \frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{M_i} < \boldsymbol{0}] \rightarrow min$$

ОТСТУП (MARGIN)

Знак отступа $M = y \cdot (w, x)$ говорит о корректности классификации на объекте.

ОТСТУП (MARGIN)

Абсолютная величина отступа M обозначает степень уверенности классификатора в ответе (чем ближе M к нулю, тем меньше уверенность в ответе)



ВЕРХНИЕ ОЦЕНКИ ЭМПИРИЧЕСКОГО РИСКА

- $L(a, y) = L(M) = [M < 0]$ **–** разрывная функция потерь

Оценим

$\boldsymbol{L}(\boldsymbol{M}) \leq \tilde{\boldsymbol{L}}(\boldsymbol{M})$, где $\tilde{L}(M)$ - непрерывная или гладкая функция потерь.

- Тогда

$$Q(a, X) = \frac{1}{n}\sum_{i=1}^{n} L(y_i \cdot (w, x_i)) \leq \frac{1}{n}\sum_{i=1}^{n} \tilde{L}\big(y_i \cdot (w, x_i)\big) \to min$$

ФУНКЦИИ ПОТЕРЬ

Минимизируя различные функции потерь, получаем разные результаты. Поэтому разные функции потерь определяют различные классификаторы.

- $L(M) = \log(1 + e^{-M})$ – логистическая функция потерь
- $\mathrm{V}(M) = (1 - \mathrm{M})_+ = \max(0{,}1 - M)$ – кусочно-линейная функция потерь (метод опорных векторов)
- $H(M) = (-\mathrm{M})_+ = \max(0, -M)$ – кусочно-линейная функция потерь (персептрон)
- $E(M) = e^{-M}$ - экспоненциальная функция потерь
- $S(M) = \frac{2}{1+e^{-M}}$ - сигмоидная функция потерь
- $[M < 0]$ – пороговая функция потерь

ОПТИМИЗАЦИЯ ФУНКЦИОНАЛА ПОТЕРЬ

- Нахождение минимума функции потерь $\boldsymbol{Q}$ происходит с помощью метода градиентного спуска:

$$\boldsymbol{w}^{(\boldsymbol{k})} = \boldsymbol{w}^{(\boldsymbol{k}-\mathbf{1})} - \boldsymbol{\eta} \cdot \nabla \boldsymbol{Q}(\boldsymbol{w}^{(\boldsymbol{k}-\mathbf{1})})$$