Логистическая регрессия

$y \in \mathbb{R}$ — **регрессия**

- линейные модели
  - Linear Regression
  - Lasso
  - Ridge
  - ......
- KNN_k

$$\hat{y}_j = \frac{\sum_{i=1}^{k} y_i \cdot \frac{1}{\rho_{ij}}}{\sum_{i=1}^{k} 1/\rho_{ij}}$$

$y \in \{0, 1\}$ $\quad y \in \{-1, 1\}$ — **бинарная классификация**

- линейные модели
  - Logistic Regression
  - SVM (нестандартно)
  - .....
- KNN

$$\hat{y}_j = \underset{h \in Y}{\operatorname{argmax}} \sum_{i=1}^{k} [y_i = h] \cdot \frac{1}{\rho_{ij}}$$

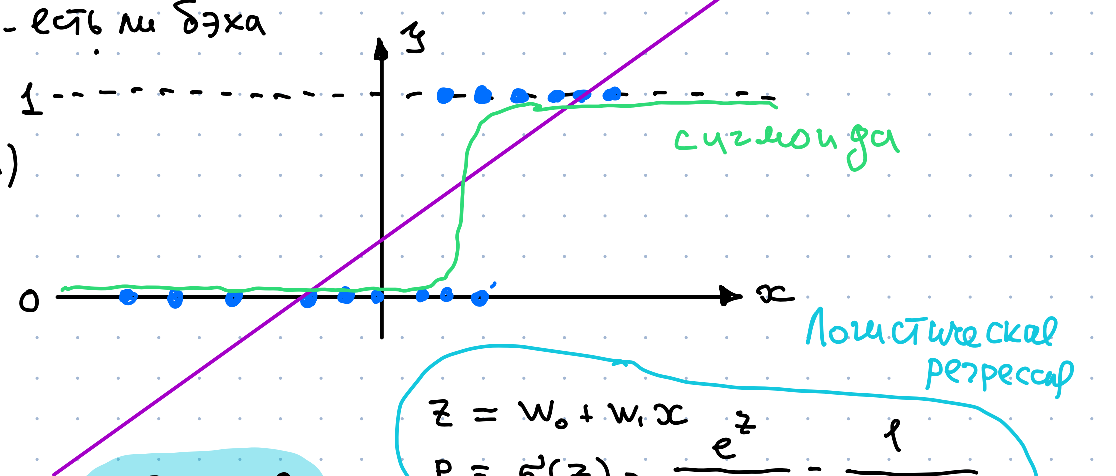
$$Q(w) = \frac{1}{n} \sum_{i=1}^{n} L(y_i, p(x_i)) + R(w) \longrightarrow \min_w$$

А нафига вообще что-то менять?

Давайте учить линейную регрессию!

$y \in \{0, 1\}$ - есть ли бэха

x - доход (центрирован)



Логистическая регрессия

$$z = w_0 + w_1 x$$

$$p = \sigma(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

сигмоида

ф. р-я для логистической сл. вел.

Модель:

$$p(x_i) = \frac{1}{1 + e^{-(w_0 + w_1 x)}} \in [0; 1]$$

можно интерпретировать для некоторых ф. потерь как $\mathbb{P}(y_i = 1 \mid x_i) = p(x_i)$

Как получить прогнозы $\hat{y} \in \{0, 1\}$?

$$\hat{y}_i = \begin{cases} 1, & \hat{p}(x_i) \geq \gamma \\ 0, & \text{иначе} \end{cases}$$

γ – как подобрать?

„бизнес"– соображения

Поговорим позже

Как её обучать?

→ можно использовать MSE

$$L(y_i, \hat{p}(x_i)) = (y_i - \hat{p}(x_i))^2$$

Да, можно но это будет немного странно

Можно придумать ф. потерь получше!

→ logloss (мы её придумаем тремя способами!)

ⓐ (x_i, y_i) – выборка (знаем) инженерный подход

$\hat{y}_i$ – $\{0, 1\}$ – прогнозирует модель

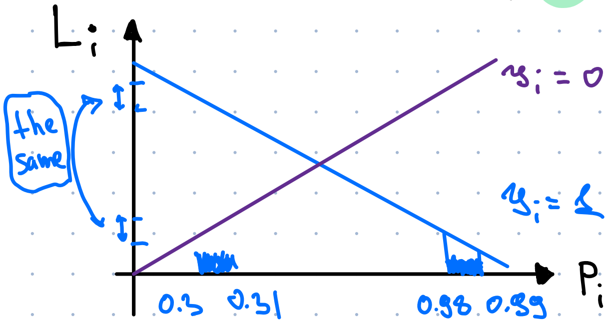
$$\hat{p}(x_i) = \mathbb{P}(y_i = 1 \mid x_i)$$

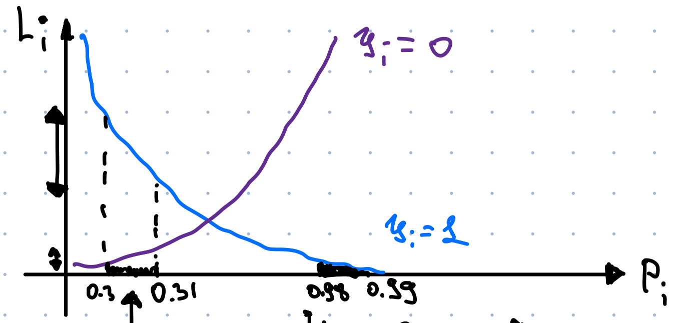
$y_i = 1$ $\quad$ $\hat{p}(x_i)$ – большое – хорошо

$y_i = 0$ $\quad$ $\hat{p}(x_i)$ – большое – плохо

$1 - \hat{p}(x_i)$ – большая

$$L(y_i, \hat{P}(x_i)) = -\left[y_i \cdot \hat{P}(x_i) + (1-y_i)\cdot(1-\hat{P}(x_i))\right] \to \min_w$$





P_i изменилось также, но ошибка ↓ сильнее

подняли P_i ↑ штраф ↓ мало

$$\text{logloss} = -\left[y_i \ln \hat{p}(x_i) + (1-y_i)\ln(1-\hat{p}(x_i))\right] \to \min_w$$

$P(x_i) = 0.98 \quad y_i = 1 \mapsto$ 0.99 хочу быть

$P(x_i) = 0.3 \quad y_i = 1 \mapsto$ 0.31 молодцом

сильнее

8 Вероятностный подход

$$L(w) = \mathbb{P}(y_1, \dots, y_n \mid x, w) = \mathbb{P}(y_1 \mid w, x) \cdot \ldots \cdot \mathbb{P}(y_n \mid x, w) =$$

$$\prod_{i=1}^{n} p(x_i)^{y_i}(1-p(x_i))^{1-y_i}$$

$$y_1 = 1 \quad \frac{1}{1+\exp(-(w_0 + w_1 x_i))} = p(x_i)$$

$$y_1 = 0 \quad 1 - P(x_i)$$

$$\| $$

$$= p(x_1)^{y_1} \cdot (1-p(x_1))^{1-y_1} \cdot \ldots \cdot p(x_n)^{y_n} \cdot (1-p(x_n))^{1-y_n}$$

$$\ln L(w) = \sum_{i=1}^{n}\left[y_i \ln p(x_i) + (1-y_i)\ln(1-p(x_i))\right] \to \max_{w_0, w_1}$$

$$logloss = -1 \cdot \frac{1}{n} \cdot \ln L(w) \to \min_{w_0, w_1}$$

Очень часто какая-то вероятностная модель приводит нас к ф. потерь

$\cancel{n \to \infty}$ $\boxed{\frac{n}{d} \to \infty}$

$\hat{\theta}^{ML}$ – клёвая

- состоятельная
- асимпт. эфф.
- асимпт. несмещ.
- асимпт. норм

Фишер

машинное обучение: $\frac{n}{d} \ll 1$ $\Leftrightarrow$ модели сильно перепараметризованы

$\Rightarrow$ Регуляризация – игрушка дьявола

можно связать её с Байесовскими методами

$$Q(w) = logloss + \lambda \cdot \sum_{j=1}^{d} w_j^2 \to \min_{w}$$

⑥ инженерный, но немного другой

$$\text{Error rate} = 1 - [\hat{y}_i = y_i] = \begin{cases} 0, & \text{да} \\ 1, & \text{нет} \end{cases} = [M_i < 0]$$