Классификация

[redacted]

$X \subset \mathbb{R}^d$

Y — классы объектов

$Y \subset \mathbb{R}^k$ $\quad$ k — кол-во классов

$y = (y_1 \dots y_n)$ $\quad$ $\sum y_i = 1$ $\quad$ y_i — шанс i-ого класса

$f: \mathcal{X} \to Y$ ищем по выборке $(X_1, Y_1) \dots (X_n, Y_n)$

[redacted]

Логистическая регрессия

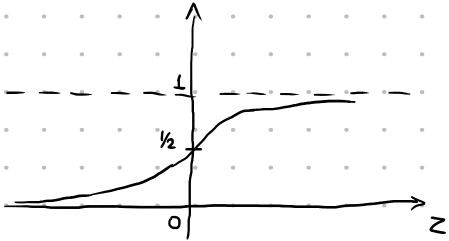
$Y = \{0, 1\}$ — двуклассовая регрессия

$y_1 \dots y_n$ — объекты, где $y_i \sim Bern(\mu_\theta(x_i))$

[redacted]

$\mu_\theta(x_i) = \sigma(\theta^T x_i)$

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



Свойства логистической регрессии

$\{\sigma(\theta^T x) = \frac{1}{2}\} = \{\theta^T x = 0\}$ – линейная модель

где-то здесь разделяющая плоскость

$\{x = 0\}$ – мн-во x, удовл. условию

σ растёт $\Rightarrow$ растёт вероятность класса 1

Св-ва сигмоиды:

1) $\sigma(-z) = 1 - \sigma(z)$

2) $\sigma^{-1}(s) = z(s) = \ln \frac{s}{1-s}$ – логит функция

3) $\sigma'(z) = \sigma(z)(1-\sigma(z))$

$$\frac{(a+b)}{d}c = \frac{e+f}{k} = \frac{l}{n} = t$$ (кста)

Оценим θ

$(x_1, y_1) \dots (x_n, y_n)$ — выборка $\quad y_i \sim Bern(\mu_\theta(x_i))$

Найдём ОМП

$$L_x(\theta) = \prod_{i=1}^{n} \sigma(\theta^T x_0)^{y_i} (1 - \sigma(\theta^T x_i))^{1-y_i}$$

$$L_x(\theta) = \Sigma \; y_i \ln \sigma(\theta^T x_0) + (1-y_i) \ln(1 - \sigma(\theta^T x_i))$$

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^{n} \left[y_i (1 - \sigma(\theta^T x_i)) + (1-y_i)(-\sigma(\theta^T x_i)) \right] x_i =$$

$$= \sum_{i=1}^{n} \left(y_i - \sigma(\theta^T x_i) \right) x_i = X^T (y - S(\theta)) \quad \Bigg| \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad S(\theta) = \begin{pmatrix} \sigma(\theta^T x_1) \\ \vdots \\ \sigma(\theta^T x_n) \end{pmatrix}$$

Градиентный спуск (подъём)

$$\theta_{k+1} = \theta_k + \gamma \nabla F(\theta_k)$$

↑ + т.к. мы макс ф-цию правдоподобия

$$\theta_{k+1} = \theta_k + \gamma X^T (y - S(\theta_k)) \qquad \text{GD}$$

$$\theta_{k+1} = \theta_k + \gamma \sum_{i \in I} \frac{n}{|I|} (y_i - \sigma(\theta^T x_i)) x_i \qquad \text{SGD}$$

[redacted]

Переобучение

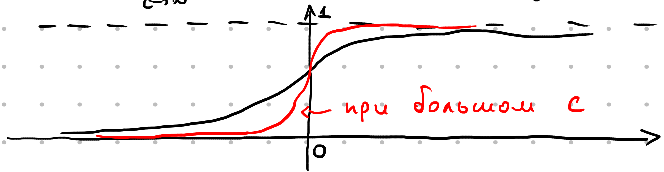
Пример: классы линейно разделимы

среди пр-ков есть единичный

$\exists \Theta$, т.ч. $\{\Theta^T X = 0\}$ — разделяет классы

Тогда $\forall c > 0 \in \mathbb{R}$ $\{c\,\Theta^T X = 0\} = \{\Theta^T X = 0\}$ — тоже раздел. гиперплоскость

$\sigma(c\,\Theta^T X) \xrightarrow[c\to\infty]{} 1$ т.е. вер-ть одного класса повышается



Проблемы н.н. ≈ 0, либо н.н. ≈ 1

$\theta^T X \approx 0 \Rightarrow$ сложно классифицировать

Чтобы избавиться от проблем применим

Регуляризацию

IRLS

$F(\theta) \to \min\limits_{\theta}$ — выпуклая

Метод Ньютона $\theta_{k+1} = \theta_k - \nabla\nabla F_k(\theta_k)^{-1} \cdot \nabla F(\theta_k) =$

Линейная регрессия $F(\theta) = \| y - X\theta \|^2$

$$\nabla F(\theta) = -2X^T(y - X\theta)$$

$$\nabla\nabla F(\theta) = 2X^TX$$

$= (X^TX)^{-1}X^Ty$ — т.е. за один шаг метод Ньютона сойдётся

[redacted]

Логистическая регрессия

$$F(\theta) = -l_{x,y}(\theta) = -\sum_{i=1}^{n}\left[y_i \ln \sigma(\theta^T x_i) + (1-y_i)\ln(1-\sigma(\theta^T x_i))\right]$$

[redacted]

$$\nabla F(\theta) = -\sum_{i=1}^{n}\left(y_i - \sigma(\theta^T x_i)\right)x_i = -X^T(y - S(\theta))$$

$$\nabla\nabla F(\theta) = \left(\sum_{i=1}^{n}\sigma(\theta^T x_i)x_i\right)' = -\sum_{i=1}^{n}\left[x_i^T\sigma(\theta^T x_i)\left(1-\sigma(\theta^T x_i)\right)x_i\right] =$$

$$= X^T V(\theta) X \qquad V(\theta) = \mathrm{diag}\left[\sigma(\theta^T x_i)(1-\sigma(\theta^T x_i))\right]$$

$$\theta_{k+1} = \theta_k - (X^T V(\theta_k) X)^{-1} X^T (S(\theta_k) - y) =$$

$$= \left(X^T V(\theta_k) X\right)^{-1} X^T V(\theta_k) Z(\theta_k)$$

zge $Z(\theta_k) = \left[X\theta_k - V(\theta_k)^{-1}\left(S(\theta_k) - y\right)\right]$