$\boxed{1}$ MVE440 / MSA 220

Lecture 2, 2019-03-27

Classification rule for 0-1 regression

Since $p(1|x) \approx x^T\beta$ it follows that $p(0|x) \approx 1 - x^T\beta$.

By Bayes' rule: $\hat{c}(x) = \arg\max_i p(i|x)$, i.e.

class 0 is chosen when $1 - x^T\beta \geq x^T\beta \Rightarrow x^T\beta \leq \frac{1}{2}$

Dummy coding to seq. of 0-1-regressions

Encode $\underset{\text{class}}{i_\ell} \rightsquigarrow \underset{\text{vector in } \mathbb{R}^K \text{ in dummy coding}}{z_\ell}$. The resulting regression problem is

$$(*) \quad \arg\min_{B \in \mathbb{R}^{K \times p}} \sum_{\ell=1}^{n} \| z_\ell - Bx_\ell \|_2^2$$

(coeff. matrix since we are regressing a vector)

Euclidean/ℓ_2 norm

But since

$$\| z_\ell - Bx_\ell \|_2^2 = \sum_{j=1}^{K} (\underset{j\text{-th component}}{z_\ell^{(j)}} - \beta_j^T x_\ell)^2$$

β_j: j-th row of B

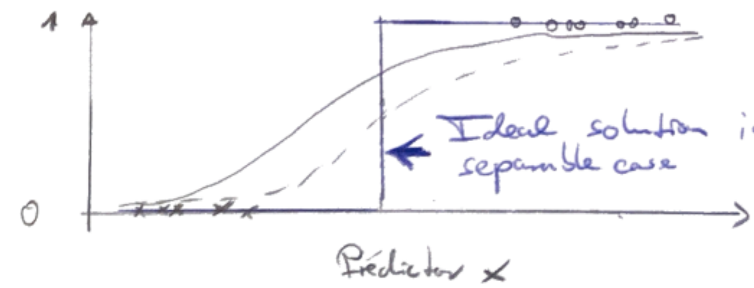
$$\rightsquigarrow (*) = \arg\min_{\beta_1, \ldots, \beta_K \in \mathbb{R}^p} \sum_{\ell=1}^{n} \sum_{j=1}^{K} (z_\ell^{(j)} - \beta_j^T x_\ell)^2 =$$

$$= \arg\min_{\beta_1, \ldots, \beta_K} \sum_{j=1}^{K} \left[\sum_{\ell=1}^{n} (z_\ell^{(j)} - \beta_j^T x_\ell)^2 \right]$$

$$z_\ell^{(j)} = \begin{cases} 0 & i_\ell \neq j \\ 1 & i_\ell = j \end{cases}$$

$\Rightarrow K$ independent 0-1 regressions

Numerical instability of logistic regression in two-class-case for perfect separability



To achieve the step like function in logistic regression:

$$\frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))} \quad \text{for } \beta_0 \to -\infty, \ \beta_1 \to +\infty$$

Softmax identifies only K-1 parameters

Two ways to show this

$\boxed{1}$ Since $\sum_{i=1}^{K} \sigma_i(z) = \frac{\sum_{i=1}^{K} e^{z_i}}{\sum_{\ell=1}^{K} e^{z_\ell}} = 1 \Rightarrow$ Fixing $K-1$ values determines the K-th value

$\boxed{2}$ It holds that

$$\sigma_i(z) = \frac{e^{z_i - z_K}}{1 + \sum_{\ell=1}^{K-1} e^{z_\ell - z_K}}, \quad \sigma_K(z) = \frac{1}{1 + \sum_{\ell=1}^{K-1} e^{z_\ell - z_K}}$$

$\Rightarrow K-1$ values $(z_i - z_K)$, $i = 1, \ldots, K-1$ enough to determine all values of $\sigma_i(z)$

LDA and $\Sigma_i = \sigma^2 I$

$$\delta_i(x) = \log \pi_i - \frac{1}{2}(x - \mu_i)^T \sigma^{-2} I (x - \mu_i) - \underbrace{\frac{1}{2} \log |\sigma^2 I|}_{\text{constant w.r.t. } i} =$$

$$= \log \pi_i - \frac{1}{(2\sigma^2)} \underbrace{\| x - \mu_i \|_2^2}_{\text{nearest centroids}}$$