Machine Learning Exercise 3.1

Raphael Schönball & Simeon Doetsch

1 LDA Derivation from Least Squares Error

Decision boundary: $\hat{\omega}^T x + \hat{b} = 0$

decision rule: $\hat{y} = \mathrm{sgn}(\hat{\omega}^T x + \hat{b})$

Training data: N annotated datapoints, with assumption that data is balanced: $N_1 = N_{-1} = \frac{N}{2}$

LSQ-Criterion: $\hat{\omega}, \hat{b} = \underset{\omega, b}{\mathrm{argmin}} \sum_{i=1}^{N} (\omega^T x_i + b - y_i)^2$

a) find optimal $\hat{b}$:

$$\frac{\partial}{\partial b} \sum_{i=1}^{N} (\omega^T x_i + b - y_i)^2 \overset{!}{=} 0 = \sum_{i=1}^{N} 2(\omega^T x_i + b - y_i)$$

$$\Rightarrow \sum_{i=1}^{N} \omega^T x_i + \sum_{i=1}^{N} b - \sum_{i=1}^{N} y_i = 0$$

$\sum_{i=1}^{N} y_i = 0$ because $y_i \in \{-1, 1\}$, and in same amount because data is balanced

$$\Rightarrow N\hat{b} = -\sum_{i=1}^{N} \omega^T x_i = -\omega^T \sum_{i=1}^{N} x_i$$

$$\Rightarrow \hat{b} = -\frac{\omega^T \sum_{i=1}^{N} x_i}{N}$$

$$\Rightarrow \sum_{i=1}^{N} x_i = \sum_{i: y_i = -1} x_i + \sum_{i: y_i = 1} x_i = (\mu_{-1} + \mu_1) \cdot \frac{N}{2}$$

$$\Rightarrow \hat{b} = -\omega^T \frac{\mu_{-1} + \mu_1}{2}$$

b) transform $\frac{\partial}{\partial \omega} \sum_{i=1}^{N} (\omega^T x_i + \hat{b} - y_i)^2 = 0$ into $(S_W + \frac{1}{4} S_B)\hat{\omega} = \mu_1 - \mu_{-1}$

With $\mu_{-1} = \frac{1}{N_{-1}} \sum_{i: y_i = -1} x_i = \frac{2}{N} \sum_{y_i = -1} x_i$; $\mu_1 = \frac{1}{N_1} \sum_{i: y_i = 1} x_i = \frac{2}{N} \sum_{y_i = 1} x_i$

$S_B = (\mu_1 - \mu_{-1})(\mu_1 - \mu_{-1})^T$ $\quad S_W = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^T$

$$\Rightarrow \frac{\partial}{\partial \omega} \sum_{i=1}^{N} (\omega^T x_i + \hat{b} - y_i)^2 = \frac{\partial}{\partial \omega} \sum_{i=1}^{N} \left(\omega^T x_i - \omega^T \frac{\mu_{-1} + \mu_1}{2} - y_i\right)^2$$

$$= \frac{\partial}{\partial \omega} \sum_{i=1}^{N} \left(\omega^T \left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right) - y_i\right)^2 = 2 \sum_{i=1}^{N} \left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right)\left(\omega^T \left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right) - y_i\right)$$

$$= 2 \sum_{i=1}^{N} \left[\left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right) \omega^T \left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right) - x_i y_i + y_i \frac{\mu_{-1} + \mu_1}{2}\right]$$

Vector relation on assignment

$$= 2 \sum_{i=1}^{N} \left[\left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right)\left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right)^T \omega - x_i y_i + y_i \frac{\mu_{-1} + \mu_1}{2}\right]$$

$$= 2\left[\left(\sum_{i=1}^{N} \left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right)\left(x_i - \frac{\mu_{-1} + \mu_1}{2}\right)^T\right)\omega - \sum_{i=1}^{N} x_i y_i + \frac{\mu_{-1} + \mu_1}{2} \underbrace{\sum_{i=1}^{N} y_i}_{=0}\right] \overset{!}{=} 0$$

$$\Rightarrow \sum_{i=1}^{N} x_i y_i = \sum_{i: y_i = 1} x_i + \sum_{i: y_i = -1} (-x_i) = \frac{N}{2}(\mu_1 - \mu_{-1})$$

$$\Rightarrow \underbrace{\left[\frac{2}{N}\sum_{i=1}^{N}\left(x_i - \frac{\mu_1+\mu_{-1}}{2}\right)\left(x_i - \frac{\mu_1+\mu_{-1}}{2}\right)^T\right]}_{S_W + \frac{1}{4}S_B}\hat{w} = \mu_1 - \mu_{-1}$$

$$\Rightarrow \left(S_W + \frac{1}{4}S_B\right)\hat{w} = \mu_1 - \mu_{-1}$$