Fundamentals of Machine Learning - Exercise 3

Eugen Dizer

Luis Pazos Clemens

Robert Freund

3. LDA - Derivation:

3.1: start with: $\frac{\partial}{\partial b} \sum_{i=1}^{N} (w^T x_i + b - y_i)^2 = 0$

using b is a scalar: $\sum_{i=1}^{N} 2 (w^T x_i + b - y_i) = 0$

$\Rightarrow Nb = \sum_{i=1}^{N} y_i - w^T \sum_{i=1}^{N} x_i$

using: $\sum_{i=1}^{N} y_i = N_1 y_1 + N_{-1} y_{-1} = \frac{N}{2}(1-1) = 0$

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

$\Rightarrow b = -\frac{1}{2} w^T (\mu_1 + \mu_{-1})$

3.2: $\frac{\partial}{\partial w} \sum_{i=1}^{N} (w^T x_i + b - y_i)^2 = 0$

One can carry this out by component but end up with:

$\sum_{i=1}^{N} 2 (w^T x_i + b - y_i) x_i = 0$

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

using vector rule: $(b^T \cdot c) a = (a \cdot c^T) b$

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

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

$$\Rightarrow \left[\sum_{i=1}^{N}(x_i x_i^T) - \frac{N}{4}(\mu_1+\mu_{-1})(\mu_1+\mu_{-1})^T\right]\omega = \frac{N}{2}(\mu_1-\mu_{-1})$$

using relation $(*)$.

$$\Rightarrow \Big[\overbrace{\frac{1}{N}\sum_{i=1}^{N}(x_i x_i^T) - \frac{1}{2}(\mu_1\mu_1^T+\mu_{-1}\mu_{-1}^T)}^{(**)} + \frac{1}{4}\underbrace{(\mu_1\mu_1^T-\mu_1\mu_{-1}^T-\mu_{-1}\mu_1^T+\mu_{-1}\mu_{-1}^T)}_{=(\mu_1-\mu_{-1})(\mu_1-\mu_{-1})^T=S_B}\Big]\omega$$

$$= \frac{\mu_1-\mu_{-1}}{2}$$

only focusing on $(**)$:

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

$$= \frac{1}{N}\Big[\sum_{i=1}^{N}(x_i x_i^T) + \underbrace{\frac{N}{2}(\mu_1\mu_1^T+\mu_{-1}\mu_{-1}^T)}_{=\sum_{i=1}^{N}\mu_{y_i}\mu_{y_i}^T} \underbrace{-\mu_1\frac{N}{2}\mu_1^T-\mu_{-1}\frac{N}{2}\mu_{-1}^T}_{=-\sum_{i=1}^{N}\mu_{y_i}x_i^T} \underbrace{-\mu_1\frac{N}{2}\mu_1^T-\mu_{-1}\frac{N}{2}\mu_{-1}^T}_{=-\sum_{i=1}^{N}x_i\mu_{y_i}^T}\Big]$$

using $(*)$ and analog derivations.

using: $\sum_{i=1}^{N}(x_i x_i^T - x_i\mu_{y_i}^T - \mu_{y_i}x_i^T + \mu_{y_i}\mu_{y_i}^T) = \sum_{i=1}(x_i-\mu_{y_i})(x_i-\mu_{y_i})^T$

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

$$\Rightarrow \left(S_W + \frac{1}{4}S_B\right)\omega = \frac{\mu_1-\mu_{-1}}{2}$$