Advanced Data Analysis
Homework Week 4

Aswin Vijay

May 15, 2023

Homework 4

Given the least squares classification objective as,

$$\hat{\theta} = arg\min_{\theta} \frac{1}{2}\sum_{i=1}^{n}(f_{\theta}(x_i) - y_i)^2$$

with mean zero inputs,

$$\frac{1}{n}\sum_{i=1}^{x_i} = 0 \tag{1}$$

and linear input model,

$$f_{\theta}(x) = \theta^T x$$

Given the solution to the least squares objective as,

$$(X^T X)\hat{\theta} = X^T y \tag{2}$$

The class means and variance are given as,

$$\mu_- = \frac{1}{n_-}\sum_{i:y_i=-1} x_i$$
$$\mu_+ = \frac{1}{n_+}\sum_{i:y_i=1} x_i$$
$$\hat{\Sigma}_- = \frac{1}{n_-}\sum_{i:y_i=-1}(x_i - \mu_-)(x_i - \mu_-)^T$$
$$\hat{\Sigma}_+ = \frac{1}{n_+}\sum_{i:y_i=1}(x_i - \mu_+)(x_i - \mu_+)^T$$
$$\hat{\Sigma} = \frac{1}{n}\left(n_+\hat{\Sigma}_+ + n_-\hat{\Sigma}_-\right)$$

where $\hat{\Sigma}$ is the MLE of the common covariance matrix.

Next we try to express Eqn (1) and (2) in terms of the means and covariances.

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{x_i} &= 0 \\
\frac{1}{n}(n_-\mu_- + n_+\mu_+) &= 0 \\
\mu_- &= -\frac{n_+\mu_+}{n_-}
\end{aligned}$$

Let the y labels be $\{\frac{-1}{n_-}, \frac{1}{n_+}\}$, then for RHS of (2) we have

$$\begin{aligned}
X^T y &= -\frac{1}{n_-}\sum_{i:y_i=-1} x_i + \frac{1}{n_+}\sum_{i:y_i=1} x_i \\
&= \mu_+ - \mu_- \\
&= \mu_+ \frac{n}{n_-}
\end{aligned}$$

For LHS of (2) we have,

$$\begin{aligned}
\hat{\Sigma} &= \frac{1}{n}\left(\sum_{i:y_i=-1}(x_i-\mu_-)(x_i-\mu_-)^T + \sum_{i:y_i=1}(x_i-\mu_+)(x_i-\mu_+)^T\right) \\
&= \frac{1}{n}\left(\sum_{i:y_i=-1}(x_i x_i^T - x_i\mu_-^T - \mu_- x_i^T + \mu_-\mu_-^T) + \sum_{i:y_i=1}(x_i x_i^T - x_i\mu_+^T - \mu_+ x_i^T + \mu_+\mu_+^T)\right) \\
&= \frac{1}{n}\left(X^T X + -n_-\mu_-^T\mu_- - n_+\mu_+^T\mu_+\right) \\
X^T X &= n\hat{\Sigma} + n_-\mu_-^T\mu_- + n_+\mu_+^T\mu_+ \\
X^T X &= n\hat{\Sigma} + \frac{n^2}{n_-}\mu_+\mu_+^T + n_+\mu_+\mu_+^T
\end{aligned}$$

Using the above results in Eqn (2) we get,

$$\begin{aligned}
\left[n\hat{\Sigma} + \frac{n^2}{n_-}\mu_+\mu_+^T + n_+\mu_+\mu_+^T\right]\hat{\theta} &= \mu_+\frac{n}{n_-} \\
\left[\hat{\Sigma} + (\frac{n}{n_-} + n_+)\mu_+\mu_+^T\right]\hat{\theta} &= \mu_+\frac{1}{n_-} \\
\hat{\Sigma}\hat{\theta} + (\frac{n}{n_-} + n_+)c\mu_+ &= \mu_+\frac{1}{n_-} \quad using\ vv^T\theta = cv \\
\hat{\Sigma}\cdot\hat{\theta} &= \mu_+(\frac{1}{n_-} - c(\frac{n}{n_-} + n_+)) \\
\hat{\Sigma}\cdot\hat{\theta} &= (\mu_+ - \mu_-)\frac{n_-}{n}(\frac{1}{n_-} - c(\frac{n}{n_-} + n_+)) \\
\hat{\Sigma}\cdot\hat{\theta} &= (\mu_+ - \mu_-)(\frac{1}{n} - c(1 + \frac{n_+n_-}{n})) \\
\hat{\theta} \propto \hat{\Sigma}^{-1}(\mu_+ - \mu_-)
\end{aligned}$$

Thus we get he desired result