Lec 11 p12

Fisher's Approach

Result 11.3

The linear combination $\hat{y} = \hat{a}'x = (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} x$ maximizes the ratio

$$\max_a \frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = \max_a \frac{(a'\bar{x}_1 - a'\bar{x}_2)^2}{\boxed{a' S_{pooled} a}} = \max_a \frac{(a'd)^2}{\boxed{a' S_{pooled} a}}$$

over all possible coefficient vectors a, where $d = \bar{x}_1 - \bar{x}_2$.

The maximum of this ratio is $D^2 = (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} (\bar{x}_1 - \bar{x}_2)$.

Linear discriminant analysis (LDA)

Q: What is the intuition of separation?
Does the D^2 looks familiar?

Lec 12 p70

Model Selection

- Any model selection criterion (AIC, likelihood ratio, BIC) can be used to select the best fitting model.
- Mclust uses the Bayesian Information Criterion (BIC) to choose the best model

$$BIC = -2\log(L) + \log(n)m$$

  where L is the likelihood function, m is the number of free parameters to be estimated and n is the number of observations. A model with low BIC fits the data better than one with high BIC.

- Akaike information criterion (AIC)

$$AIC = -2\log(L) + \boxed{2m}$$

- Note that the formulas in textbook and Mclust use reversed sign. So the higher the better.