2016-03-18 class-notes (partial)

(continued)

- generate 15000 observations from 3 sub-populations using the estimated mean vectors $\hat{\boldsymbol{\mu}}_{\mathbf{1}}, \hat{\boldsymbol{\mu}}_{\mathbf{2}}, \hat{\boldsymbol{\mu}}_{\mathbf{3}}$ and covariance matrices $\boldsymbol{C}_{\mathbf{1}}, \boldsymbol{C}_{\mathbf{2}}, \boldsymbol{C}_{\mathbf{3}}$
- classification rate every similar: 1.2% of simulated observations are classified differently.

Alternative mehtods for classification

So far: have data $(g_1, \boldsymbol{x}_{\mathbf{1}}), \ldots, (g_n, \boldsymbol{x}_{\boldsymbol{n}})$
Model:$(G, X) \rightarrow P(G = j) = \lambda_j$ and conditional on $G = j$, X has density $f_j(X)$.

$$P(G = j|X = \boldsymbol{x}) = \frac{\lambda_j f_j(\boldsymbol{x})}{\sum_{l=1}^{k} \lambda_l f_l(\boldsymbol{x})}$$

quantity of interest

LDA, QDA: Assume $f_1(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x})$ multivariates normal densities -- use data to estimate unknowns.

Key point: Explicitly model distributions of X.
But in prctice, this is difficult to do

- discrete variables
- P may be very large

Alternative approach: model $P(G = j|X = \boldsymbol{x})$ directly.

- analogous to regression modeling
  - G is the response
  - X is the predictors
- we implicityly assume that the distribution of X (i.e. not condtional on $G = j$) is not particularly informative.

Multiple regression: $y_i = \boldsymbol{x}_{\boldsymbol{i}}^T \boldsymbol{\beta} + \epsilon_i$ --> look at conditional distributions of response given predictors.

Special case: $k = 2$

$$P(G = 1|X = \boldsymbol{x}) = 1 - P(G = 2|X = \boldsymbol{x})$$

where the LHS is $\theta(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{\beta})$, and $\boldsymbol{\beta}$ is unknown parameters.

Logistic regression model

$$\theta(\boldsymbol{x}) = \frac{\exp(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta})}{1 + \exp(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta})}$$

$$1 - \theta(\boldsymbol{x}) = P(G = 2|\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{1 + \exp(\beta_0 + \boldsymbol{x}^T \boldsymbol{\beta})}$$

Note that $0 < \theta(\boldsymbol{x}) < 1$ for any $\beta_0, \boldsymbol{\beta}$

- **logit transform**: