LINEAR DISCRIMINANT ANALYSIS

Plug-In Methods

A plug-in classifier is obtained by estimating the terms appearing in the formula for the Bayes classifier. Recall

$$f^*(x) = \arg\max_k \; \pi_k \, g_k(x)$$

LDA and Naive Bayes

$$= \arg\max_k \; \eta_k(x)$$

Logistic regression

LDA

Suppose we have training data $(X_1, Y_1), \ldots, (X_n, Y_n)$.

In LDA, we assume $X \mid Y = k \sim N(\mu_k, \Sigma)$, i.e.

$$g_k(x) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left(-\tfrac{1}{2}(x-\mu_k)^T \Sigma^{-1} (x-\mu_k)\right)$$

LDA is the classifier obtained by plugging the following estimates into the Bayes classifier formula:

$$\hat{\pi}_k = \frac{n_k}{n}, \quad n_k = |\{i : y_i = k\}|$$

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i : y_i = k} x_i$$

$$\underbrace{\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^T}_{\text{pooled covariance estimate}}$$

maximum likelihood estimates

Questions

Is LDA generative or discriminative?
Parametric or nonparametric? Linear or nonlinear?

To answer the last question, we can write the LDA classifier as

$$\hat{f}(x) = \arg\max \; \hat{\pi}_k \, \hat{g}_k(x)$$

$$= \arg\max \; \log \hat{\pi}_k - \frac{1}{2}(x - \hat{\mu}_k)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_k)$$

Focus on $K=2$:

$$(x-\hat{\mu}_1)^T \hat{\Sigma}^{-1} (x-\hat{\mu}_1) - 2\log \hat{\pi}_1 \gtrless_{1}^{2} (x-\hat{\mu}_2)^T \hat{\Sigma}^{-1} (x-\hat{\mu}_2) - 2\log \hat{\pi}_2$$

$$(x-\hat{\mu}_1)^T \hat{\Sigma}^{-1}(x-\hat{\mu}_1) - 2\log \hat{\pi}_1 \underset{1}{\overset{2}{\gtrless}} (x-\hat{\mu}_2)^T \hat{\Sigma}^{-1}(x-\hat{\mu}_2) - 2\log \hat{\pi}_2$$

$$\Longleftrightarrow \quad \underbrace{a^T x + b}_{\text{discriminant function}} \underset{1}{\overset{2}{\gtrless}} 0$$

Key point: quadratic terms cancel because of assumed common covariance matrix

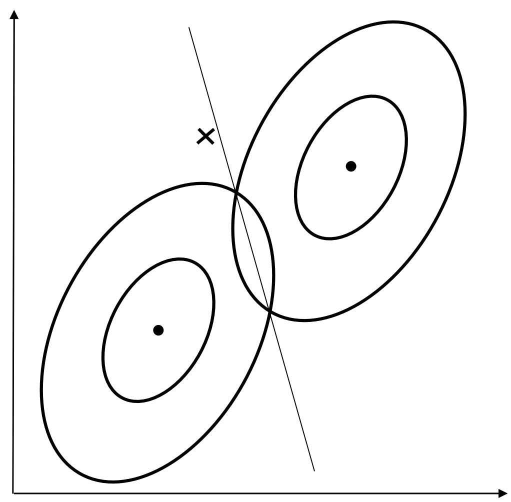

Figure assumes $\hat{\pi}_1 = \hat{\pi}_2$.

Which $\hat{\mu}_k$ is the "x" closer to?

Interpretation: nearest Mahalanobis distance classifier

What do the decision regions look like when $K > 2$?

What are the drawbacks of LDA?

$\boxed{\text{QDA}}$

Quadratic discriminant analysis results when the generative model is $X \mid Y=k \sim N(\mu_k, \Sigma_k)$.

Then the discriminant function is quadratic.