Recall: Logistic Regression involves modeling directly $P(Y=k \mid X=x)$
Using logistic function

$$p(x) = \frac{e^{\beta_0+\beta_1 x_1+\dots\beta_p x_p}}{1+e^{\beta_0+\beta_1 x+\dots+\beta_p x_p}}$$

LDA (Linear Discriminant Analysis)

- we model the distribution of predictors X separately in each of the response classes.
- Then $\Rightarrow$ use Bayes' Theorem to estimate $P(Y=k \mid X=x)$

Using Bayes' Theorem for Classification

Assume that we want to classify an observation into one of K classes, where $K \geq 2$

Let π_k = overall or "prior probability" that a randomly chosen observation comes from the k^{th} class.

Let $f_k(x) = P(X=x \mid Y=k)$ = "density function of X" for an observation that comes from k^{th} class

Bayes' Theorem:

$$P(Y=k \mid X=x) = \frac{\pi_k f_k(x)}{\sum_{i=1}^{K} \pi_i f_i(x)}$$

Again $p_k(x) = P(Y=k \mid X=x)$

→ Instead of computing $p_k(x)$ directly, we can use estimates π_k and $f_k(x)$

π_k - easy to compute (take fraction of training data that belongs to k^{th} class

$f_k(x) \rightarrow$ more difficult to estimate

$p_k(x) =$ "posterior probability" that an observation $X = x$ belongs to class k.

<u>LDA for $p = 1$</u>

Assume $f_k(x)$ is normal (Gaussian)

$$\Rightarrow f_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k} e^{-\frac{1}{2\sigma_k^2}(x-\mu_k)^2}$$

Assume $\sigma_1^2 = \sigma_2^2 \cdots = \sigma_k^2$

$$\Rightarrow p_k(x) = \frac{\pi_k\, e^{-\frac{1}{2\sigma^2}(x-\mu_k)^2}}{\sum_{i=1}^{K} \pi_i \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu_i)^2}}$$

The Bayes classifier assigns an observation $X = x$ to the class for which $p_k(x)$ is largest.

Take $\log[p_k(x)] = \log[\quad] - \log[\quad]$

This is equivalent to assigning the observation to the class for which:

$$\delta_k(x) = x\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

is largest.

Ex: if $K=2$ and $\pi_1 = \pi_2$, then Bayes classifier assigns an observation to class 1, if

$$\delta_1(x) > \delta_2(x)$$

$$x\frac{\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \cancel{\log(\pi_1)} > x\frac{\mu_2}{\sigma^2} - \frac{\mu_2^2}{2\sigma^2} + \cancel{\log(\pi_2)}$$

$$\Leftrightarrow \quad 2x(\mu_1 - \mu_2) > \mu_1^2 - \mu_2^2$$

The Bayes decision boundary corresponds to a point where $\delta_1(x) = \delta_2(x)$

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2}$$

On slide: $\mu_1 = -1.25$; $\mu_2 = 1.25$, $\sigma_1^2 = \sigma_2^2 = 1$

$\pi_1 = \pi_2 = 0.5$

$\Rightarrow$ if $x < 0 \Rightarrow$ class 1

$x > 0 \Rightarrow$ class 2

LDA for $p > 1$

Assume $X = (x_1, x_2 \dots x_p)$ - multivariate Gaussian
with class-specific mean vector and common covariance matrix.

$X \sim N(\mu, \Sigma)$ $E[x] = \mu$
$Cov[x] = \Sigma$

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

Bayes classifier assigns an observation $X = x$
to the class for which

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log(\pi_k)$$

is largest.

QDA (Quadratic Discriminant Analysis)

- Assumes that each class has its own covariance matrix

$X \sim N(\mu_k, \Sigma_k)$

Under this assumption, the Bayes Classifier assigns an observation $X=x$ to the class for which

$$\delta_k(x) = -\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\Sigma_k| + \log(\pi_k) =$$

$$\delta_k(x) = -\frac{1}{2}x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^T \Sigma_k^{-1}\mu_k - \frac{1}{2}\log|\Sigma_k| + \log \pi_k$$

is the largest

in this case we have a quadratic function