Chp. 4: Discriminant Analysis

Discriminant Analysis (LDA and QDA)
- classification w/$>$2 classes
- classes are well-separated
- small *n* when distribution of predictors is approximately normal in each of the classes

Discriminant Analysis

Instead of estimating $P(Y|X)$ directly, the approach is to estimate

1. $\widehat{P}(X|Y)$, the distribution of X in each of the k categories separately
2. $\widehat{P}(Y)$, the likelihood of each category

And use Bayes rule to obtain $P(Y|X)$:

$$P(Y = k|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Discriminant Analysis

We write this in a little different form for discriminant analysis:

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)} \text{ , where}$$

- $f_k(x) = P(X|Y)$ is the density for X in class k, assuming a normal (Gaussian) distribution for each class and
- $\pi_k = P(Y)$ is the prior probability for class k.

Linear Discriminant Analysis

$f_k(x)$ is large if there is a high probability that an observation in class k has $X \approx x$ and small if it is unlikely $X \approx x$.

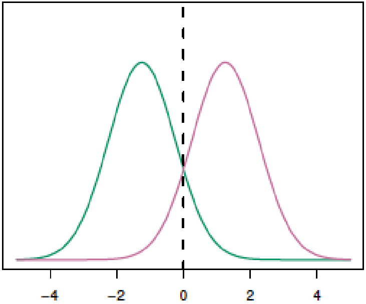


Figure 1: Fig 4.4a

Fig. 4.4: One-dimensional normal density functions, $f_k(x)$, for two different classes (green and pink). Dashed line: optimal decision boundary from Bayes classifier

Linear Discriminant Analysis

In practice, we can estimate f_{green} and f_{pink} based on observed data.

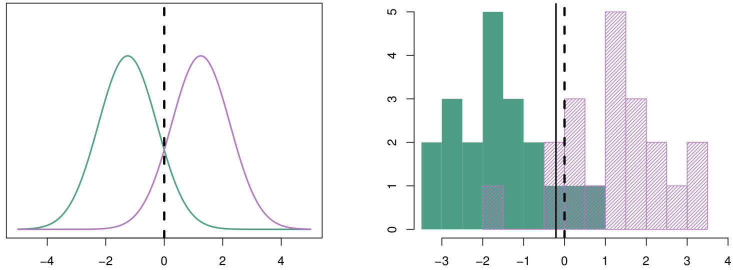
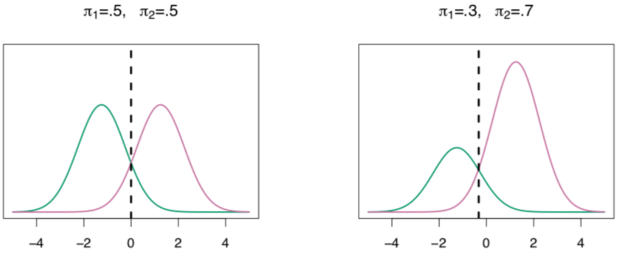


Figure 2: Fig 4.4

Fig. 4.4. Solid line: LDA decision boundary

Linear Discriminant Analysis

We also can estimate $\pi_{green}(x)$ and $\pi_{pink}(x)$ based on the proportion of training observations in each class.



We classify a new point according to which density is highest.

When the priors are different, we take them into account as well, and compare $\pi_k f_k(x)$. On the right, we favor the pink class — the decision boundary has shifted to the left.

Figure 3: Fig 4.4

Linear Discriminant Analysis

Substituting the probability density function into the equation for Baye's theorem, taking the log, and rearranging the terms, we can obtain the discriminant function (Equation 4.13):

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

We assign an observation to the class, k, for which $\delta_k(x)$ is greatest.

With 2 classes and $\pi_1 = \pi_2$, we obtain the decision boundary in Fig. 4.4!

LDA with multiple predictors

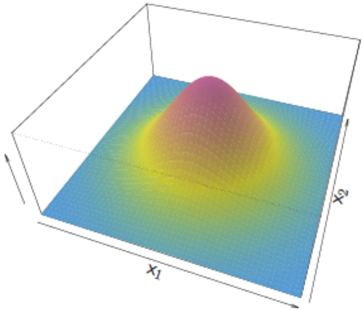


Figure 4: Fig 4.5a

Fig. 4.5: Multivariate density function with two **uncorrelated** predictors.

LDA with multiple predictors

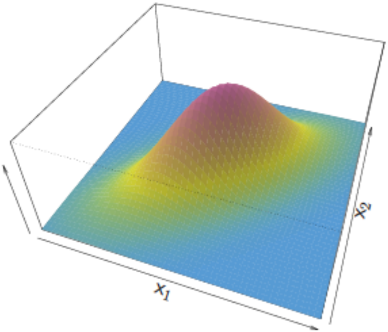


Figure 5: Fig 4.5b

Fig. 4.5: Multivariate density function with two **correlated** predictors ($r = 0.7$).

LDA with multiple predictors

Assuming 3 classes, consider the decision boundaries in Fig. 4.6:

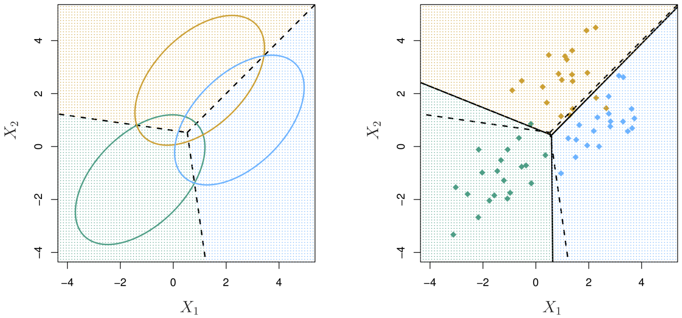


Figure 6: Fig 4.6

What is a plausible correlation coefficient for X_1 and X_2 for the green class? The blue class? The orange class?

LDA with multiple predictors

LDA assumes that the covariance structure among predictors is the same for all classes. ISL denotes the p x p covariance matrix of X as Σ.

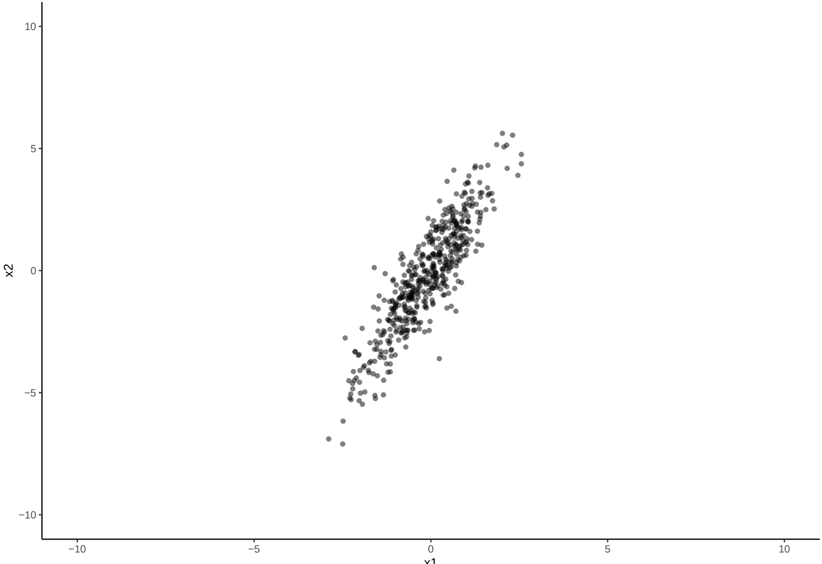
Note:

$$\mathrm{Cov}(X, Y) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N}$$

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

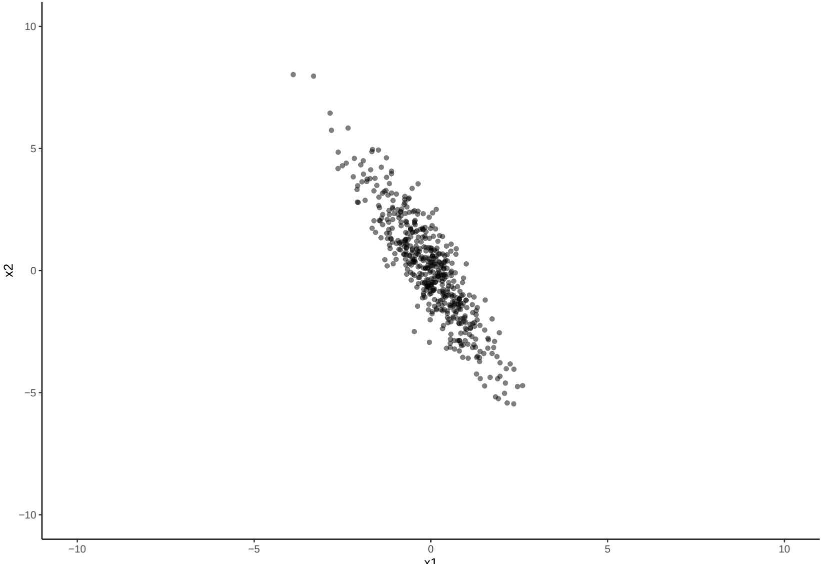
LDA with multiple predictors

$$\Sigma = \begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_1, x_2) & \mathrm{Var}(x_2) \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$



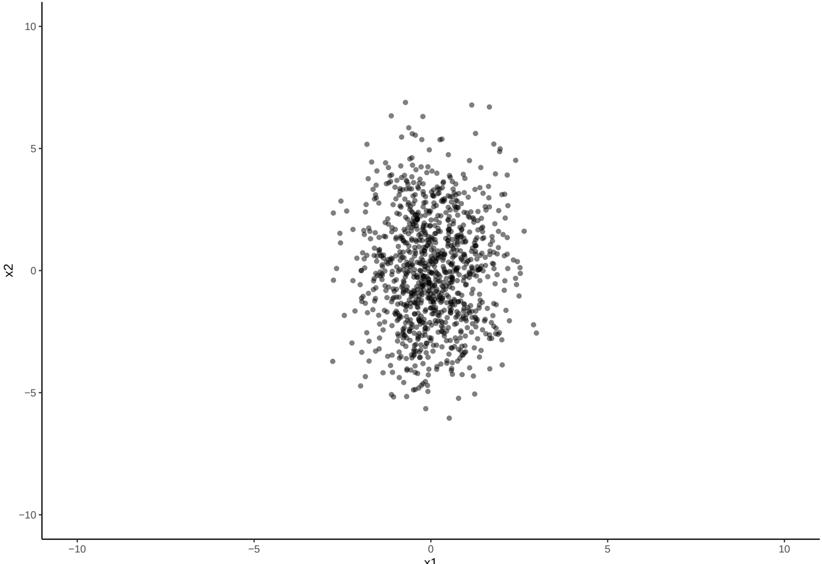
LDA with multiple predictors

$$\Sigma = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$$



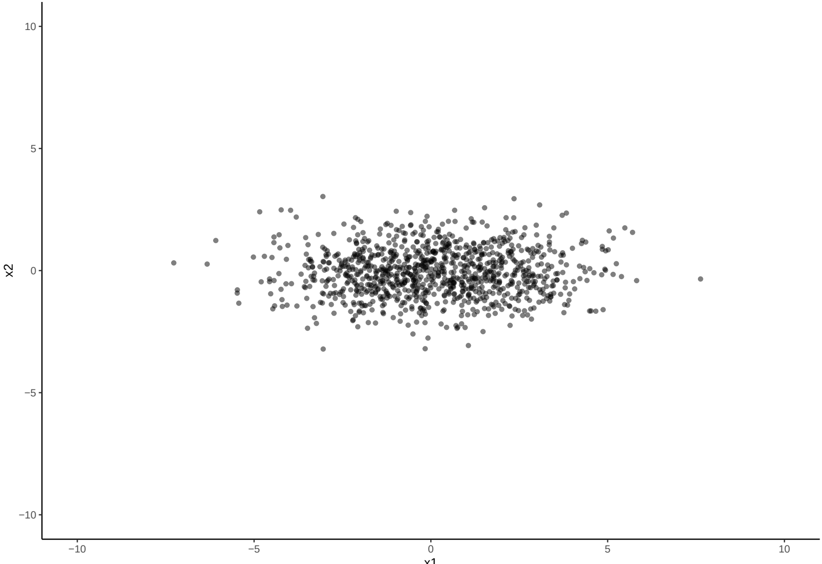
LDA with multiple predictors

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$



LDA with multiple predictors

$$\Sigma = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$$



LDA with multiple predictors

What if we have a different covariance structure among predictors for each class?

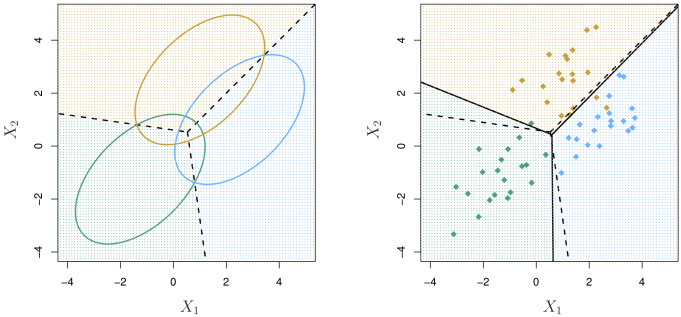


Figure 7: Fig 4.6

Quadratic Discriminant Analysis

- Unlike LDA, QDA assumes that each class has its own $p \times p$ covariance matrix.
- The resulting decision boundaries become quadratic
- Because we estimate covariances for each class separately, a larger n is needed compared to LDA
  - $p/(p+1)/2$ parameters for LDA covariance matrix vs. $Kp(p+1)/2$ for QDA.

Quadratic Discriminant Analysis

For which figure does $\Sigma_{orange} \neq \Sigma_{blue}$?

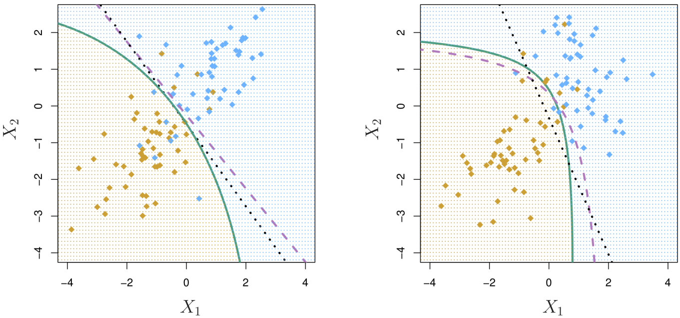


Figure 8: Fig 4.9

Fig. 4.9. Bayes (dashed purple), LDA (dotted black), and QDA (green solid) decision boundaries.