Homework 2
CS534 Machine Learning, Fall 2016

This homework explores linear classification methods.

Problem 1 - The multivariate normal (10 points)

Consider a dataset containing samples $X \in \mathbb{R}^p$ generated from two multivariate normals with means μ_1, μ_2 and equal covariance Σ. Suppose there are N_1 samples generated from class $\mathcal{N}(\mu_1, \Sigma)$, and N_2 samples generated from class $\mathcal{N}(\mu_2, \Sigma)$. Starting from $P(G = 2|X), P(G = 1|X)$, show that if a sample is more likely to have come from class $g = 2$ then

$$x^T\Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2}(\mu_2 + \mu_1)^T\Sigma^{-1}(\mu_2 - \mu_1) - log(N_2/N_1)$$

First, define the ratio where x is more likely from class $g = 2$

$$\frac{P(G = 2|X = x)}{P(G = 1|X = x)} > 1.$$

Substituting the definition of a multivariate normal

$$\frac{\frac{N_2}{N}((2\pi)^p|\Sigma|)^{-1/2}e^{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}}{\frac{N_1}{N}((2\pi)^p|\Sigma|)^{-1/2}e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}} > 1.$$

Cancel terms and moving the denominator right

$$N_2e^{-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)} > N_1e^{-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}.$$

Take the log

$$log(N_2) - \frac{1}{2}(x - \mu_2)^T\Sigma^{-1}(x - \mu_2) > log(N_1) - \frac{1}{2}(x - \mu_1)^T\Sigma^{-1}(x - \mu_1).$$

Expand terms in quadratic forms

$$log(N_2) + -\frac{1}{2}(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_2 - \mu_2^T\Sigma^{-1}x + \mu_2^T\Sigma^{-1}\mu_2)$$
$$> log(N_1) + -\frac{1}{2}(x^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu_1 - \mu_1^T\Sigma^{-1}x + \mu_1^T\Sigma^{-1}\mu_1).$$

Canceling terms not involving μ_1, μ_2, using symmetry of $x^T\Sigma^{-1}\mu, \mu^T\Sigma^{-1}x$

$$log(N_2) + x^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 > log(N_1) + x^T\Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1.$$

Gathering terms again

$$x^T\Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 - \frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + log(N_1) - log(N_2).$$

Factorizing the terms $\mu^T\Sigma^{-1}\mu$ into a quadratic

$$x^T\Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2}(\mu_2 + \mu_1)^T\Sigma^{-1}(\mu_2 - \mu_1) + log(N_1) - log(N_2).$$

Collecting the log

$$x^T\Sigma^{-1}(\mu_2 - \mu_1) > \frac{1}{2}(\mu_2 + \mu_1)^T\Sigma^{-1}(\mu_2 - \mu_1) - log(\frac{N_2}{N_1}).$$

Problem 2 - Linear discriminant analysis (10 points)

The dataset for this problem contains samples $X \in \mathbb{R}^2$ from two classes

$$X^{(1)} \sim \mathcal{N}(\begin{bmatrix}2\\2\end{bmatrix}, \begin{bmatrix}2 & -1\\-1 & 1\end{bmatrix}), \qquad X^{(2)} \sim \mathcal{N}(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0.5\\0.5 & 1\end{bmatrix}).$$

Generate these samples using your solution to Homework 1.2.

2.a. Linear discriminant
The linear discriminant function $\delta_k(x)$ is defined as

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

where π_k is the *prior probability* of class k.

Generate a scatter plot of training samples for $|\{X^{(k)}\}| = 10$, color coding class, and superimpose the following on this plot

- LDA decision boundary $\{x|\delta(x) = 0\}$
- Theoretical Bayes decision boundary: $\{x|f_{X^{(1)}} = f_{X^{(2)}}\}$
- Empirical Bayes decision boundary: $\{x|\hat{f}_{X^{(1)}} = \hat{f}_{X^{(2)}}\}$ (using sample mean, covariance from $|\{X^{(k)}\}| = 10$)

2.b. Linear discriminant - large sample
Repeat 2.a. with $|\{X^{(k)}\}| = 1000$.

To show decision boundaries in these problems, I used the contour plot and set the level to 0 to display level curves in each case. Note that as N increases, the Bayes decision boundaries converge as the sample estimates of μ_k, Σ_k approach the true values.

See plots on the following page.

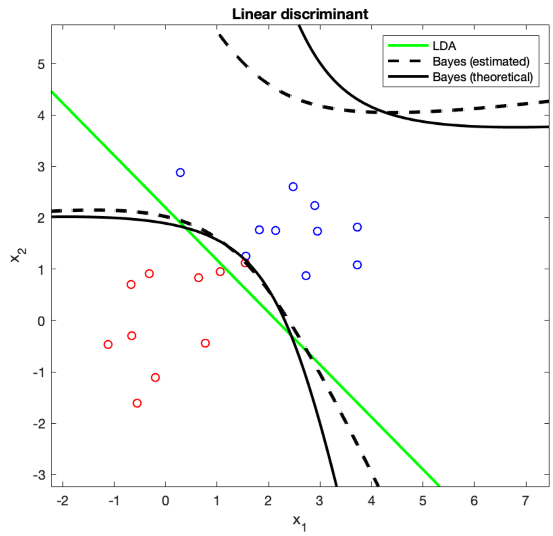

Figure 1: Problem 2.a.

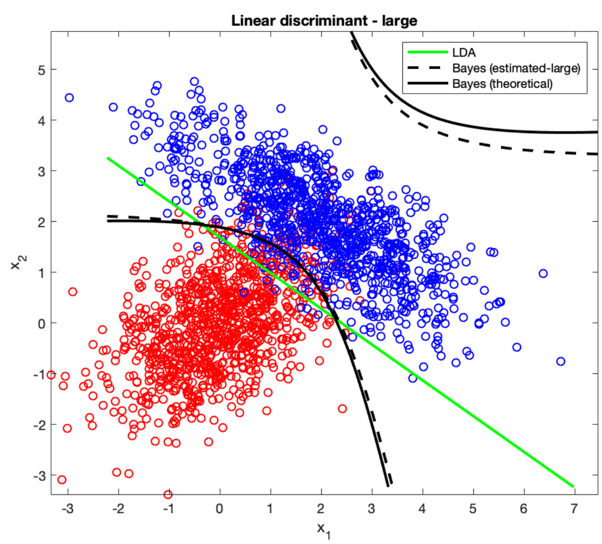

Figure 2: Problem 2.b.

Problem 3 - Quadratic discriminant analysis.

Repeat problem 2 using the quadratic discriminant

$$\delta_k(x) = -\frac{1}{2}\log|\hat{\Sigma}_k| - \frac{1}{2}(x - \hat{\mu}_k)^T \hat{\Sigma}_k^{-1}(x - \hat{\mu}_k) + \log\pi_k.$$

In this problem, the QDA boundaries conform very well to the empirical Bayes decision boundaries. As the training data increases, the QDA and empirical and theoretical Bayes boundaries all converge.

See plots on the following page.

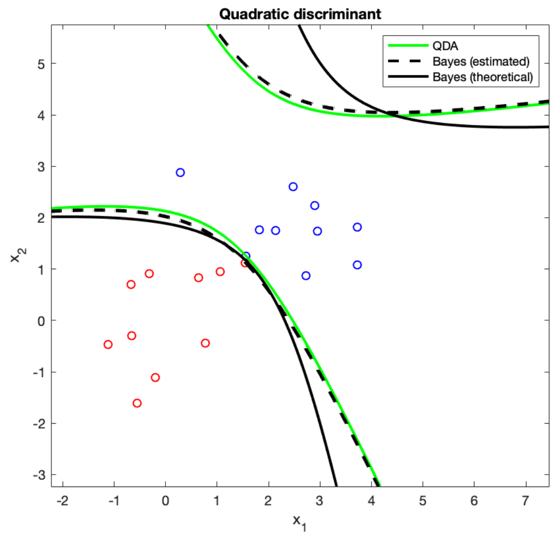

Figure 3: Problem 3.a.

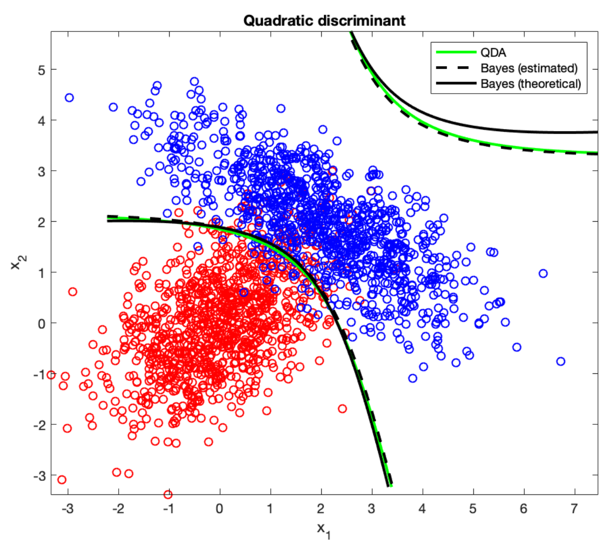

Figure 4: Problem 3.b.