

CS 475 Machine Learning: Homework 2  
 Supervised Classifiers 1,  
 Probability, Linear Algebra and Decision Trees  
 Due: Tuesday September 29, 2015, 11:59pm  
 100 Points Total      Version 1.0

## 1 Analytical (50 points)

### 1) Fisher Linear Discriminant and Logistic Regression Classifiers (15 points)

- (a) Prove that the class label  $y$  conditioning on the feature vector  $X$  follows a logistic regression model.

**Ans:**

We first use Bayes rule on  $P(y|X)$ :

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Since the two classes has equal prior, we have  $P(y) = 1/2$  for each  $y$ . In addition,  $P(X)$  can be represented as  $\sum_y P(X|y)P(y)$ . So, the original equation can be changed as below:

$$\begin{aligned}
 P(y|X) &= \frac{P(X|y)P(y)}{\sum_y P(X|y)P(y)} \\
 &= \frac{\frac{1}{\sqrt{(2\pi)^k \Sigma}} e^{-\frac{1}{2}(X-\mu_1)^T \Sigma^{-1}(X-\mu_1)} \times \frac{1}{2}}{\frac{1}{\sqrt{(2\pi)^k \Sigma}} e^{-\frac{1}{2}(X-\mu_1)^T \Sigma^{-1}(X-\mu_1)} \times \frac{1}{2} + \frac{1}{\sqrt{(2\pi)^k \Sigma}} e^{-\frac{1}{2}(X-\mu_2)^T \Sigma^{-1}(X-\mu_2)} \times \frac{1}{2}} \\
 &= \frac{e^{-\frac{1}{2}(X-\mu_1)^T \Sigma^{-1}(X-\mu_1)}}{e^{-\frac{1}{2}(X-\mu_1)^T \Sigma^{-1}(X-\mu_1)} + e^{-\frac{1}{2}(X-\mu_2)^T \Sigma^{-1}(X-\mu_2)}} \\
 &= \frac{1}{1 + e^{-\frac{1}{2}(X-\mu_2)^T \Sigma^{-1}(X-\mu_2) + \frac{1}{2}(X-\mu_1)^T \Sigma^{-1}(X-\mu_1)}} \\
 &= \frac{1}{1 + e^{-\frac{1}{2}(X^T \Sigma^{-1} X - X^T \Sigma^{-1} \mu_2 - \mu_2^T \Sigma^{-1} X + \mu_2^T \Sigma^{-1} \mu_2 - X^T \Sigma^{-1} X + X^T \Sigma^{-1} \mu_1 + \mu_1^T \Sigma^{-1} X - \mu_1^T \Sigma^{-1} \mu_1)}}
 \end{aligned}$$

Since  $X^T \Sigma^{-1} \mu_1$  and  $X^T \Sigma^{-1} \mu_2$  are numbers, we can transpose it. ( $\Sigma$  is a symmetric matrix)

$$\begin{aligned}
 (X^T \Sigma^{-1} \mu_1)^T &= \mu_1^T (\Sigma^{-1})^T X = \mu_1^T (\Sigma^T)^{-1} X = \mu_1^T \Sigma^{-1} X \\
 (X^T \Sigma^{-1} \mu_2)^T &= \mu_2^T (\Sigma^{-1})^T X = \mu_2^T (\Sigma^T)^{-1} X = \mu_2^T \Sigma^{-1} X
 \end{aligned}$$

So the equation becomes:

$$= \frac{1}{1 + e^{-\frac{1}{2}(-2\mu_2^T \Sigma^{-1} X + 2\mu_1^T \Sigma^{-1} X + \mu_2^T \Sigma^{-1} \mu_2 - \mu_1^T \Sigma^{-1} \mu_1)}}$$

$$= \frac{1}{1 + e^{-(\mu_1^T - \mu_2^T) \Sigma^{-1} X - \frac{1}{2}(\mu_2^T \Sigma^{-1} \mu_2 - \mu_1^T \Sigma^{-1} \mu_1)}}$$

The equation shown above is also the logistic regression form. Thus, we proved that the class label  $y$  conditioning on the feature vector  $X$  follows a logistic regression model.

- (b) Prove that the classifier based on the logistic regression model obtained in (a) is equivalent to the optimal Fisher linear discriminant classifier. The optimal Fisher linear discriminant classifier is obtained using the population means and covariance matrix; see section 4.1.4 in Bishop.

**Hint:** You only need to show that both classifiers use the same decision rule.

## 2) Linear Models (10 points)

## 3) Regularization and Overfitting. (5 points)

- 4) **Decision Tree (10 points)** Let's investigate how accurately decision trees can learn. We start

## 5) Conjugate Prior (10 points)