

## Homework Set 6

### Problem 1 (LDA)

Consider the Fisher's linear discriminant analysis method.

Sample	$y_1$			$y_2$		
	$x_1$	$x_2$	$x_3$	$x_1$	$x_2$	$x_3$
1	-0.4	0.58	0.089	0.83	1.6	-0.014
2	-0.31	0.27	-0.04	1.1	1.6	0.48
3	0.38	0.055	-0.035	-0.44	-0.41	0.32
4	-0.15	0.53	0.011	0.047	-0.45	1.4
5	-0.35	0.47	0.034	0.28	0.35	3.1
6	0.17	0.69	0.1	-0.39	-0.48	0.11
7	-0.011	0.55	-0.18	0.34	-0.079	0.14
8	-0.27	0.61	0.12	-0.3	-0.22	2.2
9	-0.065	0.49	0.0012	1.1	1.2	-0.46
10	-0.12	0.054	-0.063	0.18	-0.11	-0.49

- Write a general program to calculate the optimal direction  $\mathbf{v}$  for a linear discriminant analysis based on three-dimensional data.
- Find the optimal  $\mathbf{v}$  for the data in the table above.
- Plot a line representing your optimal direction  $\mathbf{v}$ . Mark on the line the positions of the projected points.
- Fit each distribution with a (univariate) Gaussian, and find the resulting decision boundary.
- What is the training error in the optimal subspace you found in (b)?

**Problem 2 (PCA and LDA)**

In this problem, you will implement principal component analysis (PCA) for dimension reduction, and use your PCA implementation to implement linear discriminant analysis (LDA). You will apply these techniques, and investigate and interpret the results you obtain.

- a) Perform PCA on the unlabeled points from the Fisher' Iris flower data set provided on ceiba. The data should be normalized before running the PCA. You may also load the data by typing the following command if using MATLAB:

```
>>load iris.mat
```

- 1) List the principal components explaining 95% of the total variance in the dataset.
  - 2) Plot the data points using the first two PCs as axes, distinguishing between the classes using different color or marker.
- b) Split the Fisher' Iris flower data set into a training and test set - use the first 30 points from each class for training and the last 20 points from each class for testing. Perform LDA on the training set.
  - 1) Perform LDA on the original explanatory variables. The dataset should be normalized before performing the LDA. Plot the training and test sets using the first two linear discriminant axes. These can be on separate plots. Calculate the training and test errors.
  - 2) Perform LDA on the PCs obtained above. Plot the training and test set using the first two linear discriminant axes. These can be on separate plots. Calculate the training and test error.
  - 3) Compare the results of 1) and 2). Explain the discrepancy.