Fall 2014

# ECE471/571 – Homework #2 (Due 09/22/14)

Note: UG: 100+10, G: 100

Data set used in this homework:

| X | Y | Class |
|---|---|---|
| 0.8 | 1.2 | 1 |
| 0.9 | 1.4 | 1 |
| 1.2 | 1.4 | 1 |
| 1.1 | 1.5 | 1 |
| 0.8 | 1.1 | 2 |
| 0.6 | 1 | 2 |
| 0.65 | 1.1 | 2 |
| 0.75 | 0.9 | 2 |

1) (30/20) Mahalanobis distance vs. Euclidean distance.
    a. (5/5) Show the equations to calculate these two distances.
    b. (5/5) Explain intuitively (in no more than three sentences) the differences between the two distances.
    c. (20/10) Use the following example to understand the differences these two distances make in classification. Here, the minimum distance classifier is used.
        i. Plot the above data set on the same figure.
        ii. Given a test sample $x = [0.85\ 1.15]^T$, calculate the Euclidean distance to the two classes. Based on the distances, which class should x belong to?
        iii. Use the same test sample, calculate the Mahalanobis distance to the two classes. Based on this pair of distances, which class should x belong to?
        iv. Plot the test sample x on the same figure as the data set. Just by observing the plot, which decision do you think makes more sense?

2) (70/70) Dimensionality reduction with a two-feature two-class data set.

   a. (10/10) Preprocessing steps (Need to show step-by-step details):

      i. Calculate the mean of each class (m1, m2) manually.

      ii. Calculate the covariance matrix of each class ($\Sigma_1$, $\Sigma_2$) manually.

   b. (20/20) Using Fisher's Linear Discriminant (FLD) to find the vector (w) which optimally (in the Fisher sense) separates the projections of these two classes.

   c. (10/10) Is the vector derived from FLD along the same direction as (m1-m2)? Plot m1-m2 and w on the same figure. Also explain it analytically.

   d. (20/20) Using principal component analysis (PCA) to reduce the dimension and compare it with the one derived from FLD. (You can use existing functions to find the eigenvectors and eigenvalues.)

   e. (10/10) Comment on the differences between FLD and PCA.

3) (+10/10) Using maximum likelihood method to derive the equation for mean and variance assuming the pdf (or likelihood) is modeled by 1-D Gaussian.