The University of Texas at Austin
Department of Electrical and Computer Engineering

**EE379K: Data Science Lab — Fall 2017**

Lab Five

Caramanis/Dimakis Due: Monday Oct 9th, 3:00pm 2017.

---

**Problem 1: Linear Discriminant Analysis**.

1. Generate 20 random points in $d = 3$, from a Gaussian multivariate distribution with mean $[0, 0, 0]$ and covariance matrix $\begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 0.9 \\ 0.9 & 0.9 & 1 \end{pmatrix}$. Let's call this data with label 1. Also generate 20 random points in $d = 3$ from another Gaussian with mean $[0, 0, 1]$ and covariance $\begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$. Let's call that data with label 2. Create a three dimensional plot of the clouds of data points, labeled with the two labels.

2. Perform a projection of the data on one dimension using Fischer's Linear Discriminant as explained in class (see also `http://research.cs.tamu.edu/prism/lectures/pr/pr_l10.pdf`). (no sklearn Linear Discriminant Analysis functions here, just friendly linear algebra.)
3. Use sklearn to perform Linear Discriminant Analysis. Compare the results.

For the remaining problems, we will be using the ISL Book (`http://www-bcf.usc.edu/~gareth/ISL/`). Chapters 1-4 from ISL is material we have already covered in detail. Read Chapters 5 and 6. The needed data sets are available from the book website. Note that we have uploaded Python versions of the end-of-chapter labs from Chapters 3, 5, 6 and 7 of the ISL book. Feel free to look through these. They are generally well done and useful.

**Problem 2**. Problem 10 from Chapter 4 (Classification for the 'weekly' dataset) Read the chapter to figure out how to use KNN and QDA, as we did not discuss these in class.

**Problem 3**: Problem 5 from Chapter 5 Estimating the test error of logistic regression using a validation set.

**Problem 4**: Problem 8 from Chapter 5.(Cross Validation). Use Google or any other means to find the Python counterpart to R's `rnorm`.)

**Problem O1**. (Optional) Problem 10 from Chapter 3.

**Problem O2**. (Optional) Problem 15 from Chapter 3.

**Problem O3**. (Optional) Problem 13 from Chapter 4.