**ECE 8527: Introduction to
Machine Learning and Pattern Recognition**

# HW No. 4: CLASS-DEPENDENT VS. CLASS INDEPENDENT GAUSSIAN MODELS

For this assignment, you will generate two multivariate 2D Gaussian distributions. The first class will have a mean of $\left[\begin{matrix}0&0\end{matrix}\right]$ and a covariance matrix of $\left[\begin{matrix}2&0\\0&1\end{matrix}\right]$. The second class will have a mean of $\left[\begin{matrix}1&0\end{matrix}\right]$ and a covariance matrix of $\left[\begin{matrix}1&0\\0&2\end{matrix}\right]$. Generate 10,000 training data points (2D vectors) and 10,000 evaluation data points for each class (the total amount of training data is 20,000 points and the total amount of evaluation data is 20,000 points).

The tasks to be accomplished in this homework assignment are:

1. Plot a scatter plot of the data showing the first class in blue and the second class in red.
2. Pool both classes in the training set into one set of training data. Estimate the mean and covariance. Generate a transformation matrix using Principal Components Analysis (PCA). Transform the data using this matrix and plot a scatter plot. What do you observe?
3. Classify the data using a simple classifier based on the halfway point between the two means. What is the error rate? Draw the decision surface on the scatter plot.
4. Next, do a traditional class-dependent analysis and classify the data using a maximum likelihood classifier, as you did in HW #3. Assume the priors are equal and make sure you plug in each class’ covariance matrix. What is the error rate?
5. For step 4, draw a scatter plot of the transformed data and show the decision surface. Compare and contrast this to the findings of steps 1-3.

What did you learn from this analysis? If you rotated the scatter plots by a 45° angle would the results change? Can you prove this?