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

# HW No. 1: Gausssian Distributions and MAXIMUM LIKELIHOOD DECODING

For this assignment, you will use the data set located here:

*https://www.isip.piconepress.com/courses/temple/ece\_8527/resources/data/set\_04/*

We will focus on the files *train.txt* and *dev.txt*, which containing training and development test set data.

The tasks to be accomplished in this homework assignment are:

1. Select the two dimensions of the data that have the greatest variance and do a scatter plot of each class. Estimate the mean and covariance for each class. Write an expression for a Gaussian distribution that models the data in each class. Superimpose this model on your scatter plot (hint: using support regions).
2. Repeat this for the two dimensions that have the greatest overlap. Smallest var greatest overlap.

Consider classification of the data using:

1. The feature with the largest covariance.
2. The two features with the largest covariance.
3. The feature with the smallest covariance.
4. Two features: the feature with the largest covariance and the feature with the smallest covariance.

Comment on your ability to classify the data using these features.

# Discussion:

Classification Performance from best to worse: 2 -> 4 -> 1 -> 3

A classification with a greater variance means more info, thus more distinguishable. Based on the observation of the data set. A larger covariance occurs with a greater variance. So, a higher covariance means a higher variance, thus more feature and more distinguishable.

Number 2 has two features with greatest covariance, which means that it has two very distinguishable features. 1 and 4 both have a single set with large covariance but with 4 has one additional set of features, even if it is no very distinguishable due to a small covariance, the additional set will contribute to 4 thus result in a little better performance. 3 has the smallest covariance which result in a very poor performance.

# Generated Pictures:



Figure Overlap with Two Greatest Variance Sets

Based on the result generated by the code, the two feature sets that has the greatest variance are features set 0 and features set 8. The graph above shows the scatter plot of those two feature sets. The contour graph is also generated based on each class of each feature with a normal multivariant gaussian distribution in a support region. The overlap of these two plots shows the similarity of these two features set.

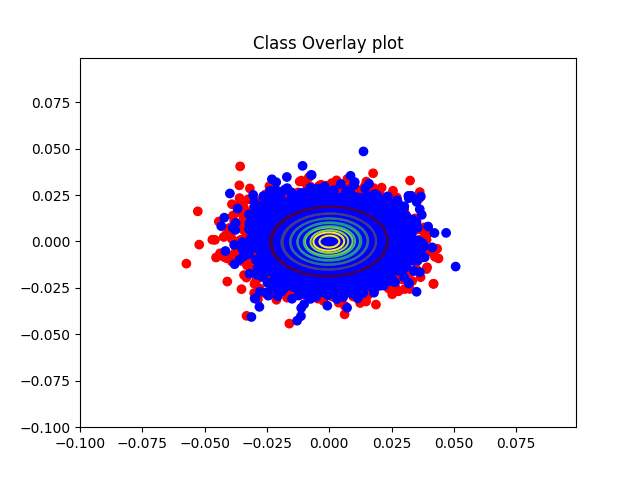


Figure Overlap with Two Smallest Sets

Based on the observation of different graphs generated by a same but modified code from the previous question, the contour graph of feature set 24 and feature set 25 shows the most overlap. By observe the information of these two feature sets, such as the variance and covariance, a conclusion is drawn due to these two feature sets has the smallest variance and covariance among all 26 feature sets that feature sets with the smallest variant and covariant will show the most overlap.