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

# HW No. 2: Bayesian DEcision Theory

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

*https://www.isip.piconepress.com/courses/temple/ece\_8527/resources/data/set\_13/*

We will focus on the files *train.csv* and *eval.csv*, which contain training and evaluation data. You can assume the loss function weighs all errors equally.

We will also focus on comparing the performance for three tools: IMLD, SKLearn in Python and JMP. Download and install IMLD following the instructions given in class.

The tasks to be accomplished in this homework assignment are:

1. Generate multivariate Gaussian distributions with a mean vector of zero and covariances matrices as follows:

Generate scatter plots of each data set, compute the eigenvalues and eigenvectors, and plot the eigenvectors over the data. Discuss why these plots make sense and what you learned about the relationship between covariance and eigenvector analysis.

1. Load the data from set\_13 into IMLD and classify the data using the algorithm named quadratic discriminant analysis (QDA). Report **the errors rates** in a table using the template below.
2. Repeat this for QDA in JMP, [QDA in SKLearn](https://scikit-learn.org/0.16/modules/generated/sklearn.qda.QDA.html), and your custom implementation of a simple Gaussian classifier, based on what we discussed in class (e.g., in Python or MATLAB). Assume the priors are equal. Discuss why these results are similar but not identical.
3. For your Python code, assume the priors are not equal. Compute and plot the error rate as you vary the prior of class “dog” over the range [0,1] (the prior for class “cat” is obviously . Write a loop that samples the prior in steps of 0.01 over the range [0,1] and plot the result. Do the same for JMP and compare the two plots. Do this only for the evaluation data.

Results should be reported using this table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Algorithm** | **Data** | **IMLD** | **JMP** | **SKLearn** | **Python** |
| **Train** | **Eval** | **Train** | **Train** | **Eval** | **Train** | **Train** | **Eval** |
| QDA | Set No. 13 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 | 99.99 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Be sure you use two decimal points of precision in your table. This is very important from a statistical significance point of view, which we will discuss later in the class.

Your plot of performance as a function of the priors should show the prior probability on a scale of [0,1] on the horizontal axis, and the error rate as a decimal in the range [0,1] on the vertical. Explain why this plot makes sense.

Include your Python (or MATLAB) code in your document. Explain what your code is doing step by step.