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

# HW No. 1: Gausssian Distributions

For all the homework assignments in this class, use this template to turn in your work:

*https://isip.piconepress.com/courses/temple/ece\_8527/resources/templates/homework\_v00.docx*

Submit a pdf file using the “Save As Reduced File Size” feature of Adobe Acrobat.

**Task 1:** Using the ISIP Machine Learning Demo (IMLD) located here:

*https://www.isip.piconepress.com/courses/temple/ece\_8527/resources/imld*

Generate a two-class data set where each class has a Gaussian distribution with an identity covariance matrix. The first class should have a mean of $[1,1]$ and the second class should have a mean of $[-1,-1]$. Generate 10,000 data points in each class.

Next, write a Python or Matlab script (use whatever programming language you feel comfortable in) that reads this data and displays it in a scatter plot similar to the IMLD plot. We will need this kind of visualization tool throughout the course.

Display your plots side-by-side using the same scales so I can easily compare them.

**Task 2:** In the programming language of your choice, generate 2D Gaussian data in the same format used by IMLD. Generate the same data described in Task 1. Load it into IMLD and show that the IMLD display of the data matches your personal code. Display your plots side-by-side using the same scales so I can easily compare them.

**Task 3:** Using your personal code, and comparing it to IMLD, generate Gaussian data that matches the shapes shown in lecture no. 2. Adjust the covariance matrix so that you achieve the shapes (e.g., support regions) shown in the review slides. Use IMLD, which does scatter plots, to verify your data. Use 10,000 data points as in the previous tasks.

There are four cases to be considered: (1) the covariance matrix is an identity matrix; (2) the covariance matrix is diagonal by the variances are unequal; (3) the covariance matrix is not a diagonal matrix (but still symmetric) but has equal variances, and (4) an unconstrained matrix (aside from being symmetric).

Make sure you understand how the values in the covariance matrix influence the shapes of the support regions. That is the most important goal of this assignment.

**Summary:** We use Gaussian distributions and similar simple test data to debug and evaluate algorithms. This assignment attempts to build your comfort level with how Gaussian distributions work and how to visualize these distributions.