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

# HW No. 4: ML vs. Bayesian Estimation

The tasks to be accomplished in this homework assignment are:

1. Generate $10$ independent sets of random data consisting of $10^{6}$ points from a 1D GRV with a mean of $1$ and a variance of $1$ (this is a Gaussian in which the argument, or feature vector values, is a scalar).
2. For the first of these sets, estimate the mean value using a maximum likelihood estimate. Plot the estimated mean for the range $N = [1, 10^{6}]$. Use a $log\_{10}$ scale for the horizontal axis (the number of points).
3. Repeat no. 2 but this time for each $N$, compute the average of each of the first $N$ points in the $10$ independent sets, and then compute the average of those averages. For example, for $N=10$, compute the average of the first $10$ points in each set, then average the averages. Plot this on the same plot as no. 1. Explain any differences you see in these two plots (e.g., analyze the results and draw some conclusions as to what aspects of the two algorithms influences the results).

Note that in this case, for $N=10$, you are using $10x10=100$ points, to compute the average. How does that compare to $N=100$ for the plot of problem no. 1?

1. Now assume an initial mean guess of 2. Construct the Bayesian estimate of the mean assuming both the mean and variance are unknown, but the data obeys a Gaussian distribution (which it does of course). Plot the error as a function of $N$. Explain how this plot compares to the previous two plots.

In this assignment, you are comparing a traditional ML estimate of the mean to a Bayesian estimate of the mean with a really bad initial guess. How many data points does it take to overcome this bad initial guess? At what number of data points are the two estimates equivalent?