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

# HW No. 5: Nonparametric Models

1. Replicate the demonstration of a Parzen window shown in lecture 13, slide 6.

2. For the remaining problems, we will focus on a data set that consists of four overlapping Gaussian distributions. Generate four 2D Gaussian GRVs with identity covariance matrices and means of [*0.25*, *0.25*], [*-0.25*, *0.25*], [*-0.25*,*-0.25*] and [*0.25*, *-0.25*]. Generate *10,000* vectors per class to use as training data, and generate an additional 1,000 vectors for testing.

3. K-Nearest Neighbor (kNN): Train a *k*-nearest neighbor classifier for each class using *M* neighbors (giving a total of *M* x *4* vectors). Classify the data by choosing the class that corresponds to the vector closest to the test vector. Plot the probability of error for the training and test data as a function of *k*. Justify your results.

4. Gaussian Mixture Model (GMM): Repeat problem 3 using a GMM. Plot the probability of error as a function of the number of mixture components per class. Explain your result.

5. Support Vector Machines (SVM): Repeat problem 3 using an SVM model. Plot the probability of error as a function of the number of support vectors and the kernel function.

6. Generate your favorite complex 2D data set (e.g., concentric toroids, yin-yang, overlapping elliptical Gaussians). Compare performance of kNN, GMM and SVM on this data from both an error rate and computational complexity point of view.

You are free to use whatever toolboxes are available in MATLAB. You should not have to code any of these algorithms.