EE 559 **Homework 7 (Week 14)** Posted: Mon., 4/18/2022
Jenkins Due: Mon., 4/25/2022, 11:59 PM PDT

1. (a) Two classes are described by normal densities as follows

$$p(\underline{x}|S_i) = N\left(\underline{x}, \underline{m}_i, \underline{\underline{\Sigma}}_i\right), i = 1,2$$

$$P(S_1) = P(S_2) = 0.5$$

$$\mathrm{m}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \ \mathrm{m}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2 = \begin{bmatrix} 2 & 3 \\ 3 & 6.5 \end{bmatrix}$$

Solve, algebraically, for the Bayes minimum error classifier; *i.e.*, give the resulting decision rule algebraically, in simplest form.

Plot (by hand or by computer) the decision boundary and label the decision regions in 2D nonaugmented feature space.

(b) Repeat part (a) with the same numbers except $P(S_1) = 0.1$, $P(S_2) = 0.9$.

2. This problem can be done by hand. If prefer you may use a computer to help with the plots, as long as you know how to plot them by hand.

In this problem, use Kernel Density Estimation (KDE) with binary kernel (window) functions given (in unnormalized form) by:

$$\Phi\left(\frac{\mathrm{u}}{\mathrm{h}}\right) = \left[\left[-1 \le \frac{\mathrm{u}}{\mathrm{h}} < 1\right]\right]$$

in which $\left[[X]\right]$ denotes indicator function of *X*.

You are given the following prototypes for a 2-class problem in 1D feature space:

$$S_1\text{: } 0,\ 0.4,\ 0.9,\ 1.0,\ 6.0,\ 8.0$$

$$S_2\text{: } 2.0,\ 4.0,\ 4.5,\ 5.0,\ 5.8,\ 6.7,\ 7.0$$

For parts (a)-(c) below, choose $h = 0.5$ so that the windows have width of 1.

(a) Graph the KDE estimates of the density functions $p(x|S_1)$ and $p(x|S_2)$. Be sure to label pertinent values on both axes.

(b) Estimate the prior probabilities based on frequency of occurrence of the prototypes in each class.

(c) Use the estimates you have developed in (a)-(b) above to find the decision boundaries and regions for a Bayes minimum-error classifier based on KDE. Only the part of feature space where at least one density is nonzero need be classified.

(d) Repeat (a)-(c) for $\mathrm{h} = 1$.

(e) Repeat (a)-(c) for $\mathrm{h} = 2$.

3. This problem can be solved by hand (using a calculator). However, if you prefer, you may use a computer for all or part of your solution, as long as you know how to solve it by hand.

For the following data in 1D feature space (listed as $(x_i, y_i),\ i = 1, N$):

(-.9, .81), (-.7, .49), (-.5, .25), (-.3, .09), (-.1, .01), (.1, .01), (.3, .09), (.5, .25), (.7, .49), (.9, .81):

(a) Use an (unweighted) *kNN* regressor with $k = 4$ to estimate $\hat{y}(x)$ at the following points:

$x = 0,\ x = .4$

(b) You are now given that the true target function is in fact $y(x) = x^2$. Calculate the MSE on the "test set" predictions of part (a). (Use $y(x)$ to compute its correct outputs.)

(c) Repeat parts (a)-(b) except using *kernel regression* with $k = 4$. For the kernel function, use:

$$\mathrm{k}(\underline{\mathrm{x}} - \underline{\mathrm{x}}_\mathrm{n}) = 1 - \frac{\mathrm{d}(\underline{\mathrm{x}}, \underline{\mathrm{x}}_\mathrm{n})}{\mathrm{d}_\mathrm{max}}$$

in which d_{max} = distance to the 5th nearest neighbor, and $d(\underline{x}, \underline{x}_n)$ denotes Euclidean distance.