1. In a 2-class classification problem, the class conditional densities are:

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

(a) This part uses a computer. Given the following class means and covariance matrices for a 2D problem:

$$\underline{m}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \underline{m}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$\underline{\underline{\Sigma}}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}, \quad \underline{\underline{\Sigma}}_2 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$$

On a single 2D plot, plot the 2 class means, and make a filled-contour plot of constant Mahalanobis distances from the mean for each class, as follows: $d_M^2\left(\underline{x}, \underline{m}_1\right) = B^2$ and $d_M^2\left(\underline{x}, \underline{m}_2\right) = B^2$, for values $B = 0.5, 1.0, 1.5, 2.0$. (End result should be 4 curves for $d_M^2\left(\underline{x}, \underline{m}_1\right) = B^2$ and 4 curves for $d_M^2\left(\underline{x}, \underline{m}_2\right) = B^2$, with colored fills between the curves.)

Tips:

(1) Use `matplotlib.pyplot.contourf` for the filled-contour plot; use the `levels` parameter to input the values of B^2; be mindful of B vs. B^2.

(2) Here is one way to correctly plot the figure. You can calculate the value of the multivariate Gaussian by using $d_m^2 = B^2$; for each value of B^2, the corresponding output value of the multivariate Gaussian gives you a level value. Then, you can concatenate all level values together and utilize such values to assign the `levels` parameter in plt.contourf().

See https://www.tutorialspoint.com/matplotlib/matplotlib_contour_plot.htm and https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.contourf.html for more info.

(b) This part is to be done by hand. Suppose we implement a Bayes classifier for minimum error, with decision rule:

$$\ln\left[p\left(\underline{x} \,\middle|\, S_1\right)P(S_1)\right] > \ln\left[p\left(\underline{x} \,\middle|\, S_2\right)P(S_2)\right] \;\Rightarrow\; \underline{x} \in \Gamma_1$$

$$\ln\left[p\left(\underline{x} \,\middle|\, S_1\right)P(S_1)\right] < \ln\left[p\left(\underline{x} \,\middle|\, S_2\right)P(S_2)\right] \;\Rightarrow\; \underline{x} \in \Gamma_2$$

Give an expression for the decision boundary for the normal densities given at the beginning of this problem, by plugging in for $N\left(\underline{x} \,\middle|\, \underline{m}_i, \underline{\underline{\Sigma}}_i\right), \; i = 1,2$. Give your answer in terms of $d_M\left(\underline{x}, \underline{m}_i\right)$, $\underline{\underline{\Sigma}}_i$, $P(S_i)$, $i = 1,2$.

(c) This part is done by computer. For the means and covariance matrices given in (a), use your answer to (b) to plot the decision regions and boundaries of the Bayes classifier for minimum error. Also show the class means on the plot. Do this for 3 cases (1 plot for each case):

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

(ii) $\mathrm{P}(\mathrm{S}_1) = 0.3,\ \mathrm{P}(\mathrm{S}_2) = 0.7$

(iii) $P(S_1) = 0.1,\ P(S_2) = 0.9$

Tip: you can use `contourf` for this as well, or another routine of your choice.

(d) This part is done by computer. Plot the decision regions and boundaries of the Bayes classifier for minimum error, for the means given in (a), the priors given in (c)(i), and for the following covariance matrices. Also show the class means on the plot. (One plot for each part below; 4 plots total.) Note that $\underline{\underline{\Sigma}}_1$ and $\underline{\underline{\Sigma}}_2$ are given in part (a) above.

(i) $\underline{\underline{\Sigma}}_{1(i)} = 4\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix},\ \underline{\underline{\Sigma}}_{2(i)} = \underline{\underline{\Sigma}}_2.$

(ii) $\underline{\underline{\Sigma}}_{1(ii)} = 16\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix},\ \underline{\underline{\Sigma}}_{2(ii)} = \underline{\underline{\Sigma}}_2$

(iii) $\underline{\underline{\Sigma}}_{1(iii)} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},\ \underline{\underline{\Sigma}}_{2(iii)} = \underline{\underline{\Sigma}}_2$

(iv) $\underline{\underline{\Sigma}}_{1(iv)} = \underline{\underline{\Sigma}}_1\ ,\quad \underline{\underline{\Sigma}}_{2(iv)} = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$

(e) This part is done by observing and thinking. Comment on the effect of:

(i) changing the determinant of $\underline{\underline{\Sigma}}$ for one class (comparing (c)(i), (d)(i), (d)(ii)).

(ii) changing the entries of $\underline{\underline{\Sigma}}$ for one class relative to each other (comparing (c)(i) to (d)(iii))

(iii) the two covariance matrices being equal, as in (d)(iv).

2. This problem is to be done by hand. You may optionally use a computer to do the plots in (a)-(c); however, make sure you know how to do them by hand (e.g., on the final exam you will have to do everything by hand).

We will use density estimation on a 2-class classification problem.

For parts (a)-(d) of this problem, we will use a generative approach. Use Kernel Density Estimation (KDE) with binary kernel (window) functions given (in unnormalized form) by:

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

in which $[[X]]$ 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 $h = 1$ so that the windows have width of 2.

(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 data points 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.

Tips: (1) In the case of a tie, use the following ad-hoc rule: assign x to the class with the higher prior. (2) Only the part of feature space where at least one density is nonzero need be classified.

(d) Classify the points 3, 6.2, 8.5.

For parts (e)-(f) below, we will use a discriminative approach. Now consider kNN.

(e) For a k-nearest-neighbor classifier, classify the same points as in (d), for each of the following values of k. Thus, you will make a total of 9 classifications (3 classifications for each part below).

Tip: In case of a tie (e.g., 2 data points are equidistant from the point to classify, and you have to pick just one of the data points), pick the data point whose class label has the larger prior.

(i) $k = 1$

(ii) $k = 3$

(iii) $k = 5$

(f) For $k = 1$, draw a plot of the data points on the axis; also show the decision boundary and decision regions.