1. In this problem, consider a Bayes minimum-error classifier. Three classes are described by normal densities as follows:

$$p(\underline{x}|S_i) = N(\underline{x}, \underline{m}_i, \underline{\underline{\Sigma}}_i), \quad i = 1, 2, 3$$

$$P(S_1) = \pi_1, P(S_2) = \pi_2, P(S_3) = \pi_3$$

in which each π_i is a variable denoting the prior for class S_i (to save on writing). Assume the π_i are given.

In this problem, the algebra should be done by hand. The plots may be done using computer or by hand.

(a) Write an expression for the discriminant functions $g_i(\underline{x})$, in two forms:

  (i) Expressed in terms of Mahalanobis distance $d_M(\underline{x}, \underline{m}_i)$ and other given terms;

  (ii) Expressed only in terms of $\underline{x}, \underline{m}_i, \underline{\underline{\Sigma}}_i$, and $\underline{\underline{\Sigma}}_i^{-1}$, grouped as follows:

$$g_i(\underline{x}) = \underline{x}^T \underline{\underline{W}}_i \underline{x} + \underline{w}_i \underline{x} + w_0^{(i)}$$

  and give expressions for $\underline{\underline{W}}, \underline{w}$, and w_0 in terms of the same quantities.

(b) State the decision rule in terms of the discriminant functions $g_i(\underline{x})$. Is the classifier linear, quadratic, or neither?

(c) For this part you are also given:

$$\underline{m}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \underline{m}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \underline{m}_3 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2 = \underline{\underline{\Sigma}}_3 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

  (i) Give expressions for the discriminant functions $g_1(\underline{x}), g_2(\underline{x})$ and $g_3(\underline{x})$ in terms of the given numbers, in simplest forms. **Hint:** before plugging in numbers, you can simplify the $g_i(\underline{x})$ by dropping some terms. Is the classifier linear, quadratic, or neither?

  (ii) Let $\pi_i = \frac{1}{3} \ \forall i$. Give equations for the 3 decision boundaries, in simplest form.

  (iii) In (non-augmented) feature space, plot the class means, the 3 curves for $d_M^2(\underline{x}, \underline{m}_i) = 1$, and the decision boundaries. Show clearly the decision regions and final boundaries.

2. A *Naïve Bayes* classifier is a Bayes classifier in which the features, conditioned on class, are assumed independent; thus:

$$p(\underline{x}|S_i) = \prod_{j=1}^{D} p(x_j|S_i) \quad \forall i = 1,\cdots,C.$$

Most of the below is similar to your work for Problem 1.

For a Naïve Bayes (minimum error) classifier, let the given class-conditional densities and priors be:

$$p(\underline{x}|S_i) = N(x,\underline{m}_i,\underline{\underline{\Sigma}}_i), \quad i = 1, 2, 3$$

$$\underline{\underline{\Sigma}}_i = \begin{bmatrix} \left(\sigma_1^{(i)}\right)^2 & 0 \\ 0 & \left(\sigma_2^{(i)}\right)^2 \end{bmatrix}$$

$$P(S_1) = \pi_1, \, P(S_2) = \pi_2, \, P(S_3) = \pi_3$$

(a) Write an expression for the discriminant functions $g_i(\underline{x})$, in the form:

$$g_i(\underline{x}) = w_{11}^{(i)}x_1^2 + w_{12}^{(i)}x_1x_2 + w_{22}^{(i)}x_2^2 + w_1^{(i)}x_1 + w_2^{(i)}x_2 + w_0^{(i)}$$

and give expressions for all the weights in terms of given scalar quantities, in simplest form. Let the mean vector of class S_i be denoted $\underline{m}_i = \begin{bmatrix} m_1^{(i)} & m_2^{(i)} \end{bmatrix}^T$.

(b) Is the classifier quadratic, linear, or neither?

(c) For this part you are also given:

$$\underline{m}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \; \underline{m}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \; \underline{m}_3 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

$$\left(\sigma_1^{(i)}\right)^2 = \sigma_1^2 = 1, \; \left(\sigma_2^{(i)}\right)^2 = \sigma_2^2 = 2, \; \forall i$$

(i) Give expressions for the discriminant functions $g_1(\underline{x}), g_2(\underline{x})$ and $g_3(\underline{x})$ in terms of the given numbers, in simplest forms. **Hint:** before plugging in numbers, you can simplify the $g_i(\underline{x})$ by dropping some terms. Is the classifier linear, quadratic, or neither?

(ii) Let $\pi_i = \frac{1}{3} \; \forall i$. Give equations for the 3 decision boundaries, in simplest form.

(iii) In (non-augmented) feature space, plot the class means, the 3 curves for $d_M^2(\underline{x},\underline{m}_i) = 1, \; i = 1,2,3$, and the decision boundaries. Show clearly the decision regions and final boundaries.

Problem 2 continues on next page...

(d) Compare your plot of 2(c)(iii) with your plot of 1(c)(iii). Does the dependence vs. independence of features make a substantial difference in the decision boundaries and regions?

3. This problem uses the minimum risk criterion (instead of minimum error criterion) for classification (introduced in Lecture 21, covered in Discussion 12 and DHS 2.2). In this problem, please use our notation (S_i instead of ω_i for class i).

This problem is based on our 2-class tumor classification example. Suppose the class-conditional densities are known (or assumed) to be Gaussian:

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

For a Bayes minimum risk classifier, answer the parts below.

(a) Write the decision rule for a 2-class Bayes minimum risk classifier, in terms of conditional risks $R(\alpha_1|\underline{x})$ and $R(\alpha_2|\underline{x})$. Then, write the decision rule in terms of $d_M^2(\underline{x}, \underline{m}_i)$, $\underline{\underline{\Sigma}}_1, \underline{\underline{\Sigma}}_2$, $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}$, $P(S_1)$, and $P(S_2)$.

For parts (b)-(d) below, suppose also that on average, 80% of the tumors that are to be classified (based on their MRI images) are benign (class S_1), and the other 20% are cancerous (class S_2).

(b) Give estimates of the class priors $P(S_1)$ and $P(S_2)$.

For parts (c)-(e) below, suppose also that misclassifying a cancerous tumor as benign, is considered much worse than misclassifying a benign tumor as cancerous. Thus we will use for our cost coefficients:

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 10 \\ 1 & 0 \end{bmatrix}$$

Also, you are given that the mean vectors and covariance matrices for each class are known (or estimated) as:

$$\underline{m}_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad \underline{m}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$\underline{\underline{\Sigma}}_1 = \underline{\underline{\Sigma}}_2 = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 2 \end{bmatrix}$$

(c) Solve for the decision boundary and regions; that is, give an expression (in simplest form, based on the given numbers) for the decision rule.

Problem 3 continues on next page…

(d) Plot in 2D feature space, the class means, the 2 curves for $d_M^2(\underline{x}, \underline{m}_i) = 1, \ i = 1, 2$, the decision boundary, and show the decision regions (by a small arrow at the boundary pointing into Γ_1, or by labelling Γ_1 and Γ_2).

(e) If the guidelines for ordering an MRI change, and it is estimated that the data points coming in will be, on average, 50% benign and 50% cancerous, repeat part (d) (or add it to your part (d) plot, clearly labeled as (e)). How have the boundary and regions changed?