1. **Tumor classification from PET images:** A company recently developed a new positron emission tomography (PET) radiotracer for the detection of tumor regions in the brain. The PET images are collected after the injection of a radiotracer. The intensity of the pixels in the PET images correspond to the amount of radiotracer absorbed by the tissue. Tumor regions are metabolically more active and hence absorb more radiotracers compared to normal tissue.

   It is known that the pixel intensities follow an exponential distribution, whose parameter $\beta$ depends on the uptake of the radiotracer in the specific pixel. Using training data from several subjects and with biopsy as ground truth, the company estimated the $\beta$ parameters of the likelihood distributions of the tumor and non-tumor regions using maximum likelihood method.

$$p(x|\text{normal}) = \begin{cases} \frac{1}{\beta_n}\exp\left(-\frac{x}{\beta_n}\right) & \text{if} \quad x \geq 0 \\ 0 & \text{else} \end{cases}$$

$$p(x|\text{tumor}) = \begin{cases} \frac{1}{\beta_t}\exp\left(-\frac{x}{\beta_t}\right) & \text{if} \quad x \geq 0 \\ 0 & \text{else} \end{cases}$$

   with $\beta_n = e \approx 2.718$ and $\beta_t = 1$. In this problem, you can assume the above distributions to be known and fixed.

   a. Plot the two distributions. You may use python to make the plots; in the exam, you may have to do it by hand.
   b. Assuming the prior probabilities of the classes to be the same, determine the minimum error classification rule. Plot the decision regions. Check if the answer makes sense by comparing with the plot in (a).
   c. Assume that the prior probability of normal tissue to be $e^{\frac{1}{2}} \approx 1.64$ times that of the tumor tissue. Determine the minimum error classification rule with this assumption. Compare with the solution in (b) and justify why this makes sense.
   d. Determine and plot using python the posterior probabilities of the classes.
   e. The mis-classification error can be reduced by rejecting regions with posterior probabilities less than 90%. Determine and plot the decision regions with this setting.
   f. The physicians assess the risk in mis-classifying a normal pixel as tumor to be \$1, while the risk in misclassifying a tumor pixel as normal to be \$148.41 $\approx e^5$, where $e \approx 2.718$. With the above information, determine and plot the minimum risk classifier and plot the decision regions. Explain why the shift in the decision region results in a reduction of risk.

2. **Two dimensional classification (Python component)** Develop the Bayes classifier for fishes using both the features given in fishes.csv. You can start with the provided `2Dclassification.ipynb` file

   (a) Compute and plot the posterior probabilities.

   (b) Show the regions corresponding to minimum error classification. Create an image corresponding to the decision regions. Assign the salmon region to +1 and the bass region to -1 and display.

   (c) Show the regions corresponding to a reject option with an 85% threshold on the posterior distributions.

   (d) Show the regions corresponding to a minimum risk classifier with the following risk table.

| Classes | Salmon | Bass |
|---|---|---|
| Salmon | 0 | 30 |
| Bass | 1 | 0 |

3. **Decision Hypersurfaces for Gaussian Distribution (6 pts):** Consider the two class classification problem, where the likelihood probabilities and the priors for the 2-D feature vectors $\mathbf{x}$ are estimated from the training data as

$$p(\mathbf{x}|\mathcal{C}_1) = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$

and

$$p(\mathbf{x}|\mathcal{C}_2) = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2).$$

Also assume $p(\mathcal{C}_1) = p(\mathcal{C}_2) = 0.5$ and the risk matrix is specified by

$$\mathbf{R} = \begin{bmatrix} 0 & 1 \\ e & 0, \end{bmatrix} \tag{1}$$

where $e \approx 2.718$. Calculate and plot by hand for exam practice the minimum error and minimum risk separating lines for (a) & (b) below. You may double check your answer using Python. For (c), you can use Python.

(a) $\mu_1 = [2, 0]$ and $\mu_2 = [3, 1]$ with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = 3\ \mathbf{I}_{2\times 2}$

(b) $\mu_1 = [2, 0]$ and $\mu_2 = [3, 1]$ with

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{2}$$

(c) $\mu_1 = [2, 0]$ and $\mu_2 = [3, 1]$ with

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{3}$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \tag{4}$$

**Hint:** the inverse of a matrix $\mathbf{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ is given by $\mathbf{A}^{-1} = \frac{1}{ac-b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$

**Hint:** For evaluating the minimum risk separating curves, you need to equate $p(x, \mathcal{C}_1)\mathbf{R}_{12} = p(x, \mathcal{C}_2)\mathbf{R}_{21}$. You can take the log on both sides to simplify.

4. **Two class regression with shared covariance matrix (4 pts):** Assume that you are given training data points $\mathbf{x}_1, .., \mathbf{x}_{N_1}$ from class $\mathcal{C}_1$ and $\mathbf{y}_1, .., \mathbf{y}_{N_2}$ from class 2. We will model the two datasets as Gaussian distributions.

We are interested in using a linear classifier for this application. We know that the discriminant in this case is linear only if $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. We hence assume the distribution of the data to be

$$p(\mathbf{x}|\mathcal{C}_1) = \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) \tag{5}$$

$$p(\mathbf{x}|\mathcal{C}_2) = \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}) \tag{6}$$

Show that

(a) The maximum likelihood estimates of the means of the classes are given by

$$\boldsymbol{\mu}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{x}_i \tag{7}$$

$$\boldsymbol{\mu}_2 = \frac{1}{N_1}\sum_{i=1}^{N_1} \mathbf{y}_i \tag{8}$$

(b) The maximum likelihood estimate of the shared covariance matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma}_{\mathrm{ML}} = \frac{1}{N_1+N_2}\left(\sum_{i=1}^{N_1}(\mathbf{x}_i-\boldsymbol{\mu}_1)(\mathbf{x}_i-\boldsymbol{\mu}_1)^T + \sum_{i=1}^{N_2}(\mathbf{y}_i-\boldsymbol{\mu}_2)(\mathbf{y}_i-\boldsymbol{\mu}_2)^T\right) \tag{9}$$

**Hint 1:** It may be easier to derive the inverse of $\mathbf{\Sigma}_{\mathrm{ML}}$, specified by $\mathbf{P}_{ML} = \mathbf{\Sigma}_{ML}^{-1}$. Note that the determinant of the inverse of a matrix is the inverse of the determinant.

**Hint 2:** You may use these matrix calculus relations.

$$\nabla_{\mathbf{Q}} \log(|\mathbf{Q}^{-1}|) = \mathbf{Q} \tag{10}$$

$$\nabla_{\mathbf{Q}} \left(\mathbf{p}^T\mathbf{Q}\mathbf{p}\right) = \nabla_{\mathbf{Q}}\, \mathrm{trace}\left(\mathbf{Q}\,\mathbf{p}\mathbf{p}^T\right) = \mathbf{p}\mathbf{p}^T \tag{11}$$

**Note:** We had used these expressions (7), (8), and (9) to estimate the shared covariance matrices of the Salmon and Bass classes in the mini-assignment.

5. **Python component: Two class minimum distance classifiers for digit classification** You can complete `minimum_distance_digits.ipynb` to realize a minimum distance classifier. The digits are 8x8 images with 64 pixels per image; this example extends the 2-D theory we covered in class to 64 dimensional space.

   - Assume the covariance matrix of the classes to be $\Sigma_1 = \Sigma_2 = \mathbf{I}$ and realize the minimum distance classifier. Determine the discriminant $y(x)$ and estimate `class1_estimate`. While discriminating digits 0 and 1, you should only get two mis-classifications.
   - Assume the shared covariance matrix of the classes to be $\Sigma_1 = \Sigma_2$ to be a diagonal matrix with the diagonal entries as variances $\sigma_i$ of the $i^{\text{th}}$ pixels. You can compute the diagonal entries of $\Sigma_{ML}$ using (9). Determine the Mahalanobis distance classifier. Determine the discriminant $y(x)$ and estimate `class1_estimate`. While discriminating digits 0 and 1, you should only get two mis-classifications.
   - Repeat the above for digits 2 and 3 and report the mis-classifications. in each case.