Homework #2
Bayesian Decision Theory
Statistical Pattern Recognition

Due Aban 16th,1395
Format of homework file: Archive all files in a folder named as your student number and send it to mohammadhme@gmail.com. Send your emails with a subject of PR95F2_xxxxxx (replace xxxxxx by your student number)

---

1. Two normal distribution are characterized by:

$$P_1 = P_2 = 0.5$$

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \Sigma_1 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

   a. Draw the Bayes decision boundary to minimize the probability of error.
   b. Draw the Bayes decision boundary to minimize the cost with $c_{11} = c_{22} = 0$ and $c_{12} = 2c_{21}$
   c. Assume that $c_{11} = c_{22} = 0$ and $c_{12} = c_{21}$:
      i. Plot the operating characteristics.
      ii. Find the total error when Neyman-Pearson test is performed with $\epsilon_1 = 0.05$
      iii. Find the threshold value and total error for the minimax test.
      iv. Plot the error-reject curve.

---

2. Two normal distribution are characterized by:

$$P_1 = P_2 = 0.5$$

$$M_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \Sigma_1 = \begin{bmatrix} 4 & -3 \\ -3 & 4 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \qquad \Sigma_2 = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$$

   a. Find the linear discriminant function which maximize the Fisher criterion and minimize the error by adjusting the threshold.
   b. Find the optimum linear discriminant function which minimize the probability of error.

---

3. We consider a classification problem in dimension d=2, with k=3 classes where: $p(x \mid w_i) \sim N(\mu_i, \Sigma_i),\ i = 1,2,3$, with

$$\mu_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mu_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma_i = \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1/3 \end{bmatrix}$$

   a. Calculate the discriminant function $g_i(x)$ for each class.
   b. Express your discriminant functions in the form of linear discriminant functions.
   c. Determine and plot the decision boundaries.

---

4. Consider the following 2-class classification problem involving a single feature x. Assume equal class priors and 0-1 loss function.

$$p(x \mid w_1) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & otherwise \end{cases} \qquad p(x \mid w_2) = \begin{cases} 2-2x & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

   a. Sketch the two densities.
   b. State the Bayes decision rule and show the decision boundary.
   c. What is the Bayes classification error?
   d. How will the decision boundary change if the prior for class w_2 is increased to 0.6?

---

5. Consider a two-category classification problem in two dimensions with

$$p(X \mid \omega_1) \sim N(0, I),\ P(X \mid \omega_2) \sim N\left(\begin{pmatrix}1\\0\end{pmatrix}, I\right) \text{ and } P(\omega_1) = P(\omega_2) = 1/2.$$

   a. Calculate the Bayes decision boundary.
   b. Calculate the Bhattacharyya error bound.
   c. Repeat the above for the same prior probabilities, but

$$p(X \mid \omega_1) \sim N\left(0, \begin{pmatrix}2 & 0.5\\0.5 & 2\end{pmatrix}\right) \textit{ and } p(X \mid \omega_2) \sim N\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}5 & 2\\2 & 5\end{pmatrix}\right).$$

---

6. Consider a two-category classification problem in two dimensions with $p(\omega_1) = p(\omega_2)$, assume that the class-conditional densities are Gaussian with mean μ_1 and co-variance Σ_1 under class 1, and mean μ_2 and co-variance Σ_2 under class 2. Further, assume that $\mu_0 = \mu_1$.

$$\Sigma_1 = \begin{bmatrix}1 & 0\\0 & 4\end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix}4 & 0\\0 & 1\end{bmatrix}$$

For the following case, draw contours of the level sets of the class conditional densities. Also, draw the decision boundaries obtained using the Bayes optimal classifier in each case and indicate the regions where the classifier will predict class 1 and where it will predict class 2.

---

7. Consider the two-dimensional data points from two classes ω_1 and ω_2 below, and each of them come from a Gaussian distribution $p(x \mid \omega_k) \sim N(\mu_k, \Sigma_k)$.

$\boldsymbol{\omega_1}$	$\boldsymbol{\omega_2}$
(1,1)	(6,8)
(2,2)	(7,9)
(3,3)	(9,9)
(1,-1)	(6,8)
(0,1)	(9,8)
(2,2)	(9,11)
	(10,9)
	(10,7)

a. What is the prior probability for each class, i.e. $p(\omega_1)$ and $p(\omega_2)$.
b. Calculate the mean and covariance matrix for each class.
c. Derive the equation for the decision boundary that separates these two classes, and plot the boundary. (Hint: you may want to use the posterior probability)
d. Think of the case that the penalties for misclassification are different for the two classes (i.e. not zero-one loss), will it affect the decision boundary, and how?

---

8. Consider a classification problem with 2 classes and a single real-valued feature vector X. For class 1, $p(x \mid c_1)$ is uniform U(a, b) with a = 2 and b = 4. For class 2, $p(x \mid c_2)$ is exponential with density $\lambda\exp(-\lambda x)$ where $\lambda = 1$. Let $p(c_1) = p(c_2) = 0{:}5$.
   a) Determine the location of the optimal decision regions
   b) Draw a sketch of the two class densities multiplied by $P(c_1)$ and $P(c_2)$ respectively, as a function of x, clearly showing the optimal decision boundary (or boundaries)
   c) Compute the Bayes error rate for this problem within 3 decimal places of accuracy
   d) Answer the questions above but now with a = 2 and b = 22.

---

9. Consider a 2-class classification problem with d-dimensional real-valued inputs x, where the class-conditional densities, $p(\underline{x} \mid c_1)$ *and* $p(\underline{x} \mid c_2)$ are multivariate Gaussian with different means $\underline{\mu}_1$ *and* $\underline{\mu}_2$ and a common covariance matrix Σ, with class probabilities $P(c_1)$ and $P(c_2)$.
   a) Write the discriminant functions for this problem in the form of $g_1(\underline{x}) = \log p(x \mid c_1) + \log p(c_1)$ (same for $g_2(\underline{x})$).
   b) Prove that the optimal decision boundary, at $g(\underline{x}) = g_1(\underline{x}) - g_2(\underline{x}) = 0$, can be written in the form of a linear discriminant, $\underline{w}^t \underline{x} + w_0 = 0$, where $\underline{w}$ is a d-dimensional weight vector and w_0 is a scalar, and clearly indicate what are $\underline{w}$ and w_0 are in terms of parameters of the classification model.

---

10. Consider the two-dimensional data points from two classes ω_1 and ω_2 below:

ω_1	ω_2
(1,1)	(2,2)
(1,2)	(3,2)
(1,3)	(3,4)
(2,1)	(5,1)
(3,1)	(5,4)
(3,3)	(5,5)

   a. Determine and plot the optimal projection line in a single dimension using Fisher linear discriminant method.
   b. Show the mapping of the points to the line as well as the Bayes discriminant assuming a suitable distribution.

---

11. [Computer Experiment] In this problem, you will be classifying the “Iris” dataset from the UCI Machine Learning Repository. The original data describes 3 classes of Iris flowers and contains various measurements about them: 1. sepal length in cm 2. sepal width in cm 3. petal length in cm 4. petal width in cm. To make things a little easier, we left out one of

the flower types and only included the “Iris Setosa” (class 0) and “Iris Versicolor” (class 1). There are two data files for this assignment, a training data set (70 data items) and a test data set (30 data items). The training and testing datasets can be found at

http://www.ias.informatik.tu-darmstadt.de/uploads/Teaching/MachineLearning1Lecture/Iris_train.dat

and

http://www.ias.informatik.tu-darmstadt.de/uploads/Teaching/MachineLearning1Lecture/Iris_test.dat

Every line in each file describes one data item; the 5 numbers in each row contain the 4 measurements in above order plus the class label. **Use only the second measurement (sepal width) for classification and ignore all others**.

a) Load the training data. Compute a feature histogram for each class (see histc and bar in Matlab), using edges between the bins at 0.2 intervals between 1 and 5; i.e. 1 : 0.2 : 5. Show the histograms either in two figures using subplot in Matlab, and make sure both plots have the same x-axis range.
b) Estimate the class priors from the training data. Report the prior probabilities of a flower being an “Iris Setosa” or being an “Iris Versicolor”.
c) Assume that we can model the class-conditional probability density of each class using a univariate Gaussian (*remember to only use the sepal width).* What are the mean and variance of both class-conditional densities?
d) Plot the estimated class-conditional densities in a single diagram. Make a second figure plotting the class posteriors for both classes. Write down the equation you used to compute the class posteriors.
e) Compute the Bayes classification of the training data. How many “Setosas” do you mistakenly classify as “Versicolor” and how many “Versicolors” do you mistakenly classify as “Setosa”? What is the overall training error (i.e. percentage of incorrectly classified training data items)?
f) Now classify the test data. How many “Setosas” do you mistakenly classify as “Versicolor” and how many “Versicolors” do you mistakenly classify as “Setosa”? Report the number of cases that are incorrectly classified as above as well as the test error.