1. Assune that in a c-class classification problem, we have k features $X_1, X_2, \dots, X_k$ that are independent and $X_j|\omega_i \sim \text{Gamma}(p_i, \lambda_j)$, i.e. $p_{X_j|\omega_i}(x_j|\omega_i) = \frac{1}{\Gamma(p_i)}\lambda_j^{p_i}x_j^{p_i-1}e^{-\lambda_j x_j}$, p_i, $\lambda_j > 0$.

   (a) Determine the Bayes' optimal classifier's decision rule making the general assumption that the prior probability of the classes are different.

   (b) When are the decision boundaries linear functions of $x_1, x_2, \dots, x_k$?

   (c) Assuming that $p_1 = 4, p_2 = 2, c = 2, k = 4, \lambda_1 = \lambda_3 = 1, \lambda_2 = \lambda_4 = 2$, and that the prior probabilites of each class are equal, classify $\mathbf{x} = (0.1, 0.2, 0.3, 4)$.

   (d) Assuming that $p_1 = 3.2, p_2 = 8, c = 2, k = 1, \lambda_1 = 1$, and that the prior probabilites of each class are equal, find the decision boundary $x = x^*$. Also, find the probability of type-1 and type-2 errors.

   (e) Assuming that $p_1 = p_2 = 4, c = 2, k = 2, \lambda_1 = 8, \lambda_2 = 0.3$, and $P(\omega_1) = 1/4, P(\omega_2) = 3/4$, find the decision boundary $f(x_1, x_2) = 0$.

2. Assume that in a c-class classification problem, there are k independent features and $X_i|\omega_j \sim \text{Lap}(m_{ij}, \lambda_i)$, i.e. $p_{X_i|\omega_j}(x_i|\omega_j) = \frac{\lambda_i}{2}e^{-\lambda_i|x_i - m_{ij}|}, \lambda_i > 0, i \in \{1, 2, \dots, k\}, j \in \{1, 2, \dots, c\}$. Assuming that the prior class probabilities are equal, show that the minimum error rate classifier is also a minimum weighted Manhattan distance (or weighted $\mathscr{L}_1$-distance) classifier. When does the minimum error rate classifier becomes the minimum Manhattan distance classifier?

3. The class-conditional density functions of a discrete random variable X for four pattern classes are shown below:

x	$p(x\|\omega_1)$	$p(x\|\omega_2)$	$p(x\|\omega_3)$	$p(x\|\omega_4)$
1	1/3	1/2	1/6	2/5
2	1/3	1/4	1/3	2/5
3	1/3	1/4	1/2	1/5

   The loss function $\lambda(\alpha_i|\omega_j)$ is summarized in the following table, where action α_i means decide pattern class ω_i:

	ω_1	ω_2	ω_3	ω_4
α_1	0	2	3	4
α_2	1	0	1	8
α_3	3	2	0	2
α_4	5	3	1	0

   Assume $P(\omega_1) = 1/10, P(\omega_2) = 1/5, P(\omega_3) = 1/2, P(\omega_4) = 1/5$.

   (a) Compute the conditional risk for each action as:
   $R(\alpha_i|x) = \sum_{j=1}^{4} \lambda(\alpha_i|\omega_j)p(\omega_j|x)$

(b) Compute the overall risk R as:
$R = \sum_{i=1}^{3} R(\alpha(x_i)|x_i)p(x_i)$
where $\alpha(x_i)$ is the decision rule minimizing the conditional risk for x_i.

4. The following data set was collected to classify people who evade taxes:

Tax ID	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	122 K	No
2	No	Married	77 K	No
3	No	Married	106 K	No
4	No	Single	88 K	Yes
5	Yes	Divorced	210 K	No
6	No	Single	72 K	No
7	Yes	Married	117 K	No
8	No	Married	60 K	No
9	No	Divorced	90 K	Yes
10	No	Single	85 K	Yes

Considering relevant features in the table (only one feature is not relevant), assume that the features are *independent*.

(a) Estimate prior class probabilities.

(b) For continuous feature(s), assume conditional Gaussianity and estimate class conditional pdfs $p(x|\omega_i)$. Use Maximum Likelihood Estimates.

(c) For each discrete feature X, assume that the number of instances in class ω_i for which $X = x_j$ is n_{ji} and the number of instances in class ω_i is n_i. Estimate the probability mass $p_{X|\omega_i}(x_j|\omega_i) = P(X = x_j|\omega_i)$ as n_{ji}/n_i for each discrete feature. Is this a valid estimate of the pmf?

(d) There is an issue with using the estimate you calculated in 4c. Explain why the laplace correction $(n_{ji}+1)/(n_i+l)$, where l is the number of levels X can assume,[1] solves the problem with the estimate given in 4c. Is this a valid estimate of the pmf?

(e) Estimate the minimum error rate decision rule for classifying tax evasion using Laplace correction.

[1] For example, if $X \in \{apple, orange, pear, peach, blueberry\}$, then $d = 5$.