Assignment 2

1. **MLE for the Poisson distribution** (pt= 6)

The Poisson pmf is defined as

$$Poi(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \text{ for } x \in \{0, 1, 2,...\}$$

where λ > 0 is the rate parameter. Derive MLE of λ - assume you have N samples from X.

2. **Bayesian analysis of the exponential distribution** (pt=18)

A lifetime – X – of a machine is modeled by an exponential distribution with unknown parameter θ. The likelihood is

$$p(x|\theta) = \theta e^{-\theta x} \text{ for } x \geq 0, \theta > 0.$$

a. Derive the MLE of θ - assume you have N samples from X.

b. Suppose we observe $x_1 = 5, x_2 = 6, x_3 = 4$ (the lifetimes (in years) of 3 different iid machines). What is the MLE given this data?

c. Assume that an expert believes θ should have a prior distribution that is also exponential $p(\theta) = Expon(\theta|\lambda)$. Choose the prior parameter, call it $\hat{\lambda}$, such that $E[\theta] = 1/3$. Hint: recall that the Gamma distribution has the form $Ga(\theta|a,b) \propto \theta^{a-1}\, e^{-\theta b}$ and it's mean is a/b.

d. What is the posterior, $p(\theta|D,\hat{\lambda})$?

e. Is the exponential prior conjugate to the exponential likelihood?

f. What is the posterior mean, $E[\theta|D,\hat{\lambda}]$?

3. **Bayesian Linear regression** (pt = 16)

Linear models are a way of describing a response variable – y - in terms of a linear combination of predictor variables – x. The relationship between y and x can be described by the conditional distribution of $p(y|x)$. Let's assume

$$p(y|x) \sim N(y; ax + b, \sigma^2)$$

This means that the expected value of y given x is $ax + b$ and its variance is equal to σ^2.

a. Let's assume we have N pairs of (y_i, x_i) measurements – $\mathcal{D} = \{(y_i, x_i)\}_{i=1}^N$. We also assume y_i s are $\mathrm{i.i.d.}$ measurements. Derive the MLE for a, b, and σ^2.

b. Let’s assume we have the following measurements of x and y. Derive MLE for a, b, and σ^2. Plot your data points and its linear prediction.

x	y
2.0	7.89
-3.6	-16.55
2.2	6.73
4.9	17.91
1.5	2.06
3.1	12.84
2.2	8.13
-0.9	-5.35
0.9	3.97
-2.4	-12.31

c. Assume we know σ^2 and we have a prior belief about a and b. What sort of prior do you suggest for a and b.

d. Using step b, derive MAP estimate for a and b.

4. Basic and Naive Bayes Classifiers (pt = 16)

To detect a specific disease, we use two different blood measurements – an indicator measurement (0/1) and a continuous measurement. We assume the indicator measurement follows a Bernoulli distribution and the continuous measurement follows a normal distribution. We collected 100 measurements from different people, and we want to build a classifier based on the data being recorded.

The data file assignment_2_problem_4.mat provides this measurement. The first column is the indicator measurement, the second column is the continuous measurement. The last column shows whether the patient has the disease or not.

a. Plot the histogram of both measurements for healthy and patient groups.

b. Build a basic classifier based on the discrete measurement, and report how many data samples are classified correctly per each class. Provide the model parameters you derived to build this classifier – priors and conditional distribution parameters.

c. Build a basic classifier based on the continuous measurement, and report how many data samples are classified correctly per each class. Provide the parameters utilized in building this classifier.
d. Build a classifier – Naïve Bayes – using both measurements. Assume that the two measurements are independent. Provide the parameters derived in building this classifier. and report the classification result per class.
e. Compare the performance of these classifiers.

5. **Exercise 4.22 QDA with 3 classes** (pt:16)

Consider a three category classification problem. Let the prior probabilities:

$$P(Y=1)=P(Y=2)=P(Y=3)=1/3$$

The class-conditional densities are multivariate normal densities with parameters:

$$\mu_1=[0,0]^T, \mu_2=[1,1]^T, \mu_3=[-1,1]^T$$

$$\Sigma_1=\begin{bmatrix}0.7 & 0\\ 0 & 0.7\end{bmatrix}, \Sigma_2=\begin{bmatrix}0.8 & 0.2\\ 0.2 & 0.8\end{bmatrix}, \Sigma_3=\begin{bmatrix}0.8 & 0.2\\ 0.2 & 0.8\end{bmatrix}$$

Classify the following points:

a. $\boldsymbol{x}=[-0.5,0.5]$
b. $\boldsymbol{x}=[0.5,0.5]$

6. **Exercise 4.17 LDA/QDA on height/weight data** (pt: 16)

The function *discrimAnalysisHeightWeightDemo* fits an LDA and QDA model to the height/weight data. Compute the misclassification rate of both of these models on the training set. Turn in your numbers and code.

7. **Multivariate Normal Distribution** (pt: 12)

We want to learn how MVN samples are generated. Let's assume we have three classes of data, and we want to create data samples from each class. The data per each class follows an MVN distribution – the data dimension is 2, with the following mean and covariance matrices.

Class 1: $\mu_1=\begin{bmatrix}1\\ -1\end{bmatrix}$ and $\Sigma_1=\begin{bmatrix}2 & 0\\ 0 & 16\end{bmatrix}$

Class 2: $\mu_2=\begin{bmatrix}-1\\ 1\end{bmatrix}$ and $\Sigma_2=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$

Class 3: $\mu_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ and $\Sigma_3 = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix}$

The prior for class 1 is 0.4 ($\pi_1 = 0.4$), class 2 is 0.3 ($\pi_2 = 0.3$), and class 0.3 ($\pi_3 = 0.3$).

a. Let's assume we have 1000 samples in total. Draw the proper number of samples per each class and plot them in one plot with different colors.

b. Let's assume all the data points generated in step a represent one class – or category; what is the mean and covariance of the whole dataset. Draw 1000 samples from this distribution and plot them in a graph. Discuss the difference between these data points and the one you have created in step a.

Helpful Matlab functions: mvnrnd, polyfit