**Machine Learning: Assignment 3**

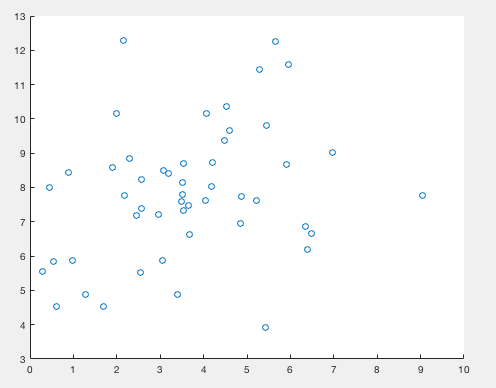
Due: Thursday, February 8, 2018

**Ang Li**

**Problem 1. Multivariate Gaussian**

Assume the pairs of real valued measurements in file ’gaussian.txt’

(a) Plot the data using the scatter plot matlab function.



(b) Calculate and report the ML estimate of the mean and the covariance matrix from the data. Please use the unbiased estimate of the variance. Plot and report the resulting Gaussian distribution. (Note: you need to plot this in 3D).

ML\_mu =

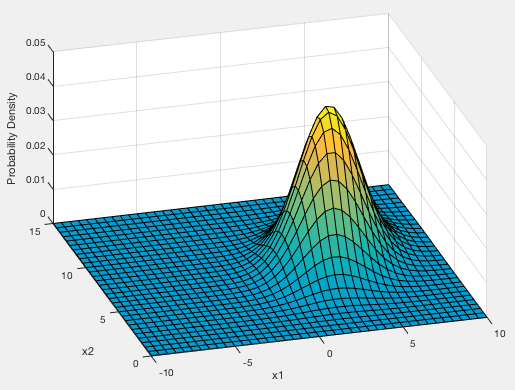
3.6377

7.8506

ML\_Sigma =

3.6414 1.0779

1.0779 3.7831



(c) Now consider each measurement in ’gaussian.txt’ separately. Calculate the ML estimate of the mean and variance of these measurements. Plot and report the individual distributions.

ML\_mu\_d1 = 3.6377

ML\_mu\_d2 = 7.8506

ML\_Sigma\_d1 = 3.6414

ML\_Sigma\_d2 = 3.7831

(d) Do you believe the mutlivariate Gaussian model is better than two separate univariate Gaussian models? Explain why yes or why not? How would you use the data to answer that question?

Yes, multivariate Gaussian will have better estimation. Multivariate Gaussian is a joint probability among all the measures, which also captured the covariance between different measures. However, two separate univariate Gaussian distribution could only model the data from two separate views, which cannot capture the covariance between the features.

So basically, assumption of P(x1, x2) = P(x1)\*p(x2) is that x1 and x2 are independent with each other.

Based on data, we can calculate the

1. P(Data| mu, sigma) based on the multivariate Gaussian distribution
2. P1(Data\_x1|mu1, sigma\_x1) and P2(Data\_x2|mu2, sigma\_x2)
3. Compare the probability P(Data| mu, sigma), and P1(Data\_x1|mu1, sigma\_x1)\* P2(Data\_x2|mu2, sigma\_x2)
4. Using multivariate Gaussian distribution can get larger probability

**Problem 2. Poisson distribution**

The Poisson distribution is used to model the **number of random arrivals** to a system over a fixed period of time. Examples of systems in which events are determined by random arrivals are: arrivals of customers requesting the service, occurrence of natural disasters, such as floods, etc. The Poisson distribution is defined as:

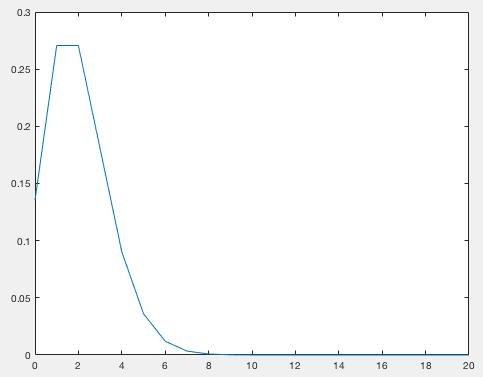


where λ is a parameter. The mean of the Poisson distribution is λ.

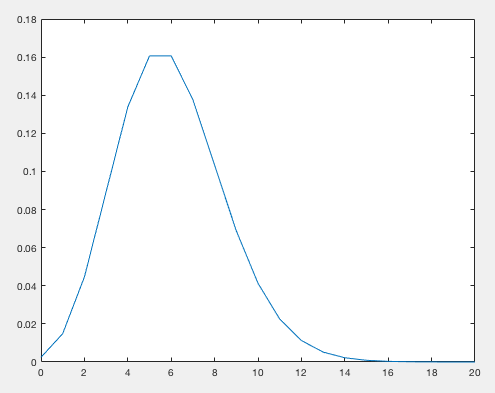
Answer the following questions:

(a) Plot and report the probability function for Poisson distributions with parameters λ= 2 and λ = 6. Note that the Poisson model is defined over nonenegative integers only.

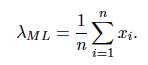
λ= 2



λ= 6



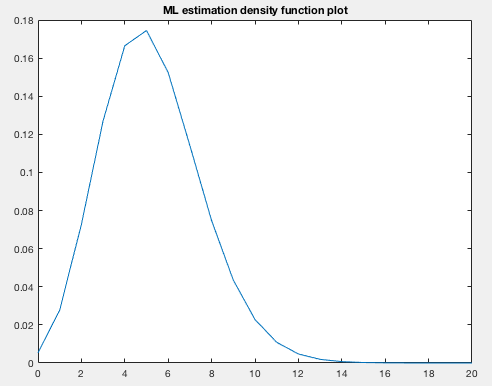
(b) Given a set of independent observations x1, x2, . . . , xn from a Poisson distribution, the ML estimate of the parameter λ is:



Assume the data in poisson.txt file that represent the number of incoming phone calls received over a fixed period of time. Compute and report the ML estimate of the parameter λ. Also plot and report the probability function for the ML parameter.

ML\_lamda =

5.2400



(c) The conjugate prior for λ defining the Poisson distribution is Gamma distribution. It is defined as:



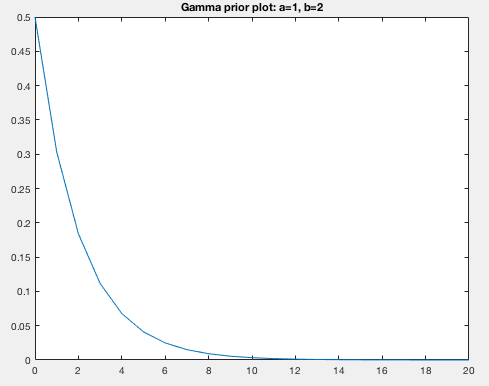
Plot and report the Gamma distribution for the following set of parameters (a = 1, b = 2) and (a = 3, b =5).

x = 0:20;

prior\_y1 = gampdf(x,1,2);

plot(x, prior\_y1)

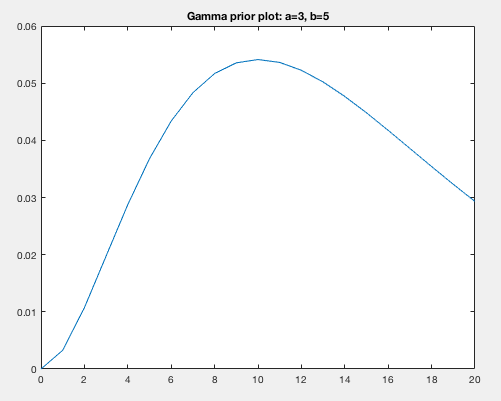
title('Gamma prior plot: a=1, b=2')



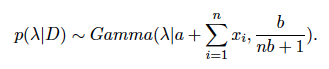
prior\_y2 = gampdf(x,3,5);

plot(x, prior\_y2)

title('Gamma prior plot: a=3, b=5')



(d) Assuming the prior distribution on λ is Gamma(λ|a, b), the posterior distribution for λ after seeing observations D = {x1, x2, . . . , xn} is again gamma distribution:



Please use data in ’poisson.txt’ to calculate and plot the posterior distributions of λ for both priors in Part c.

x = 0:12;

a1=1; b1=2;

post\_a1 = a1 + sum(poisson);

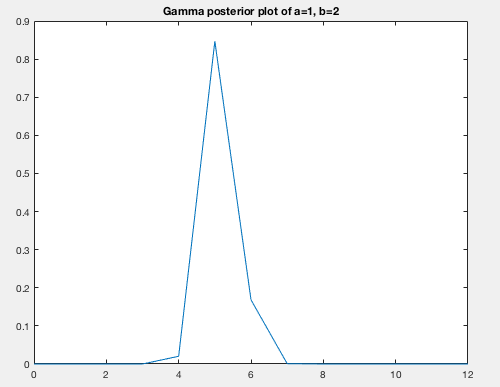
n = size(poisson,1);

post\_b1 = b1/(n\*b1 + 1);

post\_y1 = gampdf(x, post\_a1, post\_b1);

plot(x, post\_y1)

title('Gamma posterior plot of a=1, b=2')



a2=3; b2=5;

post\_a2 = a2 + sum(poisson);

post\_b2 = b2/(n\*b2 + 1);

post\_y2 = gampdf(x, post\_a2, post\_b2);

plot(x, post\_y2)

title('Gamma posterior plot of a=3, b=5')

