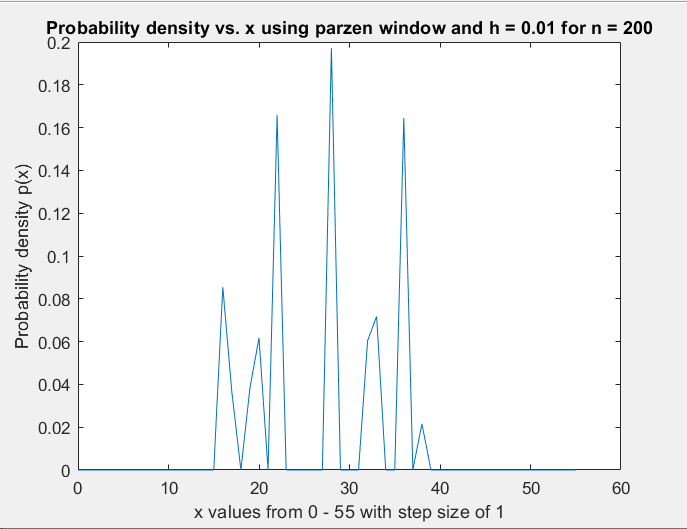
CSE 802 Homework 04

By Abhiram Durgaraju

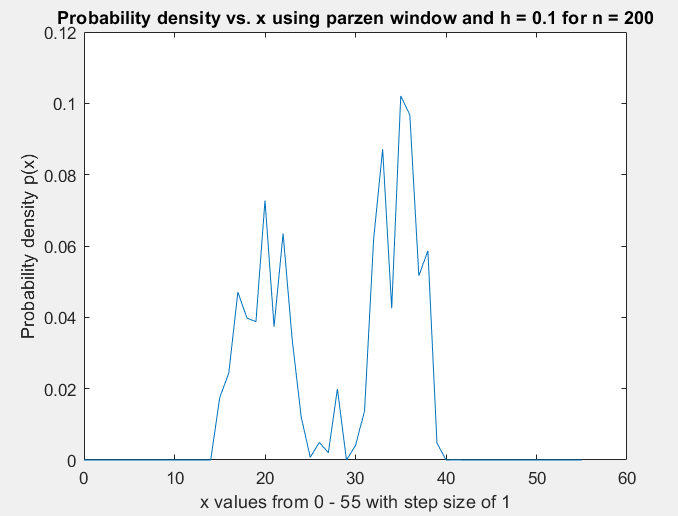
Problem 1)

200 points:

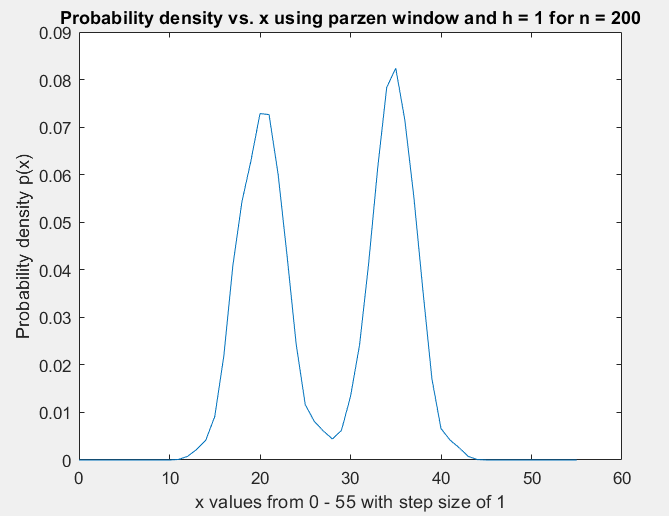
h = 0.01;



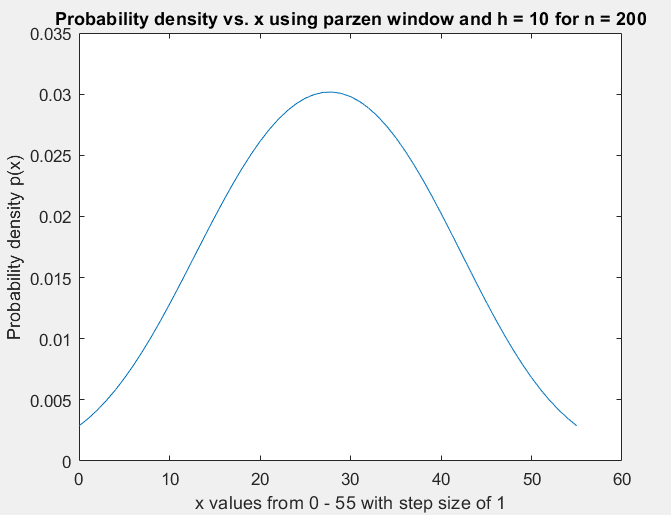
h = 0.1



h = 1

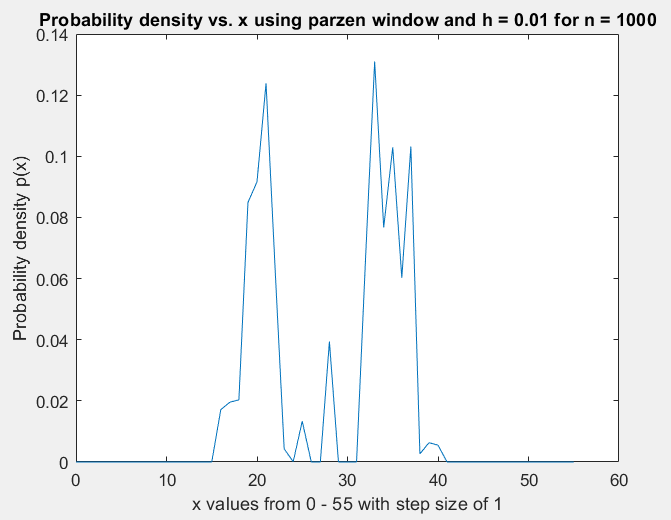


h = 10

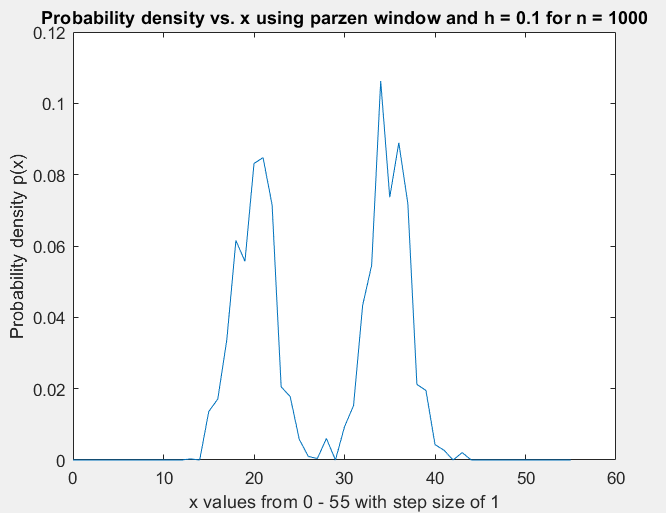


500 points:

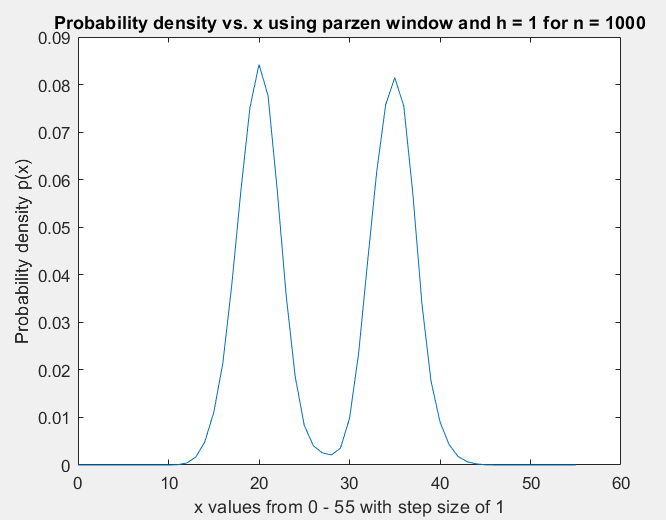
h=0.01



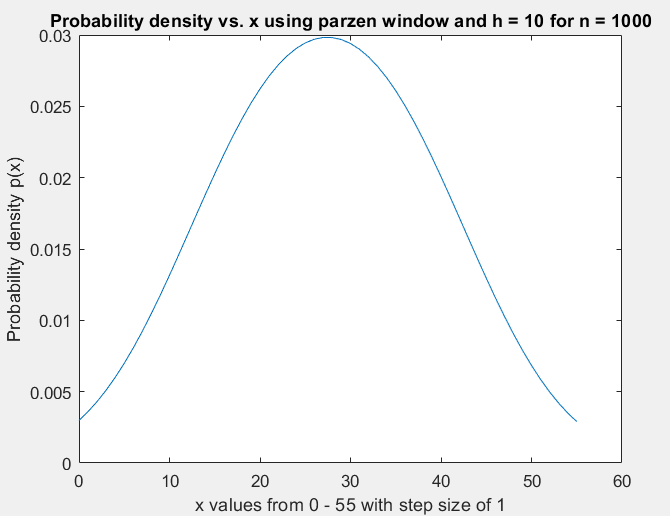
h=0.1



h=1

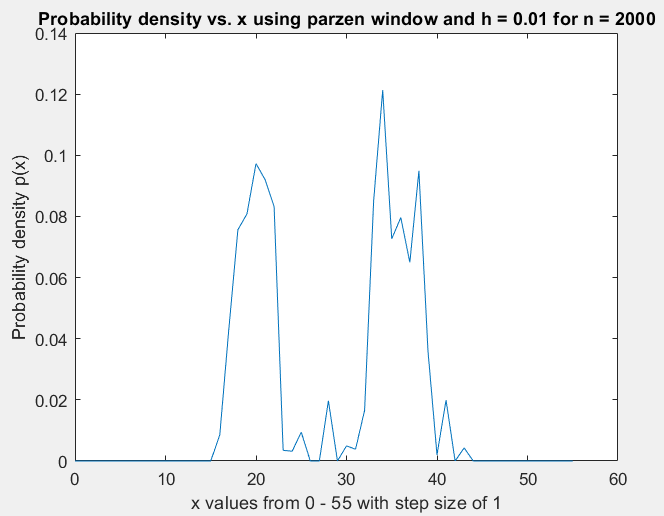


h=10

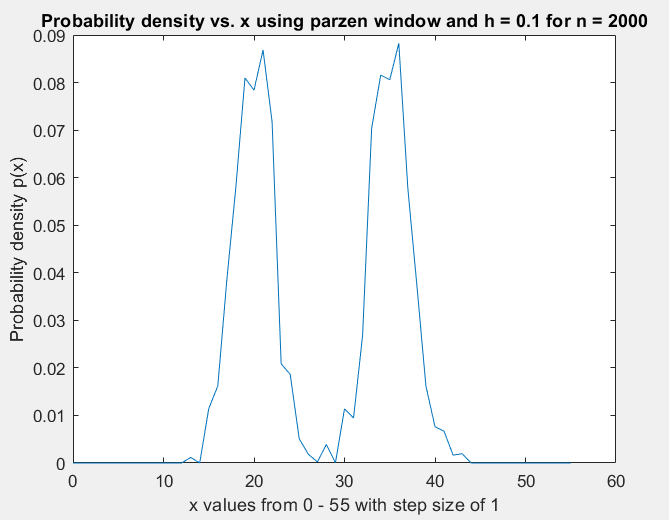


1000 points:

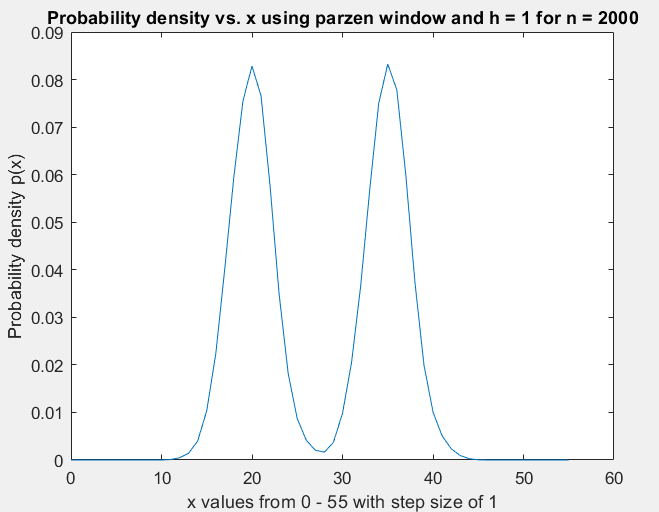
h=0.01



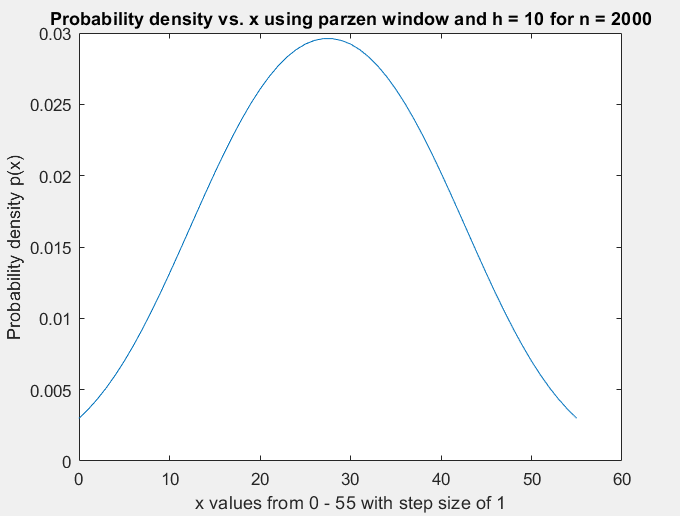
h=0.1



h=1



H=10

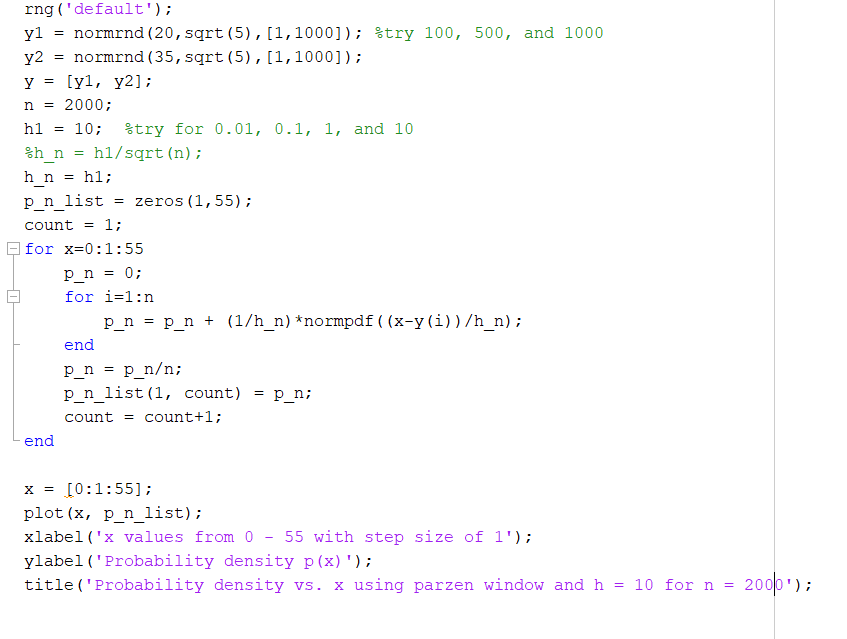


c)

As the window width increases, in general, the bimodal distribution starts to show from the estimated density. However, increasing the window width too much, to h = 10, it “smooths” out the density resulting in a false unimodal approximation. At very low window width however, we see there is sharp variation in estimated density.

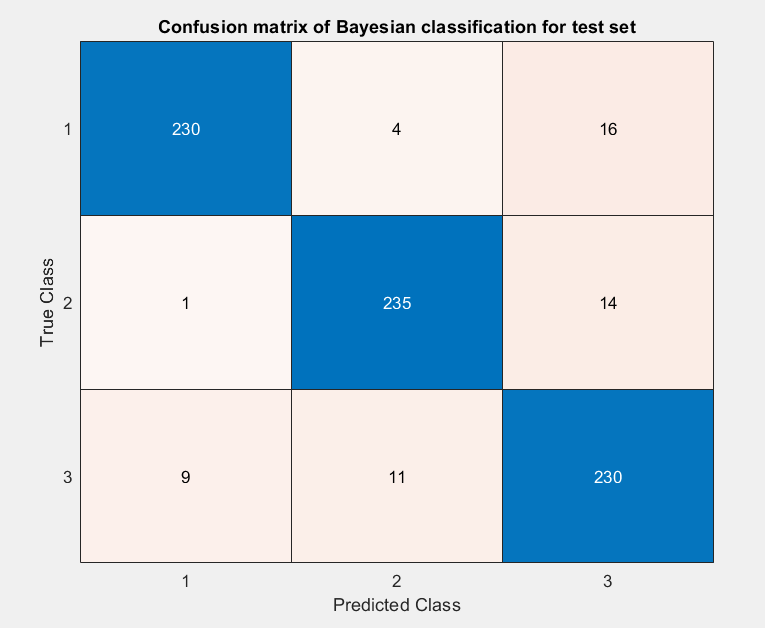
As the number of training samples increase, the effect of variation on window width is suppressed. The density curve is generally smoother for a given window width when compared to densities at lower training sample size. The bimodal distribution is seen, with less jaggedness, for h = 1. However, at h=10, regardless on the sample size, the density estimation converges to a unimodal gaussian distribution.

Code used for problem 1:



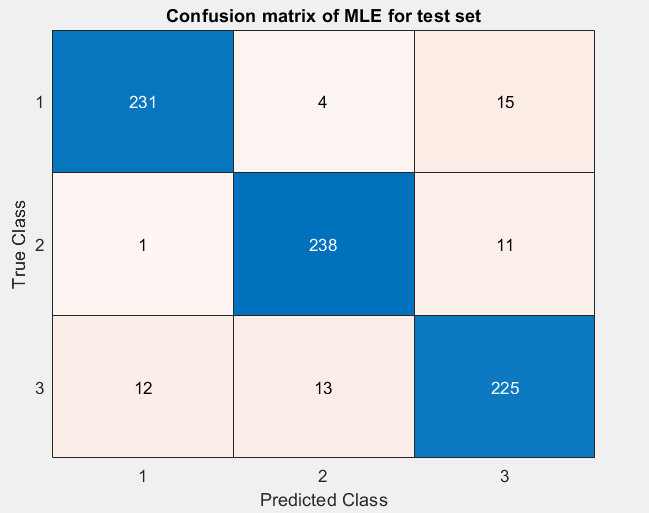
Problem 2)

a)



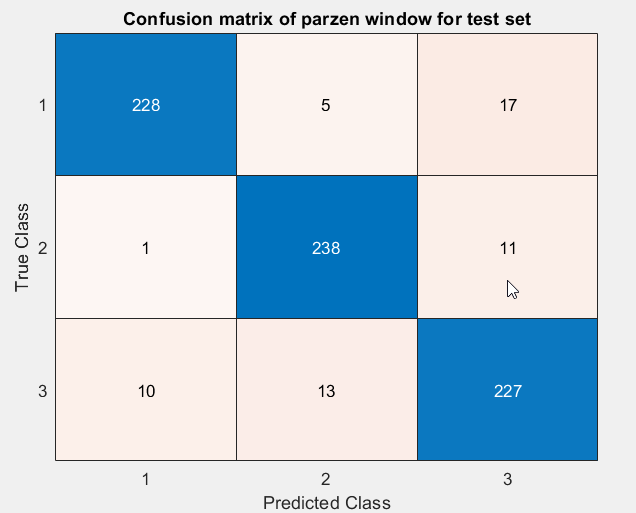
55 samples in the test set were misclassified. Error rate = 55/750 = 0.0733

b)



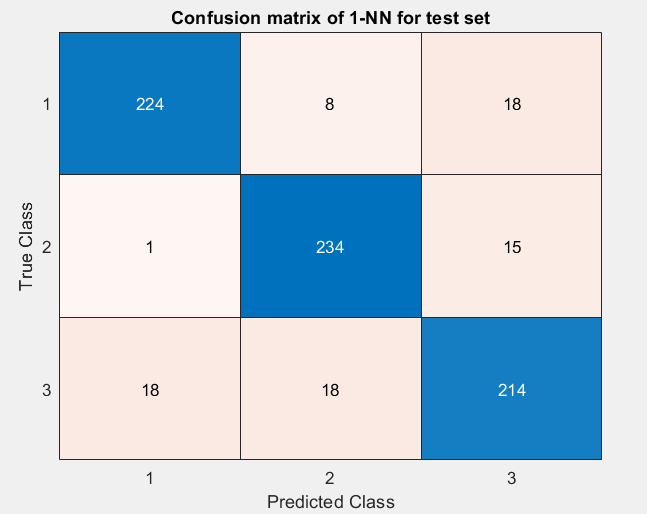
56 samples in the test set were misclassified. Error rate = 56/750 = 0.0746

c)



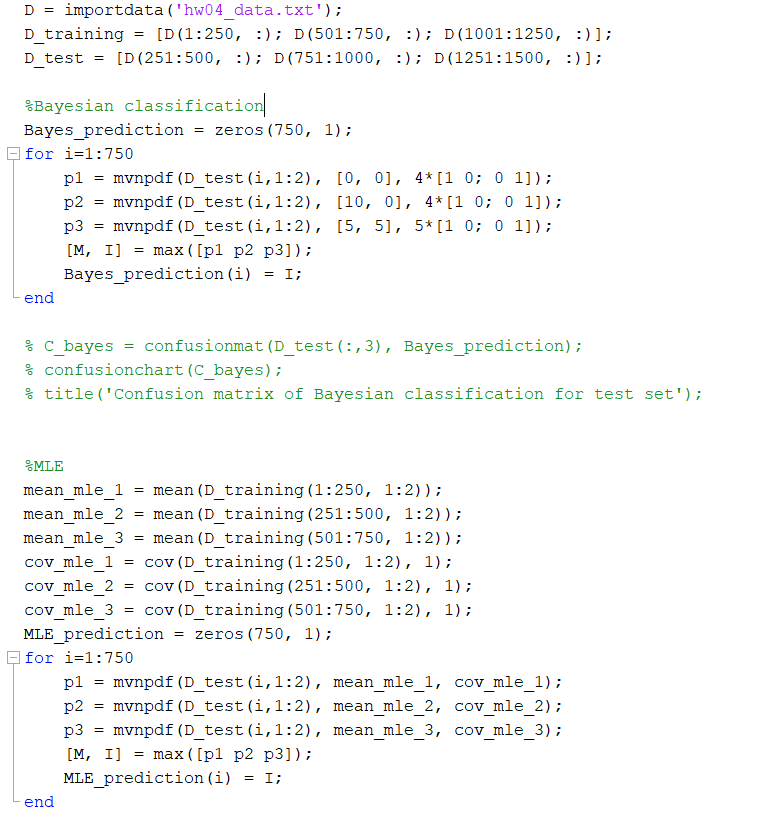
82 samples were misclassified. Error rate = 57/750 = 0.076

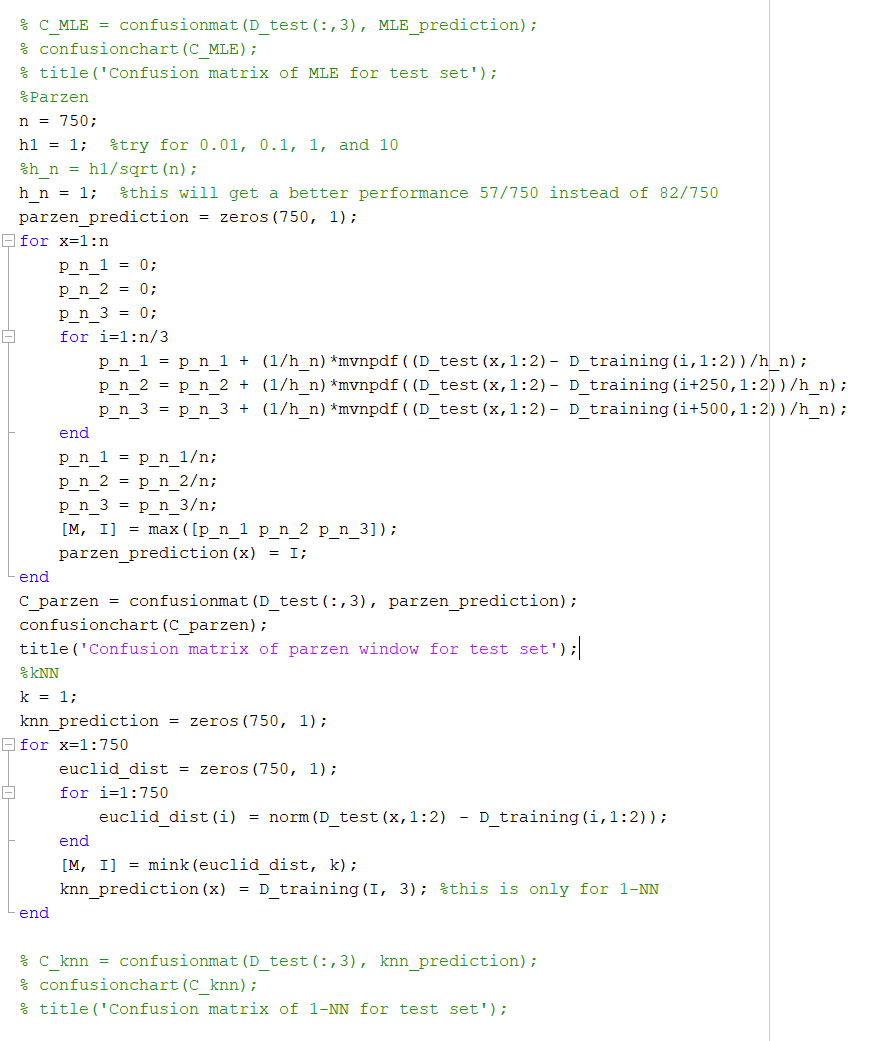
d)



78 samples were misclassified. Error rate = 78/750 = 0.1040

Code used for problem 2:

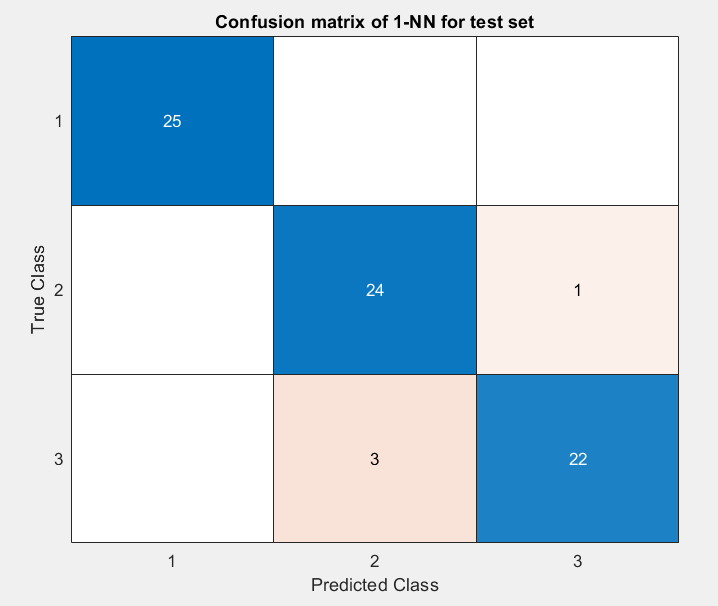




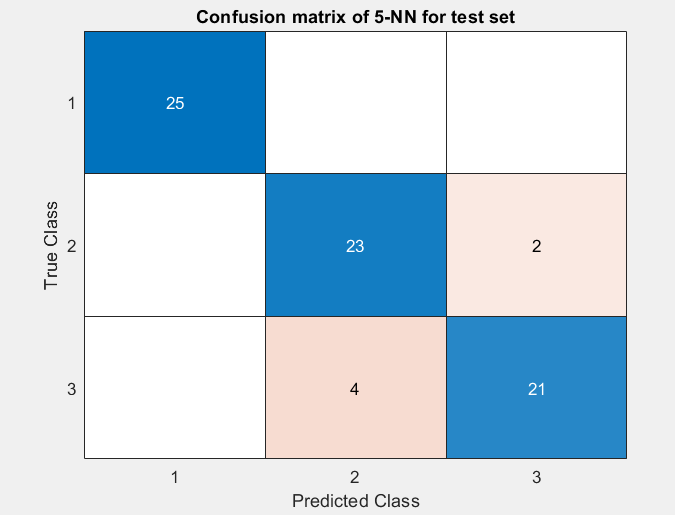
Problem 3)

a)

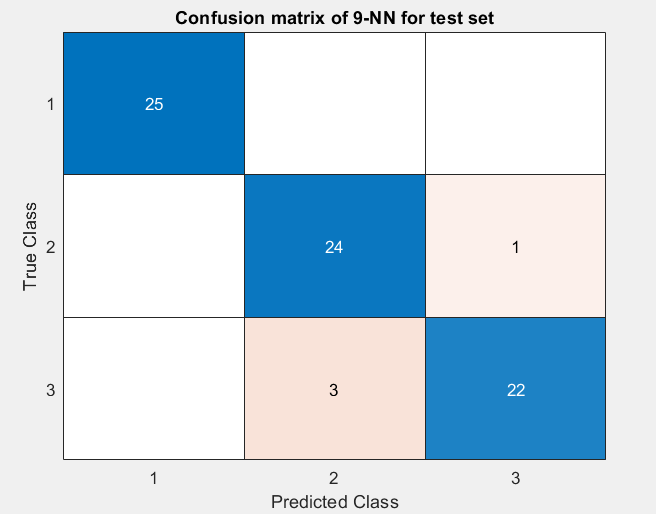
k =1



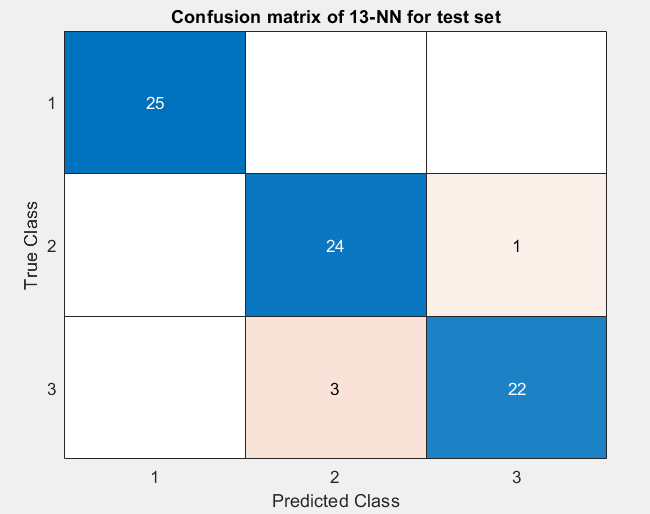
k=5



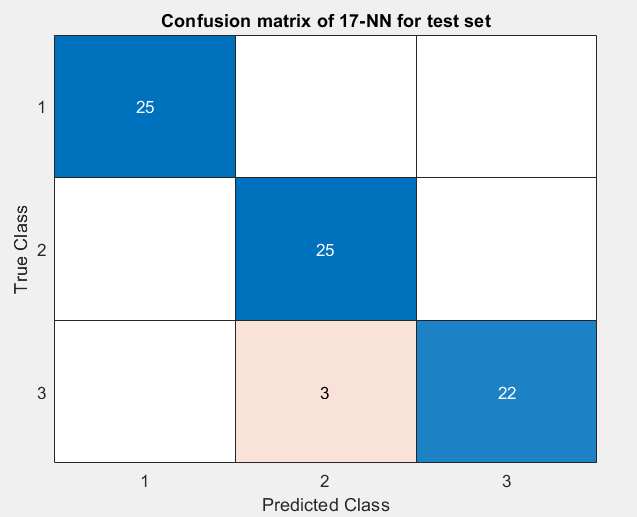
k=9



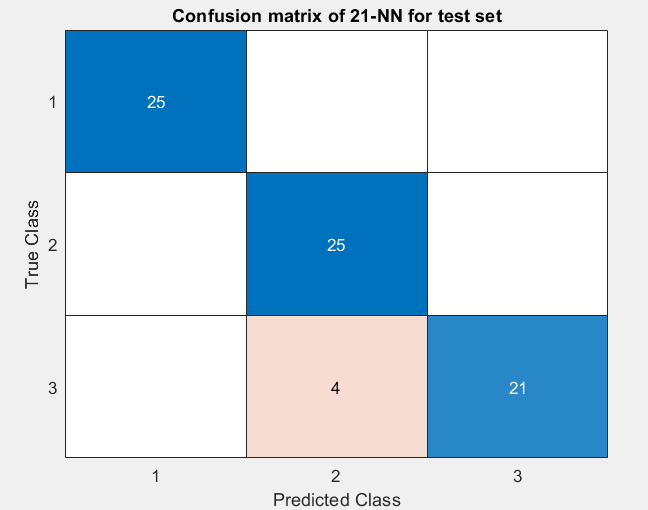
k=13



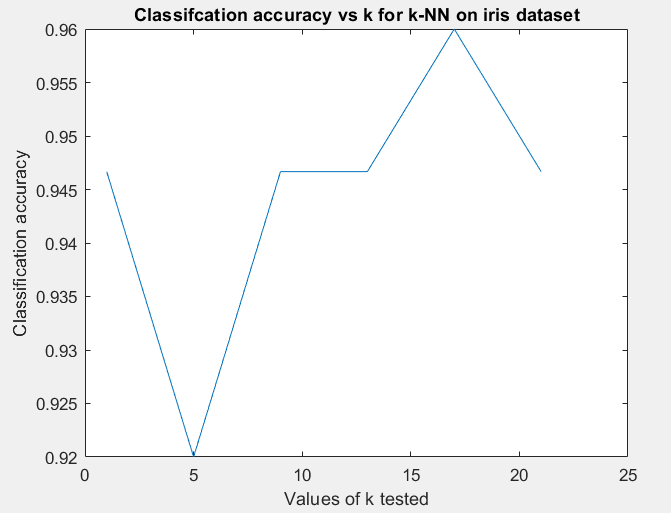
k=17



k=21



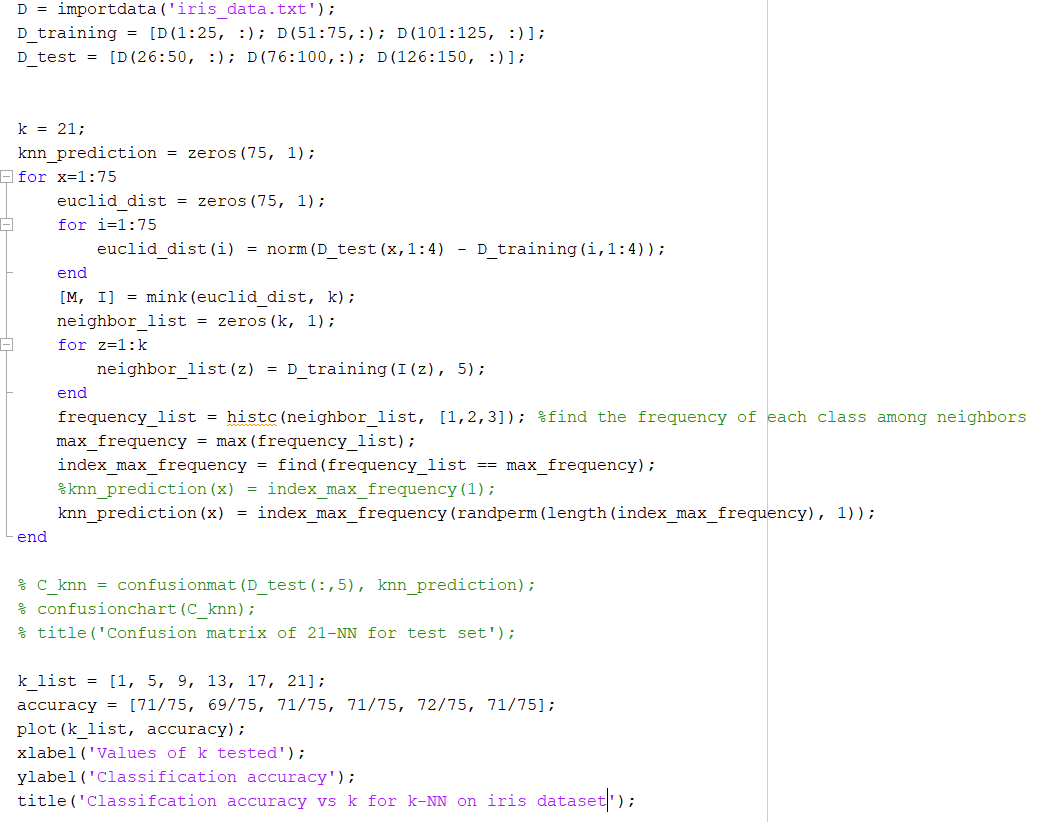
b)



Discussion:

The graph above shows that the optimal value of k for classification accuracy peaks at a certain value but isn’t necessarily always the largest or smallest k. As k increases, the classification accuracy can fluctuate. A small value of k is computationally inexpensive but may result in misclassification which is sensitive to noise. A large value of k is computationally expensive, but usually performs better until a certain inflexion point (k=17 in our example case). K value that is too large or too small may not accurately capture the underlying patterns of the features. Testing various values of k until the best classification accuracy is found is a good idea it seems.

Code used for problem 3:



Problem 4)

**SBS:**

Input: //total available samples

Output: //desired suboptimal subset features

Initialization: //Start with the full set of features

Termination: Stop when equals the number of features desired

**Step 1** (Exclusion) //Calculate the criterion function for each feature removed and choose best subset. Repeat until subset contains features

go to **Step 1**

**SBFS:**

Input: //total available samples

Output: //suboptimal subset of features

Initialization: //Start with the full set of features

Termination: Stop when equals the number of features desired

**Step 1** (Exclusion) //Calculate the criterion function for each feature removed and choose best subset.

**Step 2** (Conditional Inclusion) //Perform SFS if the corresponding subsets are better than previously evaluated

go to **Step 2**

go to **Step 1**