Assignment 3

Question 1: Row 22, 65, 97= missing value
Row 148 for the outlier.

Question 2 : This is my output
“

```
Predicted that test point  0  was  setosa
 but it is actually  versicolor

Predicted that test point  14  was  virginica
 but it is actually  setosa

The accuracy is:  86.66666666666667  percent
```

“

Question 3:
In our code, the double for loop is the most computationally demanding because it is running in runtime n^2. The need to compare each attribute in each instance to two different data sets requires this runtime.

```
 for x in test_data:
        close = -2
        closest_index = -1
        counter = -1
        for y in train_data:
            counter += 1
            distance = cosDistance(x,y)
            if close < distance:
                close = distance
                closest_index = counter

        predictions.append(train_labels[closest_index])
        print(train_labels[closest_index])

    return predictions
```

Question 4: If categorical data was introduced into our model we would have to differentiate between types using numbers for classification. This would change where the data is located on the matrices and how we measure using cosine distance.