BAIT 509 Assingment3
Liqiong Sun # 75956961

Exercise 1 (15%): Probabilistic Forecasting

Language: R

The diamonds dataset from the ggplot2 R package contains information about 53,940 diamonds. We will be using length, width, and depth (x, y, and zm respectively) to predict the quality of the cut (variable cut). Cut quality is categorical with five possible levels.
You own a shop that sells diamond, and you receive word of two new diamonds, with the following dimensions:

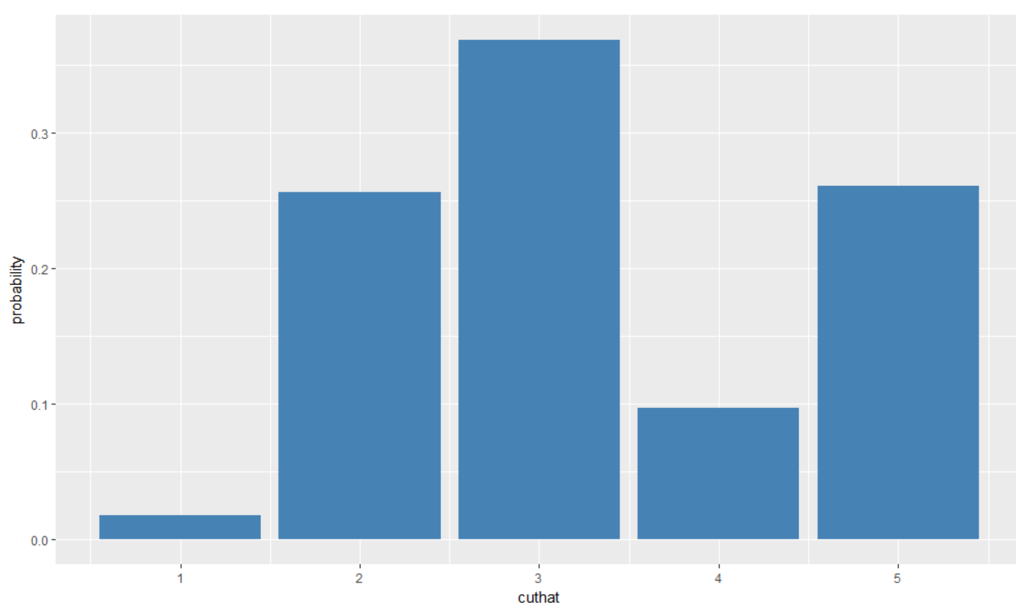
- Diamond 1: x=4, y=4, and z=3.
- Diamond 2: x=6, y=6, and z=4.

You can choose only one diamond to include in your store, but only have this information. You want the diamond with the highest cut quality.

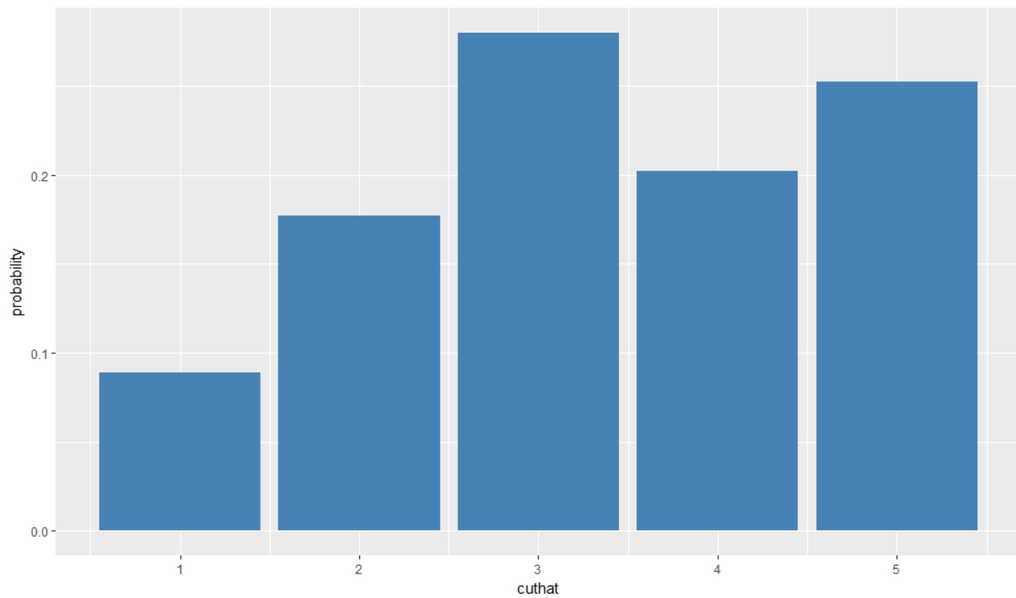
Answer the following questions.

1. **Produce a probabilistic forecast of the cut quality for both diamonds, using a moving window (loess-like) approach with a window width of 0.5. It's sufficient to produce a bar plot showing the probabilities of each class, for each of the two diamonds.**

Probabilistic Forecast Diamond 1



Probabilistic Forecast Diamond 2



2. **What cut quality would be predicted by a local classifier for both diamonds? Does this help you make a decision?**

   A local classifier would predict cut quality as 3 for both diamonds. Apparently, this does not help me make a decision as they have the same cut quality.

3. **Looking at the probabilistic forecasts, make a case for one diamond over the other by weighing the pros and cons of each.**

   Diamond 1 has higher probability of having cut quality 3 than diamond 2. However, it also has lower probability of getting cut quality higher than 3. Diamond 2 on the other hand, has higher probability of getting cut quality higher than 3 while its probability of getting the lowest cut quality is higher than diamond 1.

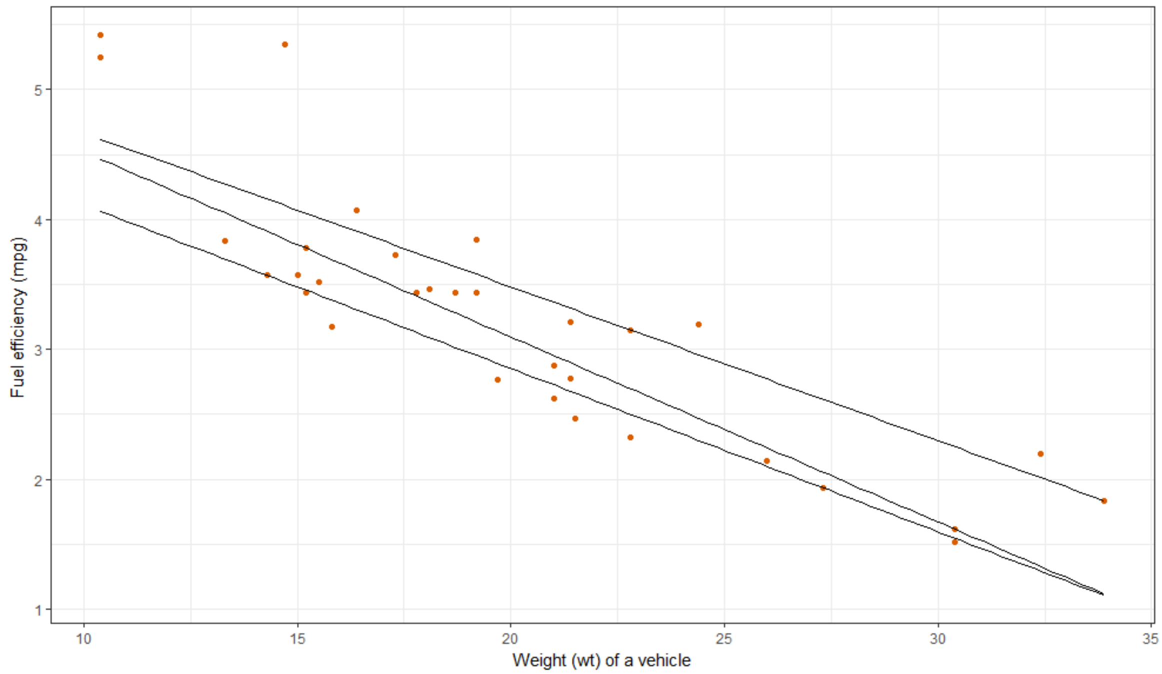
   Given the diamond 2 has higher probability of getting the lowest cut quality, if my purpose is to minimize the risk of getting a diamond with poor cut quality, I will prefer diamond 1. However, I will lose the opportunity of higher chance in getting cut quality over 3. If my purpose is to maximize the possibility of getting a diamond with cut quality higher than 3, than I will prefer diamond 2 over diamond 1. Meanwhile, I will bear higher possibility of getting the lowest cut quality.

BAIT 509 Assingment3
Liqiong Sun # 75956961

Exercise 2: Quantile Regression

2(a) (15%)

1. **Plot the data with the three lines/curves overtop. This gives us a "continuous version" of a boxplot (at least, the "box part" of the boxplot).**



2. **For a car that weighs 3500 lbs (i.e., wt=3.5), what is an estimate of the 0.75-quantile of mpg? Interpret this quantity.**

   The 0.75 quantile of mpg estimation is 20.23892 miles per gallon. It indicates that when a car weights 3500lbs, it has 25% chances of having fuel efficiency over 20.23892 miles per gallon.

3. **What problems do we run into when estimating these quantiles for a car that weighs 1500 lbs (i.e., wt=1.5)? Hint: check out your three quantile estimates for this car.**

   The quantile regression shows that a car weighted 1500 lbs has a 75% chance of getting fuel efficiency over 27.78833 miles per gallon and also has a 50% chance of getting fuel efficiency over 27.423 miles per gallon. The statement is illogical and the result is invalid.

2(b) (15%)

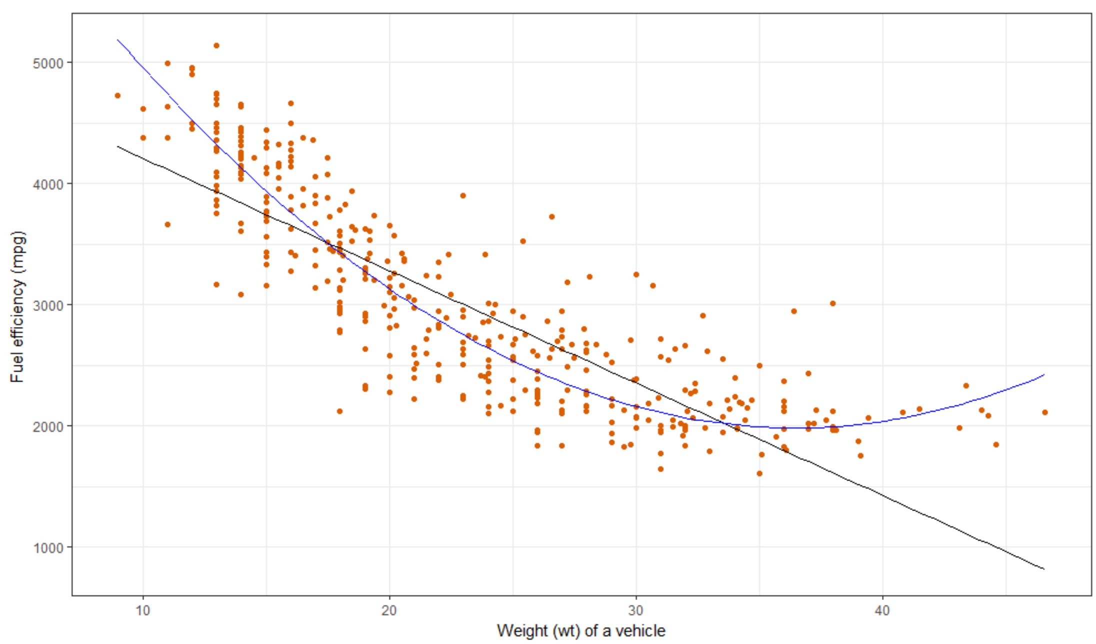
Let's now use the auto_data.csv data as our training data, and the mtcars data as a validation set. So, you can ignore your results from 2(a).
Note: In both mtcars and auto_data.csv, the fuel efficiency is titled mpg and have the same units as the mtcars data. The weight column is titled weight in the csv file, and is in lbs, but is title wt in mtcars, where the units are thousands of lbs.

1. **To the training data, fit the 0.5-quantile of mpg using the "weight" variable as the predictor. Fit two models: a linear model and a quadratic model.**

   auto_rq <- rq(mpg ~ weight, data=auto_data, tau=c(0.5))

   auto_lm <- lm(mpg ~ poly(weight, 2), data=auto_data)

2. Plot the two quantile regression curves overtop of the training data.



   The black line is the quantile regression curve of the quadratic model and the blue line represents the quantile regression of the linear model.

3. What error measurement is appropriate here? Hint: it's not mean squared error.

   Mean absolute error is appropriate here.

4. Using the validation data, calculate the error of both models. You'll first have to convert the "weight" variable to be in the same units as your training data.

The mean absolute error of the linear model is 2.894014 and of the quadratic model is 3.127625.

5. Use the error measurements and the plot to justify which of the two models is more appropriate.

   The plot form question3 shows that the linear model fits the data better than the quadratic model. Moreover, the mean absolute error of the linear model is less than that of quadratic model. Therefore, the linear model is more appropriate.

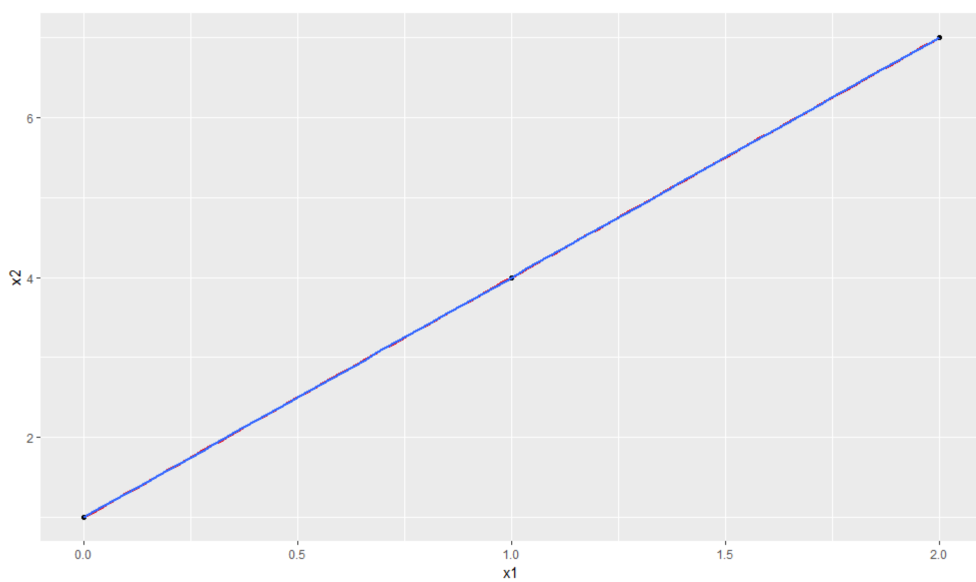
Exercise 3 (20%): SVM Concepts

Language: Only low-level programming is allowed here -- no machine learning packages are to be used.

From Section 9.7 in the ISLR book (that starts on page 368), complete the following:

- Question 1(a)

  a) Sketch the hyperplane 1 + 3X1 − X2 = 0. Indicate the set of points for which 1 + 3X1 − X2 > 0, as well as the set of points for which 1 + 3X1 − X2 < 0.



The blue lien represents 1 + 3X1 − X2 = 0

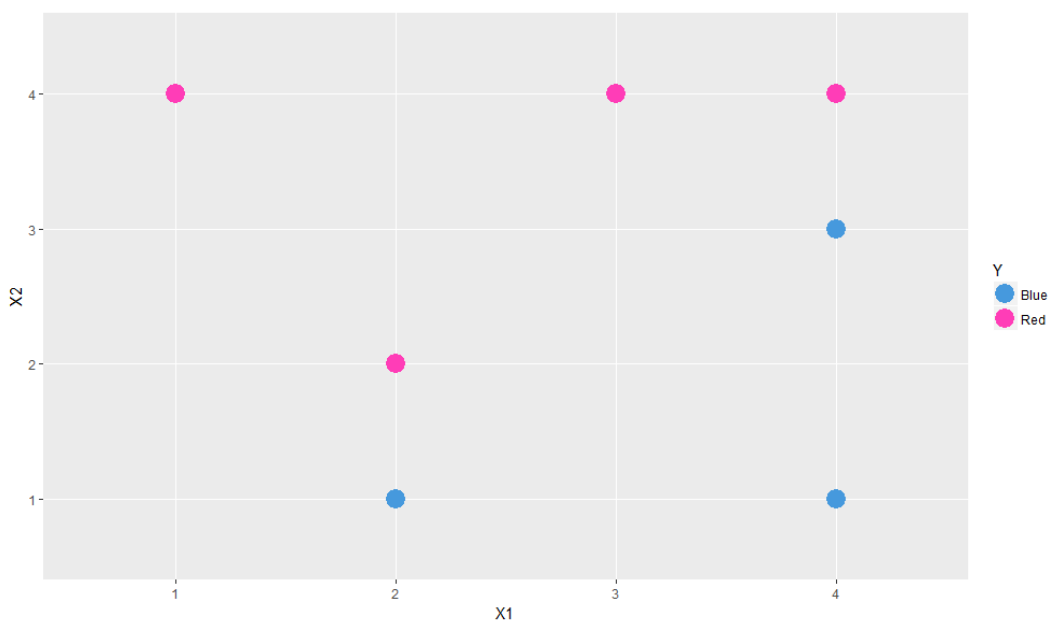
The set of points below the line is the set of points for which 1 + 3X1 − X2 > 0, the set of points above the line is the set of points for which 1 + 3X1 − X2 < 0.

- **3. Here we explore the maximal margin classifier on a toy data set.**

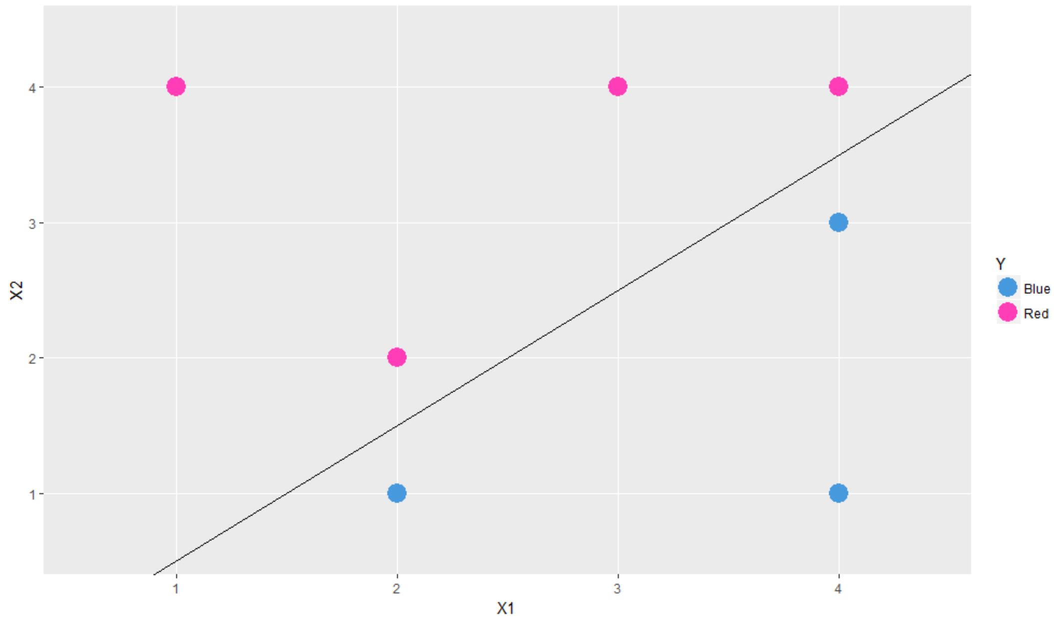
  (a) We are given n = 7 observations in p = 2 dimensions. For each observation, there is an associated class label.

Obs.	X1	X2	Y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

Sketch the observations.



- **(b) Sketch the optimal separating hyperplane, and provide the equation for this hyperplane (of the form (9.1)).**



The black line represents the optimal hyperplane: 0 =X_1-0.5- X_2

- **(c) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "Classify to Red if β0 +β1X1 +β2X2 > 0, and classify to Blue otherwise." Provide the values for β0, β1, and β2.**

  For points in the regions above the hyperplane, where 0 > X_1-0.5 -X_2, points are classified as red and for points in the regions below the hyperplane, where 0 < X_1-0.5-X_2, points are classified as blue. Therefore, **β0=-0.5, β1=1, and β2=-1.**

- **(d) On your sketch, indicate the margin for the maximal margin hyperplane**.

  The margin for the maximal margin hyperplane is 0.707.

- **(e) Indicate the support vectors for the maximal margin classifier.**

  The support vectors for the maximal margin classifier are (2,1), (2,2), (4,3) and (4,4)

- **(f) Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.**

  As the seventh observation is not the support vector of the maximal margin classifier and it is relatively far away from the maximal margin classifier, hence a slight movement of it would not affect the maximal margin hyperplane.