2016

Assignment 1 Report

THE ELEMENTS OF MACHINE LEARNING

AARON (ARUNA) TILLEKERATNE - 27345483

Part A Model Complexity and Model Selection

Question 1

Report the optimum value for K in terms of the testing error. Discuss the values of K corresponding to underfitting and overfitting based on your plot in fig.1.a.1.pdf.

The optimum value of K in terms of testing error was identified to be 11. See figure 1.a.1 below.

It seems as K approaches 1, the model becomes overfit, this is evident as the training error decreases rapidly while the testing error seems to increase after K = 11. Below this global minimum, (K >11) it seems the training error again increases, which is indicative of that the model is underfit. Please see iPython notebook for reference.

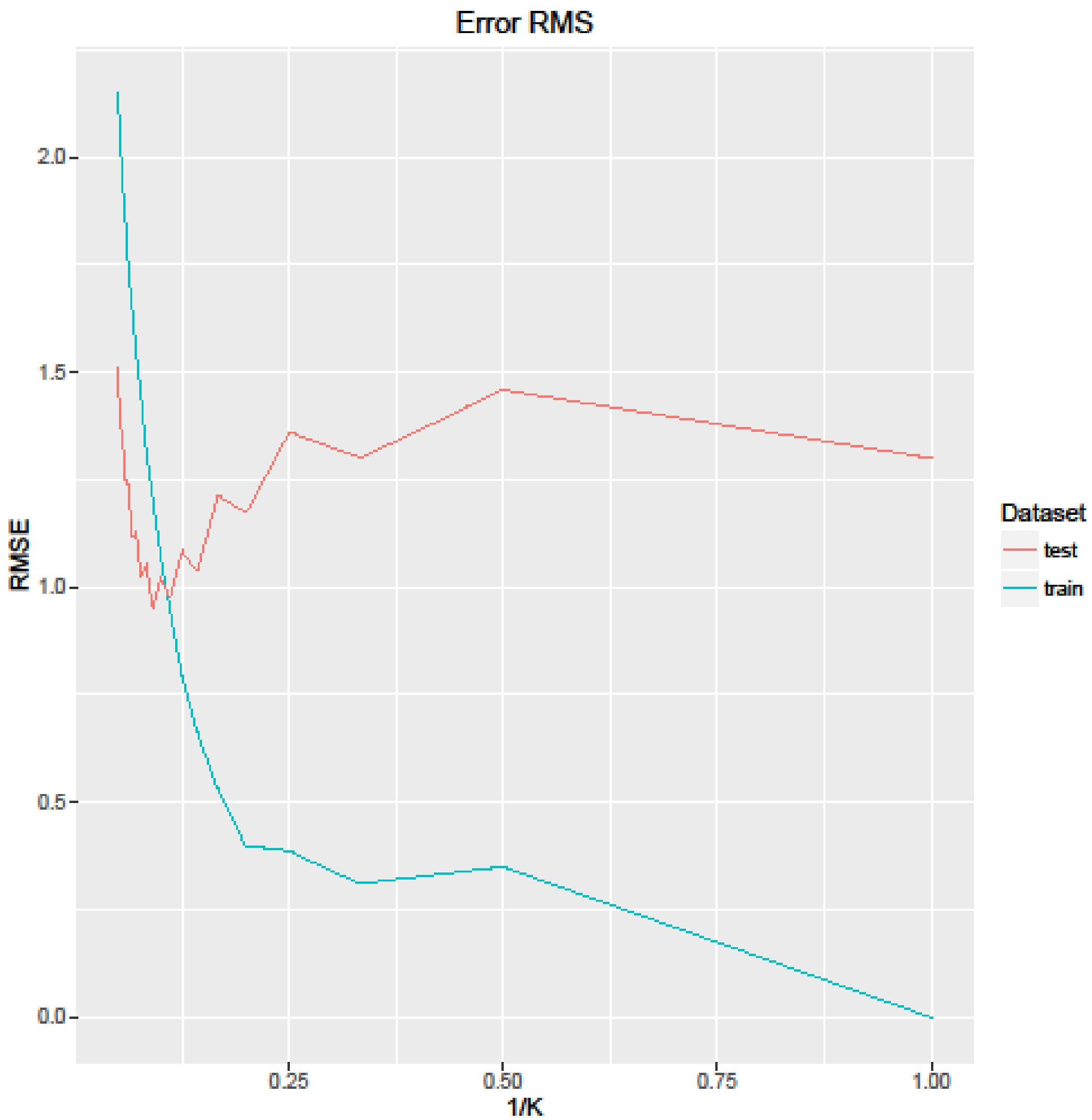

Figure 1: plot 1.a.1

Question 2

Report the values of K that result to minimum average error and minimum standard deviation of errors based on your cross validation plot in fig.1.a.2.pdf.

The k value resulting in the minimum average error was identified to be 4 and the smallest standard deviation in error was identified where K = 2. See fig1.a.2 below.

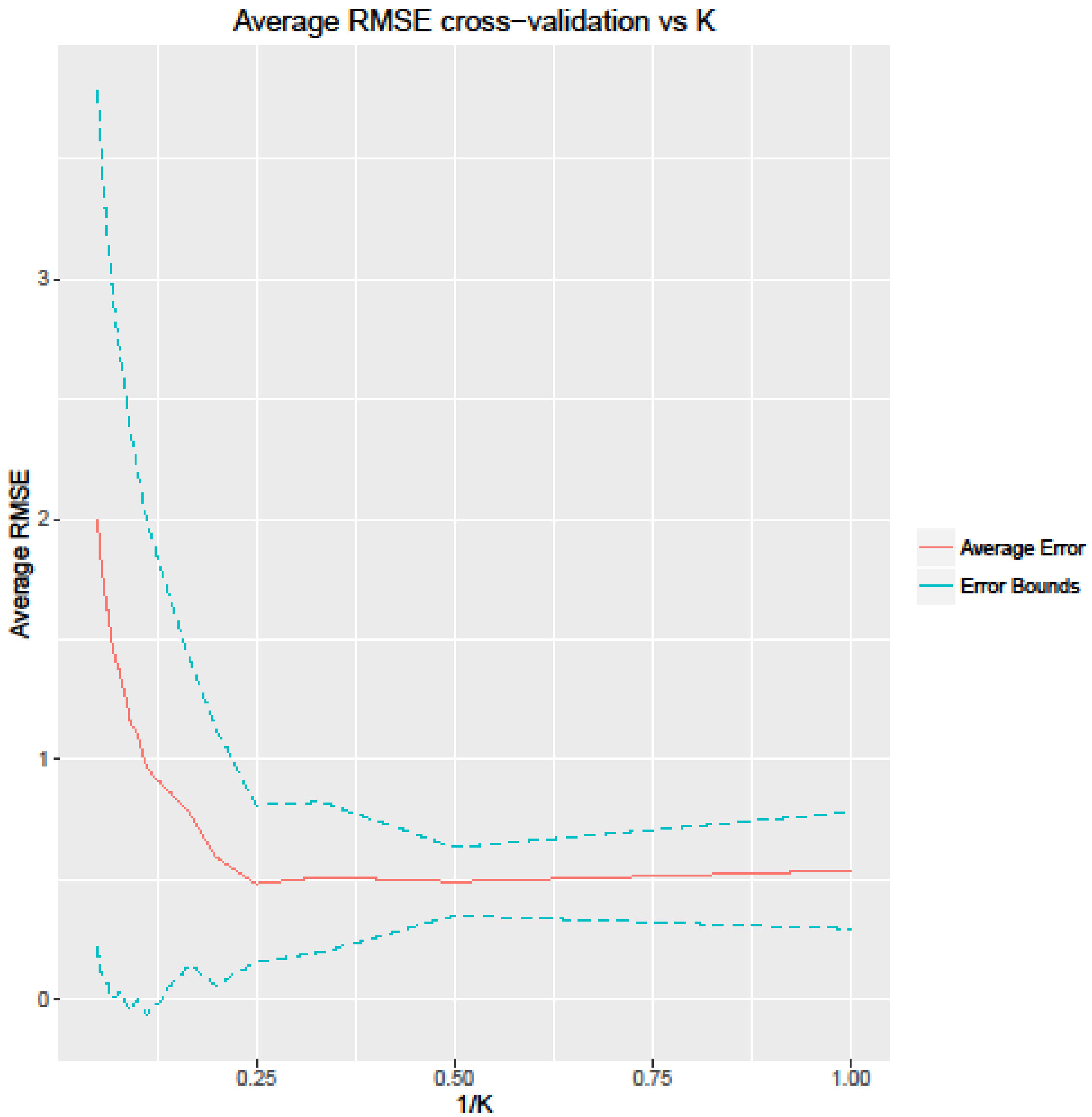

Figure 2:plot 1.a.2

Part B Prediction Uncertainty with Bootstrapping

Question 3

Based on fig.1.b.1.pdf,how does the test error and its uncertainty behave as K increases?

It is evident from the figure below that the testing error and uncertainty increase with the increase K value.

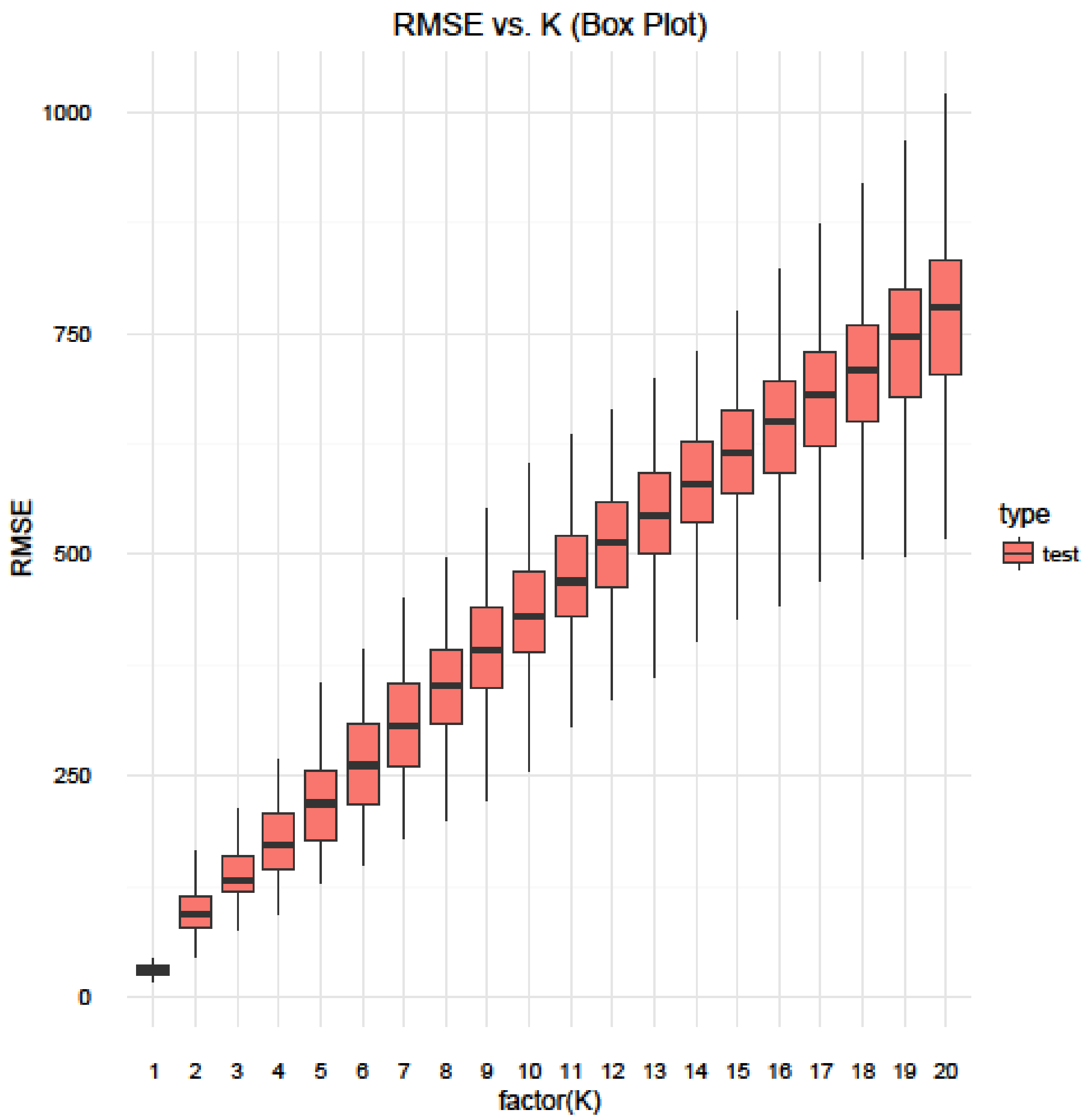

Figure 3: plot 1.b.1

Based on fig.1.b.2.pdf,how does the test error and its uncertainty behave as the number of subsets in bootstrapping increases?

It seems that the error seems to drop, but stabilises after the sampling time hits around 90. The uncertainty doesnt seem to be too greatly affected by this.

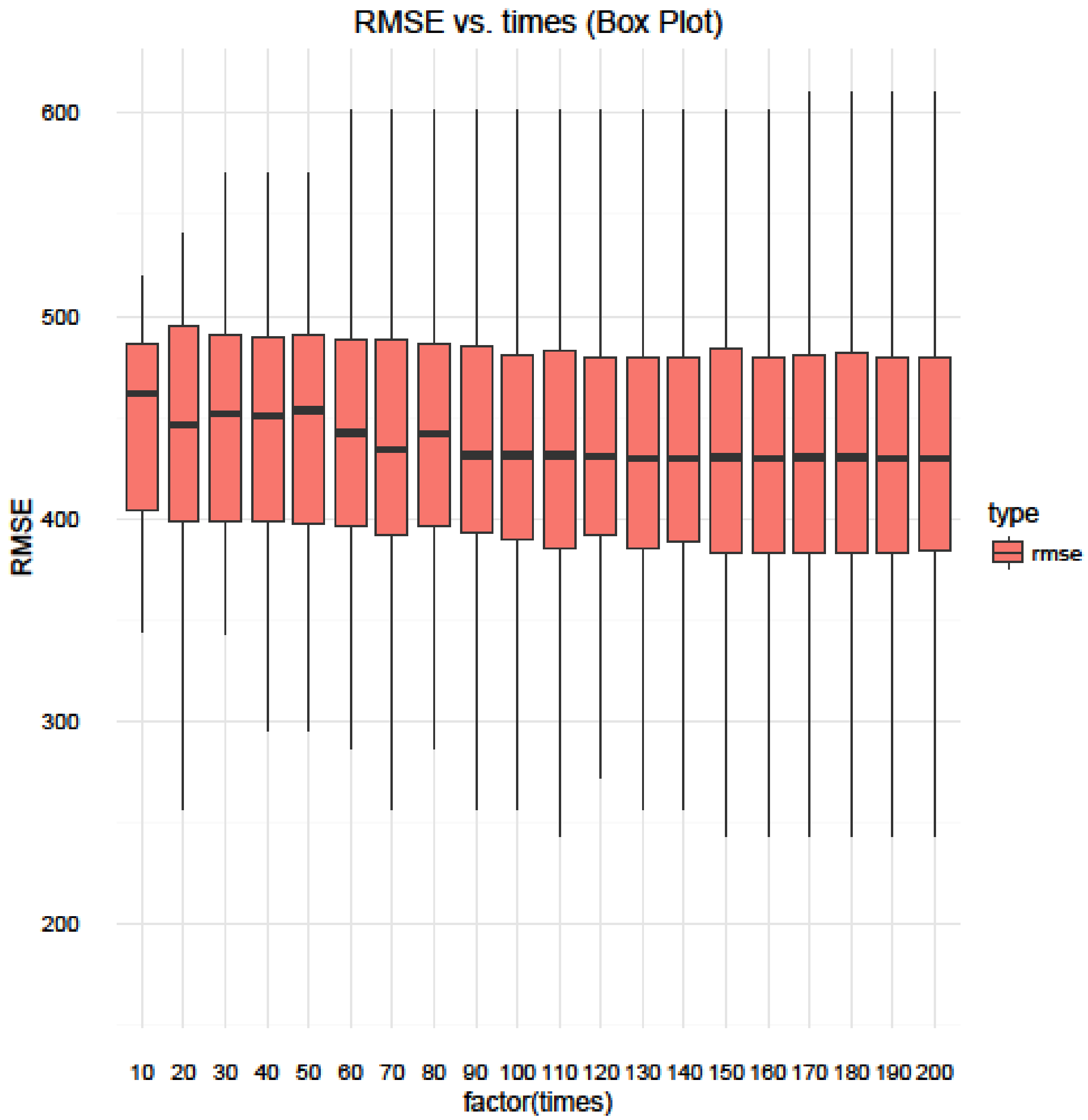

Figure 4: plot 1.b.2

Part C Probabilistic Machine Learning

Question 4

Recall the simple example from Appendix A of Module 1. Suppose we have one red and one blue box. In the red box we have 2 apples and 6 oranges, whilst in the blue box we have 3 apples and 1 orange. Now suppose we randomly selected one of the boxes and picked a fruit. If the picked fruit is an apple, what is the probability that it was picked from the blue box? Note that the chance of picking the red box is 40% and the selection chance for any of the pieces from a box is equal for all the pieces in that box.

$$P(blue|apple) = ?$$

$$P(blue) = 0.6$$

$$P(red) = 0.4$$

$$P(apple|blue) = 0.25$$

$$p(apple|red) = 0.75$$

$$P(blue|apple) = \frac{P(apple|blue) \times P(blue)}{P(apple)}$$

$$= \frac{P(apple|blue) \times P(blue)}{P(red) \times P(apple|red) + P(blue) \times P(apple|blue)}$$

$$= \frac{0.25 \times 0.6}{0.4 \times 0.75 + 0.6 \times 0.25}$$

$$= \frac{0.15}{0.45}$$

$$\therefore P(blue|apple) = \frac{1}{3} \approx 0.33$$

Question 5

As opposed to a coin which has two faces, a dice has 6 faces. Suppose we are given a dataset which contains the outcomes of 10 independent tosses of a dice: D:={1,4,5,3,1,2,6,5,6,6}. We are asked to build a model for this dice, i.e. a model which tells what is the probability of each face of the dice if we toss it. Using the maximum likelihood principle, please determine the best value for our model parameters.

$$P(D|W) = W_1 \times W_4 \times W_5 \times W_3 \times W_1 \times W_2 \times \left(1 - (W_1 + W_2 + W_3 + W_4 + W_5)\right)^3$$

$$f(w) = W_1^2 \times W_4 \times W_5^2 \times W_3 \times W_2 \times \left(1 - (W_1 + W_2 + W_3 + W_4 + W_5)\right)^3$$

$$\frac{d(w)}{dw} = 0$$

Take logarithm

$$\log f(w) = \log\left[W_1^2 \times W_4 \times W_5^2 \times W_3 \times W_2 \times \left(1 - (W_1 + W_2 + W_3 + W_4 + W_5)\right)^3\right]$$

$$\log f(w) = 2logW_1 + logW_4 + 2logW_5 + logW_3 + logW_2 + 3\log\left(1 - (W_1 + W_2 + W_3 + W_4 + W_5)\right)$$

$$\frac{dL(w)}{dw} = \frac{2}{W_1} + \frac{1}{W_4} + \frac{2}{W_5} + \frac{1}{W_3} + \frac{1}{W_2} + \frac{3}{\left(1 - (W_1 + W_2 + W_3 + W_4 + W_5)\right)} = 0$$

Therefore, estimated parameters from the data are as follows.

W_1 = 2/10

W_2 = 1/10

W_3 = 1/10

W_4 = 1/10

W_5 = 2/10

W_6 = 3/10