CPSC-6430 Machine Learning: Implementation & Evaluation
Homework 1: K-Fold Cross-Validation

Dineshchandar Ravichandran
C19657741

Contents

CPSC-6430-002-91763 Project-1
Author: Dineshchandar Ravichandran

Introduction
HW1 to implement the K-fold cross validation to choose the best regression algorithm from Liner, Quadratic and Cubic model.

1. Problem Statement

- To create a Python program to read the Women's Olympic 100m Sprint data from "W100MT.txt" file as per the user input and run this model through Liner, Quadratic and Cubic model.
- Evaluate the above models with K-fold cross validation with 5 folds.
- Generate a table as illustrated in HW prompt.
- Based on the K-fold validation of above models choose the model with least mean J value.
- Use the best model to predict and plot graph.

2. Project Screen Shot:
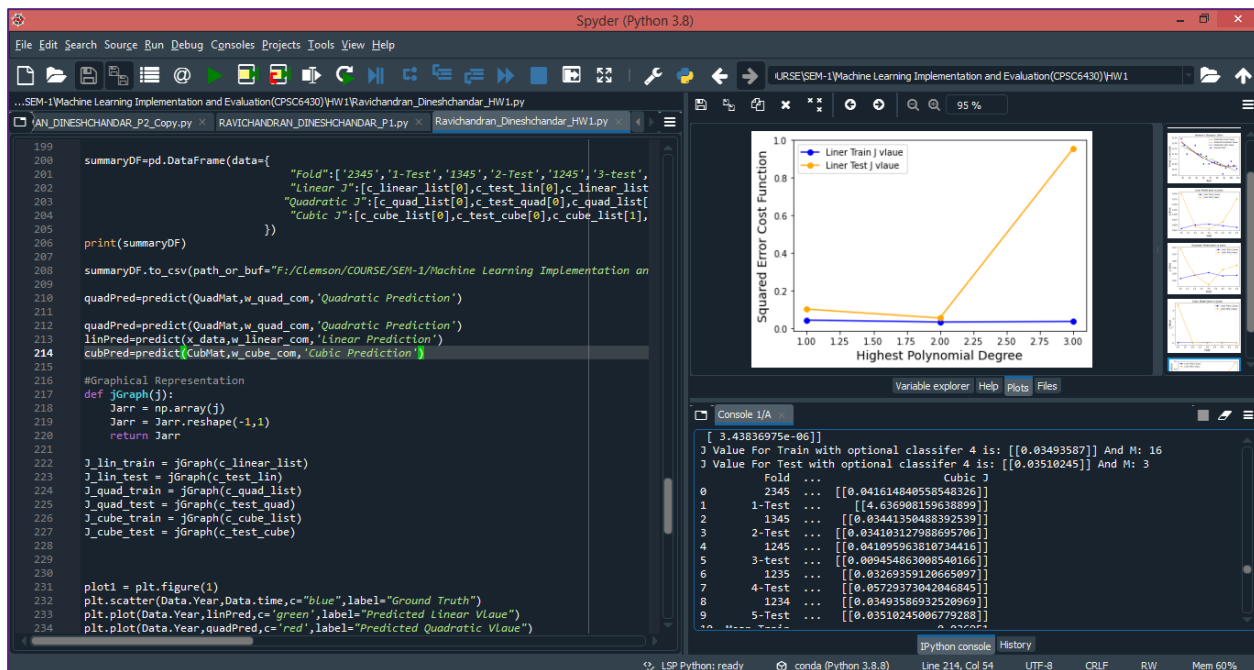
- The above screen shots represents the code in the SPYDER IDE, along with K-Fold Summary of Linear, Quadratic and Cubic model J Data in console (The data will be visible in CSV saved) and the graphical representation of J Test and J Train values for all three models.

CPSC-6430-002-91763 Project-1
Author: Dineshchandar Ravichandran

- Console Screen Shot for the same:

3. Project Input and Output

3.1. Input Data:

TXT file containing Women's Olympic 100m Sprint data from "1928 to 2008".

W100MTimes.txt

3.2. Output:

3.3. A Summary table

Fold	Linear J	Quadratic J	Cubic J
2345	[[0.01608517]]	[[1.18593825e-05]]	[[0.01559116]]
1-Test	[[0.37543116]]	[[0.00389656]]	[[0.23300417]]
1345	[[0.0620803]]	[[0.00011694]]	[[0.05359002]]
2-Test	[[0.09661618]]	[[8.37661028e-05]]	[[0.04560255]]
1245	[[0.07276117]]	[[7.06588189e-05]]	[[0.051885]]
3-test	[[0.0363714]]	[[0.00026969]]	[[0.0335505]]
1235	[[0.0784264]]	[[0.0001062]]	[[0.05769926]]
4-Test	[[0.01321215]]	[[0.0001316]]	[[0.01192363]]
1234	[[0.07260715]]	[[0.00012468]]	[[0.05693641]]
5-Test	[[0.04455494]]	[[3.66117388e-05]]	[[0.00566551]]
Mean Train	0.060392039	0.000086069	0.04714037
Mean Test	0.113237167	0.000883646	0.065949273

CPSC-6430-002-91763 Project-1
Author: Dineshchandar Ravichandran
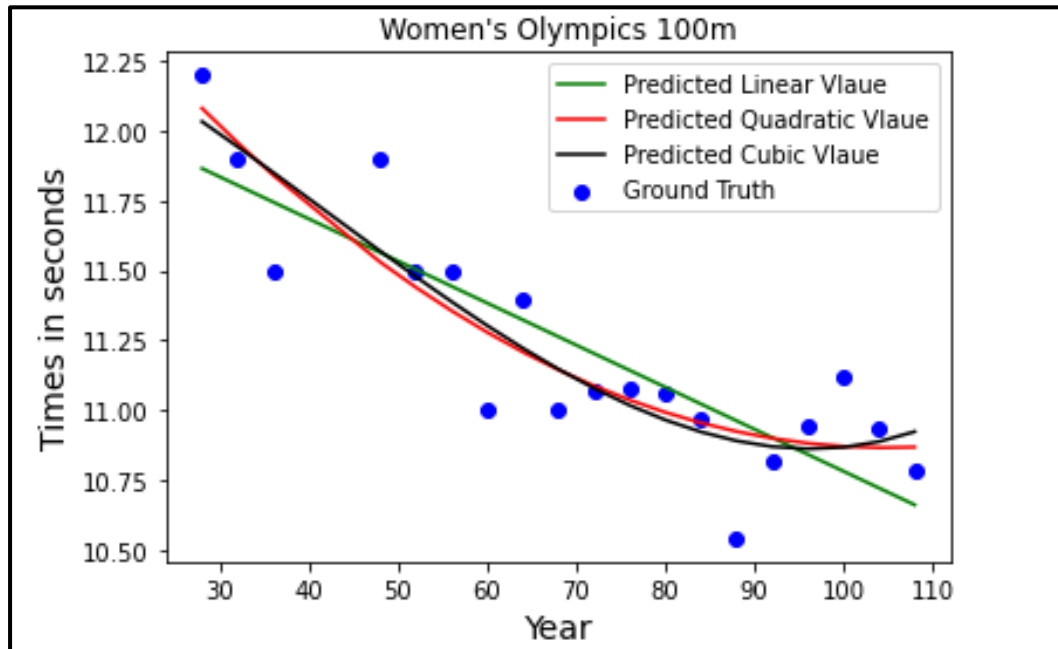
- From the above illustrated summary table and graph plot of prediction of Linear, Quadratic and Cubic model. We are able to summarize that the **Quadratic model** has the lowest mean test and train J value [J_test: 0.000086069, J train: 0.000883646] , as highlighted in yellow in Summary table.
- The below segments will cover the Weight Values for all the three models and the plot representation of Mean J training values and J testing values for all the 3 models.

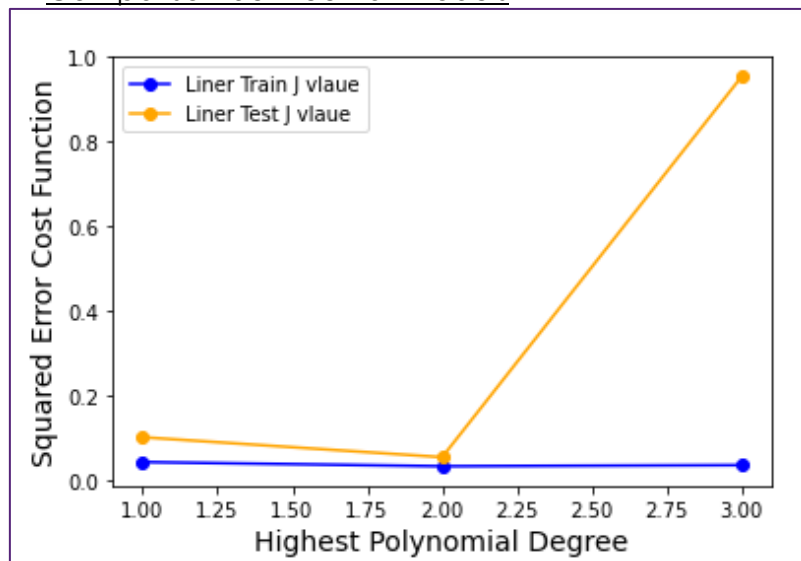### 3.3. C.   Weight Representation for all the three module

- hw(x){Linear}   = 12.28771135 + (- 0.01507181)(x)
- hw(x){Quadratic} = 13.1307195+ (- 0.0432482706)(x) + 0.00020636904 (x^2)
- hw(x){Cubic}   = 12.5156194 + (- 0.0102228099) (x) + (- 0.000322218021)(x^2) + (0.00000259521291) (x^3)

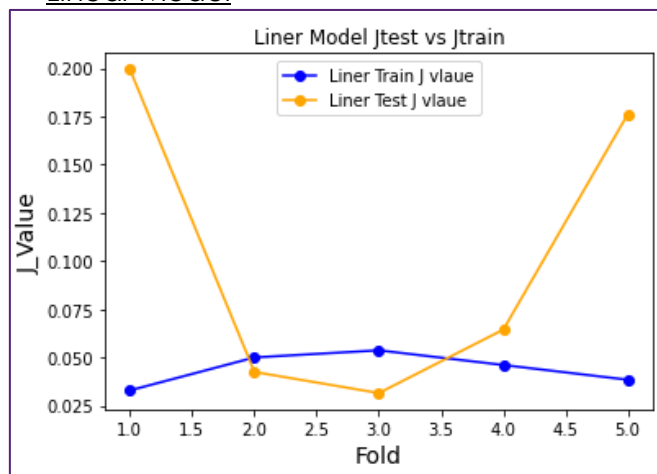### 3.3. D.   Weight Calculation using Best Model for entire data Set:

- Weight's calculated using "Quadratic Model" for entire women 100m data set are:
    - W0=13.1307195
    - W1= - 0.0432482706
    - W2= 0.000206369039

3.3. E.   J Test VS J Train Graph Representation
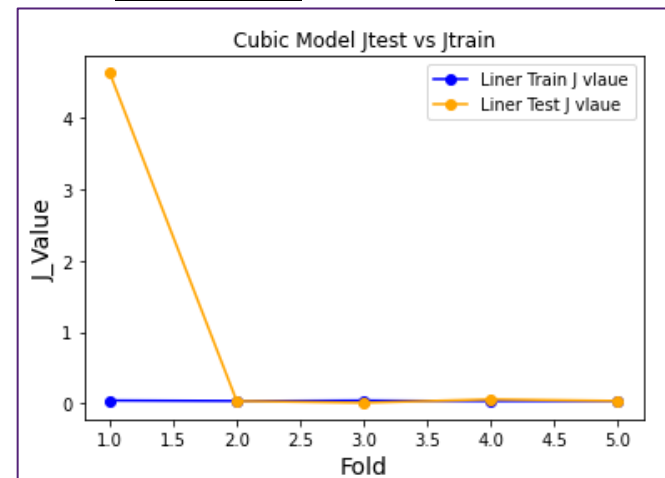
<u>Comparison between all models</u>

<u>Linear Model</u>

<u>Cubic Model</u>

<u>Quadratic Model</u>

CPSC-6430-002-91763 Project-1
Author: Dineshchandar Ravichandran