

# **CPSC-6430 Machine Learning: Implementation & Evaluation**

## **Homework 1: K-Fold Cross-Validation**

**Dineshchandar Ravichandran  
C19657741**

## Contents

Introduction.....	3
1. Problem Statement.....	3
2. Project Screen Shot: .....	3
3. Project Input and Output .....	4
3.1. Input Data: .....	4
3.2. Output: .....	4

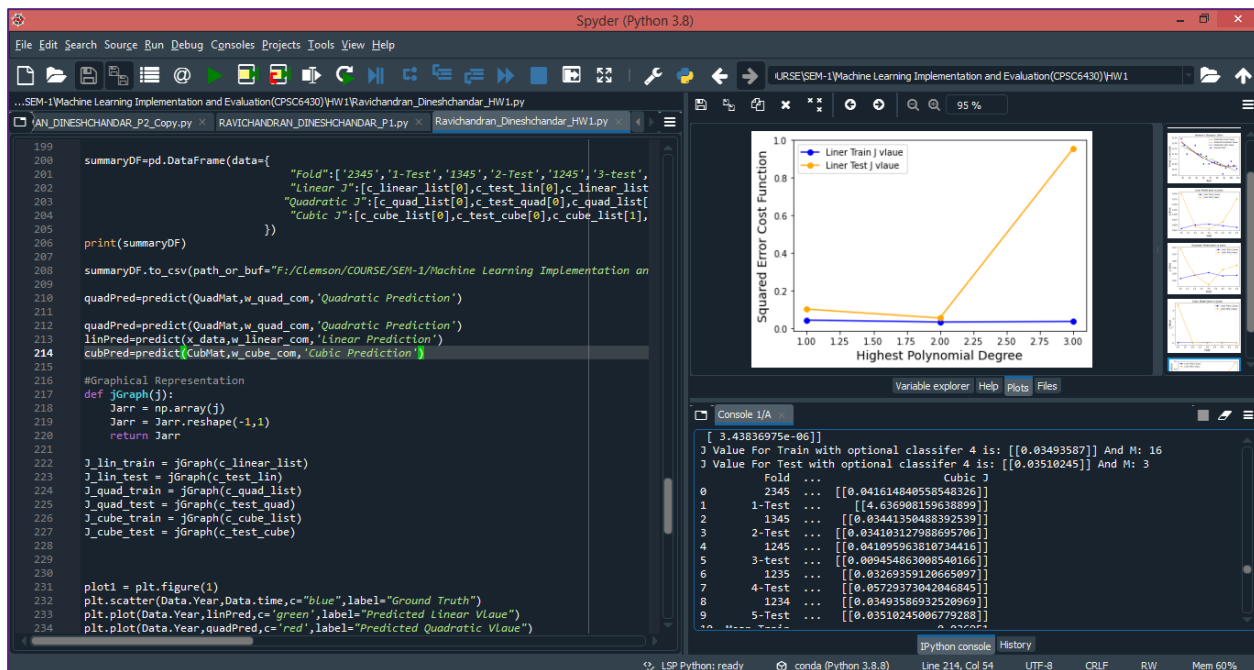
# Introduction

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.

- Console Screen Shot for the same:

```

J value for train with optional classifier 4 is: [[0.03493307]] And R: 10
J Value For Test with optional classifier 4 is: [[0.03510245]] And R: 3
Fold ... Cubic J
0 2345 ... [[0.041614840558548326]]
1 1-Test ... [[4.636908159638899]]
2 1345 ... [[0.03441368408392539]]
3 2-Test ... [[0.034103127988695786]]
4 1245 ... [[0.041095963810734416]]
5 3-test ... [[0.009454863008540166]]
6 1235 ... [[0.03269359120665097]]
7 4-Test ... [[0.05729373042046045]]
8 1234 ... [[0.034935869325209609]]
9 5-Test ... [[0.03510245006779288]]
10 Mean Train ... 0.036951
11 Mean Test ... 0.954572

[12 rows x 4 columns]
Predicted Value For Quadratic Prediction is: [[12.08156127]]
[[11.95809676]]
[[11.04123600]]
[[11.5302768]]
[[11.43983133]]
[[11.35598967]]
[[11.27875183]]
[[11.20811779]]
[[11.14400756]]
[[11.08666114]]
[[11.03583852]]
[[10.99161972]]
[[10.95400479]]
[[10.92299354]]
[[10.8958617]]
[[10.8807826]]
[[10.86958285]]
[[10.8649869]]
[[10.86699476]]

Figures now render in the Plots pane by default. To make them also appear inline in the Console, uncheck "Mute Inline Plotting" under the Plots pane options menu.

```

### 3. Project Input and Output

#### 3.1. Input Data:

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



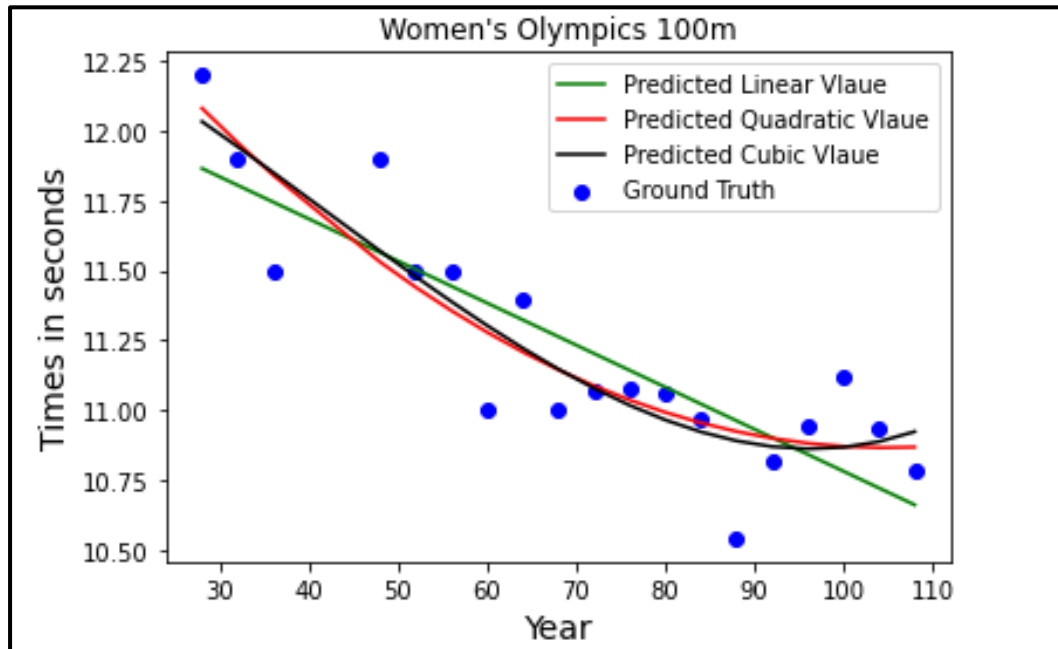
**W100MTimes.txt**

#### 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

### 3.3. A.1. Prediction Graph Representation



- 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 [ $J_{\text{test}}$ : 0.000086069,  $J_{\text{train}}$ : 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

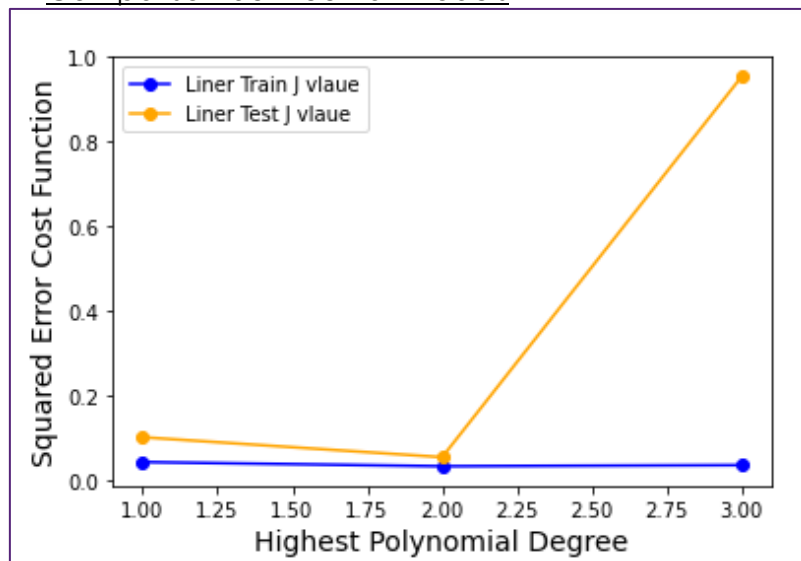
- $hw(x)\{\text{Linear}\} = 12.28771135 + (-0.01507181)(x)$
- $hw(x)\{\text{Quadratic}\} = 13.1307195 + (-0.0432482706)(x) + 0.00020636904(x^2)$
- $hw(x)\{\text{Cubic}\} = 12.5156194 + (-0.0102228099)(x) + (-0.000322218021)(x^2) + (0.00000259521291)(x^3)$

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

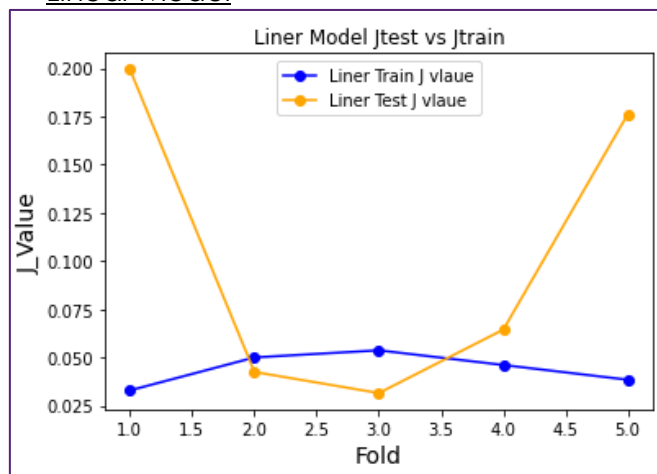
- Weight's calculated using "Quadratic Model" for entire women 100m data set are:
  - $W_0 = 13.1307195$
  - $W_1 = -0.0432482706$
  - $W_2 = 0.000206369039$

### 3.3. E. J Test VS J Train Graph Representation

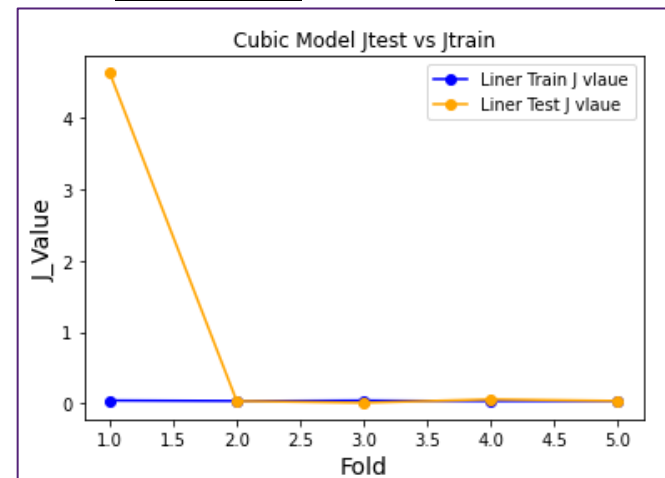
Comparison between all models



#### Linear Model



#### Cubic Model



#### Quadratic Model

