

# **SML Assignment 4**

# Question 1

For this question, the very first thing we have to do is to divide the data into training data and testing data. After doing that, I have performed k-fold test. Some of the results follows :

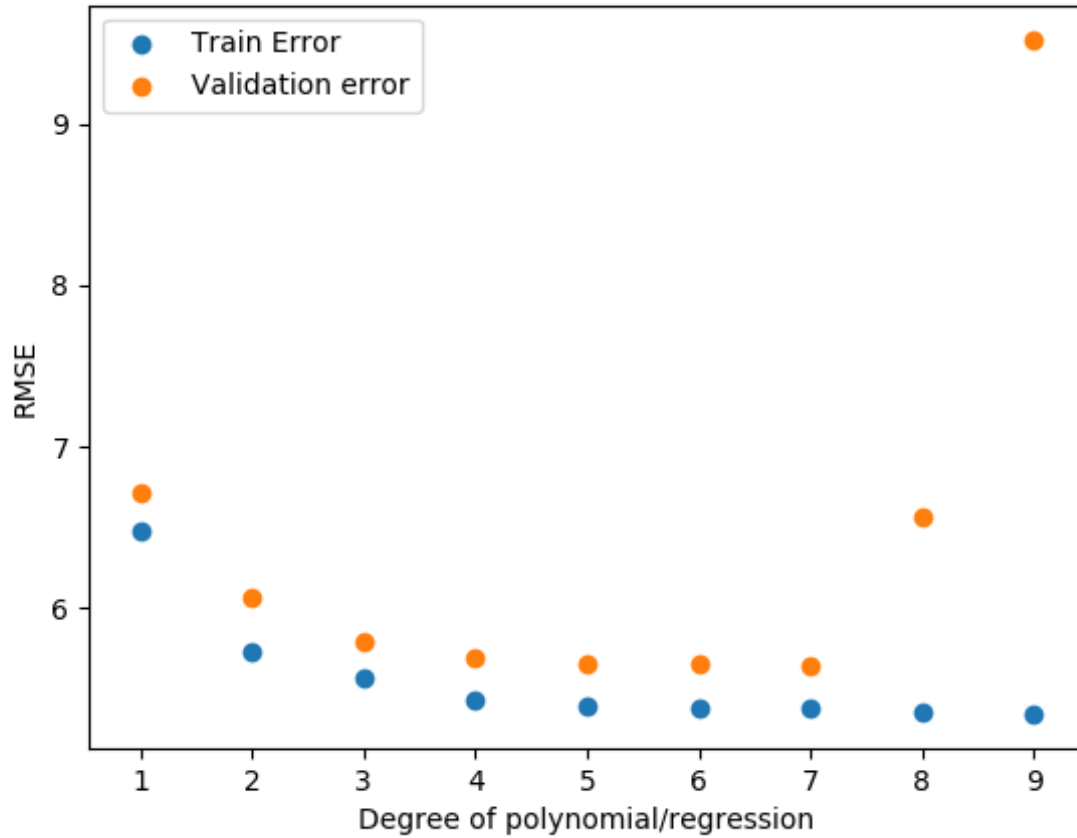
RMSE of training data of all 13 elements :

```
RMSE of training data = [[8.82883005]
[8.78386668]
[8.52632646]
[9.15585344]
[8.76328206]
[6.14585888]
[8.84525089]
[9.1736052 ]
[9.00289484]
[8.70355384]
[8.28968396]
[9.16047201]
[6.52666796]]
```

Similarly, RMSE of testing data of all 13 elements :

```
RMSE of testing data (20%) = [[4.62221533]
[5.18539992]
[4.91739483]
[5.18539992]
[4.46886002]
[5.03034064]
[4.58094656]
[4.45623141]
[4.97133698]
[4.91971789]
[5.16193037]
[4.38872091]
[3.79234266]]
```

After doing that, I have performed 5-folds on the training data with different degree of regression (polynomial). Graph for error on performing 5-fold is :



We can see that at degree of polynomial = 7, we are getting best results for validation error. So,  $m = 7$  is the best option for the degree of regression.