Exercise 0:** Which polynomial has the lowest train MSE? Which one has the lowest test MSE?* ***[0.5 mark]
Answer:

Order 9 has minimum value MSE for training (7.737735706301106e-27) and

Order 3 has minimum value MSE for testing (0.016317893338617738)

Exercise 1:** What trend do you observe when you analyse the dependence of train and test MSE on the polynomial order? First describe the observed trends, and then explain them.* ***[1 mark]
Answer:

The Mean Square Error (MSE) for both training and testing data sharply decreases with 3rd degree polynomial and reaches to an approximate value of 0.02. This trend of lower MSE continues with both Training and Testing MSE diverging in opposite directions as we increase the degree of polynomial until degree 8 and then overshoots for polynomial degree 9 and train MSE is 0 and test MSE is very high.

The convergence can be seen at polynomial degree 3 and is the best fit, before this the model is too rigid and at degree 9 we can see overfitting as the model has memorized all the points instead of learning.

Exercise 2:** Identify the models that are suffering from <u>under-fitting</u> and the ones suffering from <u>over-fitting</u>. Justify your choice based on your observations and use the theory that we have learnt to explain it.* ***[0.5 mark]
Answer:

Underfitting can be seen with first degree polynomial the model is too ridged and generalization for newer data is poor as the MSE for train is high.

Overfitting can be seen after degree 7 polynomial as the model is too generalized and as the model tries to fit all the point of the given Training data the learning capability is lost instead it is trying to memorize the training data, which is why it fails miserably and we see higher Test and Training MSE training.

Exercise 3:** Which model would you pick as the best one amongst these 9 models? What are the parameter(s) and hyper-parameter(s) of your chosen model?* ***[0.5 mark]
Answer:

Polynomial model of degree 3 and the model is

y = 1.326 x^3 - 0.09076 x^2 - 0.1449 x + 1.004

PARAMETERS: w3 = 1.326, w2 = 0.09076, w3 = 0.1449

HYPER-PARAMETERS = degree 3 polynomial

Exercise 5: Focus on order=9. Describe AND explain the trend of each of the metrics below with respect to increasing values of λ (that is, first describe what the effect of increasing λ from zero upward is it on the parameter in question and then explain briefly and clearly the reasons behind it): [1 mark]
(a) TRAIN MSE
(b) $\vec{w}^T.\vec{w}$
(c) TEST MSE

Answer:

a) Effect of λ on Train MSE, with degree 9 polynomial as we increase the penalty, the Training MSE increases. Initially when the λ is 0 the model is not learnt it has memorized all the data points and therefore Train MSE is 0. But as we penalize the model by increasing the λ, overfitting decreases and thus Train MSE also increases, the model is less overfitted than before and more generalized.

b) As the value of λ increases by a multiplication factor of 10 we see a drastic decrease in the value of $\vec{w}^T.\vec{w}$. As $\vec{w}$ are the parameters which needs to be tuned in-order for model to perform better, for degree 9 polynomial, initially as the model was overfit this was high, but as the penalty is added the loss function is minimized and we see tuned $\vec{w}^T.\vec{w}$

c) Effect of λ on Test MSE, with degree 9 polynomial as we increase the penalty, we see a sharp decrease upto 0.0100 and then a gradual increase in Test MSE. This is due to the fact that, when there was no penalty the model was highly overfitted and it performed poorly on Test data but as we penalized the model with increasing lambda the model started to learn and overfitting decreased until it reached λ is 0.0100, after that the model starts to generalizes more and underfitting is seen.

Exercise 6: Now suppose that instead of 10 training instances, we had access to 100 train instances. Run the following script and inspect the change in the test error. Describe and explain the effect of having more training data on the test error (test MSE) and over-fitting. [1 mark]

Answer:

When we increase the training samples from 10 to 100, test MSE starts to increase slowly for all the orders of polynomial with increase in λ. As large amount of data is provided, higher order polynomial models learns better than lower order as they are less tunable and better flexibility is seen. But as we penalize the model with increasing λ an opposite trend is seen, although this trend is very small. And we clearly see the impact of more training samples as there is very less over-fitting.

Exercise 7: Which is the correct way to complete the script and calculate the variable SSE? Explain what each of the three options to calculate SSE would do and justify your choice. What is the resulting train and test errors? [0.5 mark]

Answer:

To choose SSE we would use Training Data

SSE = np.sum(np.power(ytrainset - np.matmul(phitrain, w_map), 2))

We would choose SSE from Training data as enough (75%) of data is provided into model and validation data is used in-order to calculate the best hyperparameter(λ) which is 0.100 and Test MSE would be based on the selected hyperparameter.

Thus, Training MSE is 0.69638, Test MSE is 0.19468.

Exercise 8: Compare the training, validation and the test errors in both the linear model and the M5P model. First, explain the differences between the numerical values of each error separately for the linear model and for the M5P model. Then, bearing in mind that the M5P model is more complex than the linear model, explain why the numerical values of the errors seem to behave differently for the linear model and for the M5P model. [1 mark]

Answer:

In Linear Model the values over Training set, 2-fold cross validation(CV), and 50% Split over test and train has an increasing trend, the reason is quite obvious, in case of Training Set, the same data is used for Testing, which is why the MAE is lower here as the model is already aware of the test data. The 2-fold CV has a reliable MAE and RMSE as the data is shuffled and split into two and Training and Validation is done, and the error results are summed up, whereas in case of 50% Test-train split only 50% (4540 instances) of training data is used to predict the model which is why it has the largest MSE/MAE. Similar trend of increasing order for errors over Training set, 2-fold CV and 50% Test-train split is observed with M5P model.

M5P model has lower MAE and RMSE than Linear Regression model over all three different test options. This is because of the fact that, unlike linear regression which fits the data on to a single model, M5P is a decision tree which divides the prediction space on basics of standard deviation and for each region a different linear regression model is used to predict the test data and this is the reason M5P performs better than Linear regression.