

Housing Price Prediction

Question-1:

Rahul built a logistic regression model having a training accuracy of 97% while the test accuracy was 48%. What could be the reason for the seeming gulf between test and train accuracy and how can this problem be solved.

ANSWER:

Overfitting is the main reason for it, as the model memorizes the training data and it's not performing well on test data set.

Solution – Polynomial regression, regularization and selecting correct variables.

Question-2:

List at least 4 differences in detail between L1 and L2 regularization in regression.

L1 –Lasso regression

1. useful in shrinkage and selection of parameters
2. It reduces less important feature coefficient to zero
3. Higher value of alpha causes under fit
4. It adds absolute value of coefficient as penalty
5. It takes much time for execution

L2 – Ridge regression

1. It adds squared coefficient as penalty
2. Higher value of alpha causes over fit
3. Takes lesser time for execution

Question-3:

Consider two linear models

$$L1: y = 39.76x + 32.648628$$

And

$$L2: y = 43.2x + 19.8$$

Given the fact that both the models perform equally well on the test dataset, which one would you prefer and why?

ANSWER:

Based on hyper parameter value, we can go with L1. Higher value of alpha causes over fitting in L2.

Question-4:

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

ANSWER:

We can use cross validation method to check whether model is robust and generalizable.

Question-5:

As you have determined the optimal value of lambda for ridge and lasso regression during the assignment, which one would you choose to apply and why?

ANSWER:

Lasso regression has added advantage of selecting parameters over ridge regression. So we will go with lasso regression. It also penalizes loss function based on absolute value of coefficient.