Problem Statement - Part II

Question 1

Rahul built a logistic regression model with a training accuracy of 97% and a test accuracy of 48%. What could be the reason for the gap between the test and train accuracies, and how can this problem be solved?

Answer:

It seems to be the problem of overfitting with logistic regression model where model accuracy for training is 97% i.e model is very well trained on training set, however if the unseen data feed to model its accuracy is lowered drastically. Overfit model has low bias but high variance.

To overcome this problem, we can try validation or cross validation and regularization techniques that uses cost function and take the coefficients of variables to 0 and reduce cost function.

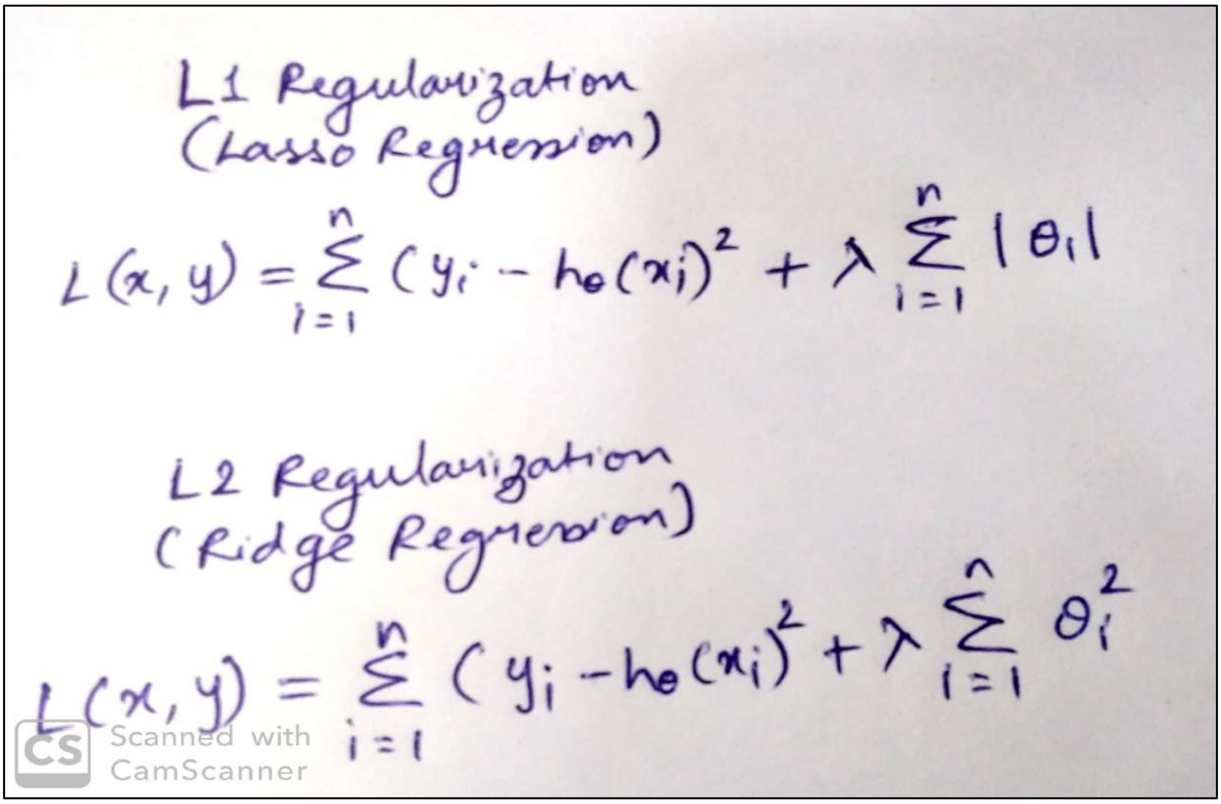
Question 2

List at least four differences in detail between L1 and L2 regularisation in regression.

Answer:

Below is the difference between L1 and L2 regularisation in regression:

L1 Regularization (Lasso Regression)	**L2 Regularization (Ridge Regression)**
L1 Regularization uses penalty as absolute value of magnitude of coefficients.	L2 Regularization uses penalty as square of magnitude of coefficients.
L1 Regularization has built in feature selection	L2 Regularization does not feature selection
L1 regularization forces the variables to be zero	L2 regularization forces the variables to be small but does not make them zero
L1 models are simple but cannot learn complex data.	L2 regularization can learn complex data.



L1 Regularization
(Lasso Regression)

$$L(x, y) = \sum_{i=1}^{n} (y_i - h_\theta(x_i)^2 + \lambda \sum_{i=1}^{n} |\theta_i|$$

L2 Regularization
(Ridge Regression)

$$L(x, y) = \sum_{i=1}^{n} (y_i - h_\theta(x_i)^2 + \lambda \sum_{i=1}^{n} \theta_i^2$$

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 data set, which one would you prefer and why?

Answer:

Given the fact that both the models perform equally well on the test data set,
Model, **L2: y = 43.2x + 19.8** will be preferred as it has less complexity and is robust.

Question 4

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

Answer:

Bias: Bias is error in model, when the model is weak to learn from the data. High bias means model is unable to learn details in the data. Model performs poor on training and testing data.

Variance: Variance is error in model, when model tries to over learn from the data. High variance means model performs exceptionally well on training data as it has very well trained on this of data but performs very poor on testing data as it was unseen data for the model.

It is important to have balance in Bias and Variance to avoid overfitting and underfitting of data.

Question 5

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:

It is important to regularize coefficients and improve the prediction accuracy also with the decrease in variance, and making the model interpretably.

Ridge regression, uses a tuning parameter called lambda as the penalty is square of magnitude of coefficients which is identified by cross validation. Residual sum or squares should be small by using the penalty. The penalty is lambda times sum of squares of the coefficients, hence the coefficients that have greater values gets penalized. As we increase the value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in final model unlike Lasso Regression.

Lasso regression, uses a tuning parameter called lambda as the penalty is absolute value of magnitude of coefficients which is identified by cross validation. As the lambda value increases Lasso shrinks the coefficient towards zero and it make the variables exactly equal to 0. Lasso also does variable selection. When lambda value is small it performs simple linear regression and as lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.