Advanced Regression Assignment Problem Statement - II Answers

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 1.

We can observe from the question that after building logistic regression model

training accuracy = 97% and

test accuracy = 48%

There is large difference between train and test accuracy, so we can say that **model is Overfitting.** This means that the model build is not generalizable from training data to test (unseen) data.

Overfitting is a phenomenon where a model becomes too specific to the data it is trained on and fails to generalize to other unseen data points

We can solve the problem of Overfitting by using

- **Cross Validation**

In this method we generate multiple mini train-test splits from training data. And then we tune our model on these splits. One such method is k-fold Cross Validation.

- **Regularization**

In regularization we try to make a model as simple as possible without compromising the performance on training data. Through regularization, we deliberately try to bring the complexity of model down. In other words, we try to have a model that is not too simple at the same time not too complex.

- **Remove (Dropping) irrelevant features**

Question 2

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

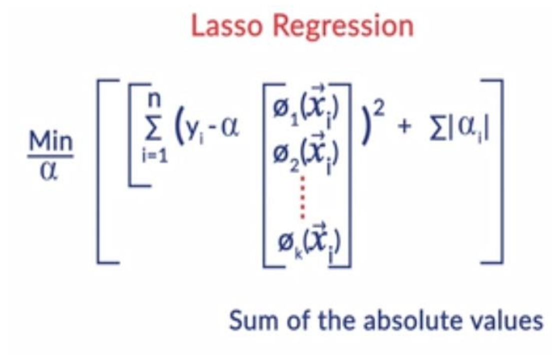
Answer 2.

L1 regularization also called as Lasso regression and L2 regularization also called as Ridge regression. Although they are both regularization techniques but there is some difference between them. Differences are as follows:

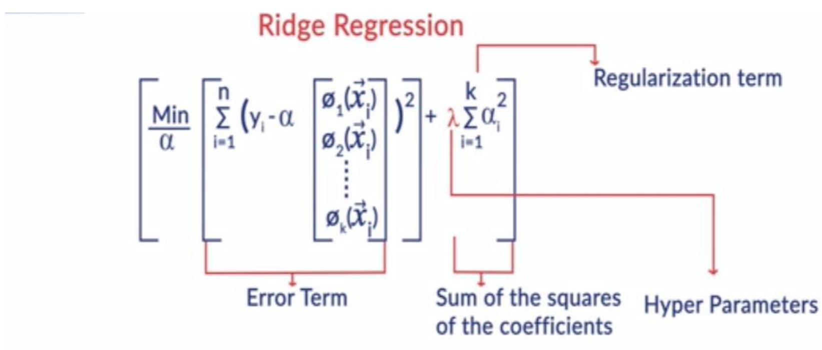
1. L1(Lasso) gives sparse models while
   L2(Ridge) gives non-sparse models.

Here, Sparse means many features has 0 coefficient value while others feature will have non-zero coefficient value.

2. The main difference between L1 and L2 regularization is in terms of penalty applied by them
   L1(Lasso) – "**Sum of the absolute value of the coefficients"** as regularization term



   L2(Ridge) – "**Sum of squares of the coefficients"** as regularization term



3. L1 (Lasso) shrinks down the coefficients of less important features/ redundant variables to zero, and hence it performs **feature selection.**
   L2 (Ridge) reduces the coefficients of features to arbitrarily low values but not zero, hence **no feature selection** takes place.

4. L1(Lasso) is robust to outliers, means it doesn't get affected by outlier's while
   L2(Ridge) is not robust to outliers, means it gets affected by outliers

5. L1(Lasso) is good for high-dimensional data, since it performs built in feature selection, whereas
   L2(Ridge) is not good for high-dimensional data.

6. L1(Lasso) can give more than one solution whereas
   L2(Ridge) can always give only one solution.

7. Shape of the constraint region is also a difference plays important role in L1(Lasso) and L2(Ridge)
For L1(Lasso) – The shape of the constraint region is diamond, and it uses L1 norm for constraint.
For L2(Ridge) – The shape of the constraint region is circle, and it uses L2 norm constraint.

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 3.

From the above two model L1 and L2, as mentioned that both the models perform equally well on test dataset, I would prefer **L2 model.**

We can represent these models in terms of their coefficients L1(36.76, 32.648628) and L2(43.2, 19.8)

The reason to choose L2 model is that it is simpler i.e. the coefficients looks simpler as compared to L1 model. And it takes a smaller number of bits to represent the coefficients of L2 model as compared to L1 model. Hence **L2 model is preferable.**

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 4.

1. Model is Robust and Generalizable

Generally, the simpler model is more robust and generalizable.

- The **robustness** of model is the property that tells that how effective is the model when it is tested on the new dataset of similar kind.
- And **generalization** of model refers to a model's ability to handle any random variations of training data.

If a model is not robust and generalizable then model may **Overfit** i.e. it performs very good on train data but its performance goes down when it is applied on test data.

A model should be able to handle any data set coming from the same distribution as the training data.

We can make sure that a model is robust and generalizable by avoiding a model to **Overfit.** To avoid the possibility of overfit we can use **cross validation & regularization technique** to avoid overfitting of model and **to make model more robust and generalizable.**

2. Implications of model robust and generalizable on accuracy of model

Simpler model is more robust and generalizable. The properties of simpler model are as follows:

- Simpler model is more generic than the complex model, and as a result generic models perform well of unseen data.
- The number of data points required for simpler model is far less.
- A simple model bound to make more errors

More accurate model leads to overfitting. A flexible one will learn from data a lot. So usually avoiding accurate model and preferring a robust one is better.

Because of this there is an **inverse implication of model being robust and generalizable on accuracy of model.** Because simpler model makes more error and more error means less accuracy.

So, to overcome this we should have a **bias-variance tradeoff**. In this we try to build a model in such a way that a model accurately captures the regularities of training data and also it is robust and generalizable to unseen data. It means we try to have **less bias and less variance in the model.**

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 5. The evaluation metrics obtained for Ridge and Lasso regression are as follows:

	Best (optimal) Alpha	R2_train	R2_test	AIC	BIC	Adjusted_R2_test
Ridge Regression	**6**	**0.939**	**0.914**	**531.66**	**8772.81**	**0.742**
Lasso Regression	**50**	**0.939**	**0.914**	**311.67**	**8124.21**	**0.864**

From the above table we can observe that –

- **Optimal value of lambda(alpha)**

  For Ridge regression is – 6

  For Lasso regression is – 50

- **For Ridge and Lasso regression R2_train and R2_test is same**
  For Ridge regression R2_train – 0.939 and R2_test – 0.914
  For Lasso regression R2_train – 0.939 and R2_test – 0.914
- **For Ridge Regression**
  AIC - 531.66
  BIC - 8772.81
  Adjusted_R2 - 0.742
  For Lasso Regression
  AIC - 311.67
  BIC - 8124.21
  Adjusted_R2 - 0.864

Based on the above metrics it is clear that **Lasso Regression with optimal alpha value – 50 is better**. And we will **choose Lasso regression** to apply for model building.