Advanced Regression

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?

Answer1
There are two scenarios in which this problem can occur.

Scenario1: Non-Regularized Logistic Model:

This clearly indicates that the model has overfit which is performing exceptionally good in training set but then it fails to generalize hence it doesn't perform well in test set.
We can try building a simpler model than the current one by reducing the number of features or by creating derived features that makes more sense and dropping the original features that has been factored into the derived features.

Scenario2: Regularized Logistic Model:

This again indicates model has overfit and we have probably used low value of alpha that has resulted in building a very complex model.

We can try increasing the value of hyperparameter - Alpha so that the coefficients are penalized more in case the learning algorithm builds a more complex model.

Adding more data can solve the problem or reducing the number of features can also help.

Question 2

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

Answer 2.

1)

L1 also known as Lasso regularization uses the error term + sum of the absolute values of the coefficients as represented in below expression(Yellow highlighted portion).

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Whereas,

L2 also known as Ridge regression uses error term + sum of the squares of the coefficients as the regularization term(Yellow highlighted portion). As mentioned below:

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

2)

L1 gives sparse solution which can also be used for feature selection. Whereas, L2 will bring the coefficients very close to zero but will not make them zero hence it produced dense result.

3)

L1 is computationally inefficient, whereas L2 is computationally efficient.

4)

L2 regularization has closed form solution and L1 regularization is not differentiable or doesn't have closed form solution.

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 3

As per the principle of Occams's Razor, it's always best to choose simpler model.
And we can see that L1 has more floating point digits hence occupy more bits as compared to L2 model. Thus, Model L2 between the two would be preferable.

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 4

In order to build a more generalisable and robust model we should always keep the regularization term(sum of the sq. of the coefficients or sum of the absolute value of the coefficients) added to the error term(sum of the squares of the difference between actual and the predicted value).
This will keep a check on the model's coefficient values if we have fine tuned our regularization term appropriately, hence ensuring both the consistency and correctness is maintained.

Accuracy is measured as degree of closeness of the predicted value is to the actual value , hence reducing the error rate. If we want more accuracy then the model has to be fairly complex to perform well when supplied with certain specific type of data. However, this model which has very high accuracy will not be generalizable and robust if we tweak the data a little bit and will result in more errors and thus reduce the accuracy. Therefore, we need to strike a balance between the right accuracy that the model can give but still is generalisable or robust enough to make sure that the error rate is not very high in case of slight deviation in the data being fed to the final model that it being tested.

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 5

As we have seen that both the Lasso and Ridge model gives almost the same accuracy. Hence, now it boils down to choosing a model that is more simpler.

As we know that Lasso always gives sparse solution hence it shrinks the values of feature coefficients to zero if they are not so important thus making the model more robust and simpler. Lasso will be better choice for the model.