**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:**

One definite reason for seeing a large difference between training and test accuracy is overfitting or high variance. The model that was built could’ve been very complex for example, with higher order polynomial features.

This can be solved by adding a regularization term to the cost function which penalizes the model accuracy as the complexity of the model increases. The regularization term can include (just like linear regression regularization term) a lambda factor and sum of squares of the coefficients of the features.

**Question-2:**

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

**Answer:**

**Regularization Term**

In L2 it is the sum of the square of the weights of the features, while in L1 it is just the sum of the weights.



**Sparsity**

Refers to that only very few entries in a matrix (or vector) is non-zero. L1-norm has the property of producing many coefficients with zero values for less important features or very small values with few large coefficients. L2-norm on the other hand doesn’t produce coefficients with zero value. Since L1 produces coefficients with zero values, it helps in feature selection automatically.

**Computational efficiency**

L1 (Lasso) does not have an analytical solution, but L2 (Ridge) does due to matrix representation. This allows the L2-norm solutions to be calculated computationally efficiently. However, L1-norm solutions does have the sparsity properties which allows it to be used along with sparse algorithms, which makes the calculation more computationally efficient.

**Solution uniqueness**

L2 has unique solutions whereas L1 doesn’t because minimizing the L2 loss corresponds to calculating the arithmetic mean, which is unambiguous, while minimizing the L1 loss corresponds to calculating the median, which is ambiguous.

Reference => <https://en.wikipedia.org/wiki/Central_tendency#Solutions_to_variational_problems>

**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:**

According to Occam’s Razor, “When in dilemma, choose the simpler model”. In this case both are simple and we can’t say one is more complex than others. Both would require the same data type (float) to store the matrix in the memory/disk.

But, if we are to consider the numbers after decimal also contributes to complexity, I would go with Model L2.

**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:**

Models are trained on a set of training data but their efficiency is determined by the ability to perform well on the unseen (test) data. This is the classic case of bias vs variance trade off.

Extremely complex models don’t generalize well since they are prone to change with small changes in the input data. Extremely simple models are likely to fail in predicting hence are prone to make errors and less accurate.

Hence, we should always select a model which is just complex enough to understand the variance in the data without much inaccuracy at the same time not too complex to overfit.

This can be achieved using regularization. Regularization is the process of deliberately simplifying models to achieve the correct balance between keeping the model simple and yet not too naïve.

**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:**

The optimal value of lambda for Ridge is 1 and for the Lasso is 0.001. These are the values at which the train and test set scores of r2 converge. The same is explained in the Jupyter notebook.

Below is the o/p of the Lasso model with lambda 0.001:



Below is the o/p of the Ridge model with lambda 1:



Even though the variables that are taken input into Ridge came from Lasso, the accuracy of Ridge is slightly better than Lasso’s.

I would apply lambda=0.001 with Lasso regression since Lasso regression helps in feature elimination by making the coefficients of insignificant/duplicate variables 0 thus giving only those variables which are significant in determining the house price.