ADVANCE REGRESSION ASSIGNMENT

ASSIGNMENT: PART 2

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?

Higher value of accuracy in training data implies that the model very well explains the training set. Here in this case we have got 97% of accuracy in training set which implies the model in hand almost explains every data point in the dataset, while we could see test accuracy is only 48%. On other terms we can say, the model is over fitting. The model would be more complex hence explaining almost all the variance of the training set hence not being a good general model that can model outside data, in this case the test dataset.

Maximizing training accuracy, rewards only complex models that won't necessarily generalise.

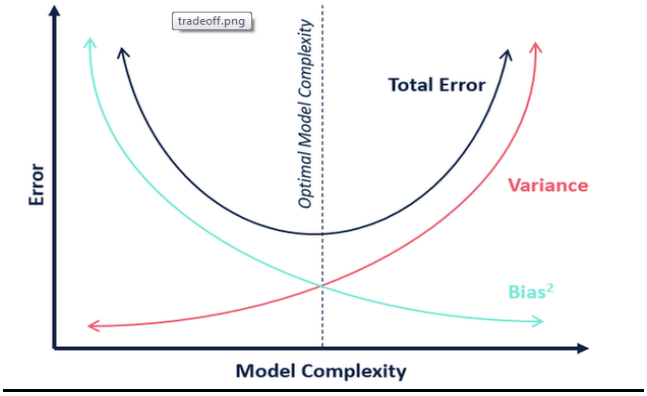
The solution here is to strike the right balance of bias variance so that the model is kept simple at the same time not making the model too naïve for any use. This indicates perfect strike of bias-variance trade-off. This can be achieved by using the approach of regularization. It is a process of deliberately simplifying the model to achieve the correct strike as mentioned above. Hyperparameter are one such parameter that we pass on to control the complexity of the model. Hyperparameter can be tuned with the help of cross validation techniques hence controlling the complexity of the model. Also the use of regularized regression like Ridge and Lasso approach and using model selection criteria as AIC/BIC which penalizes the model for more number of features can be used to help solve this issue of over fitting.

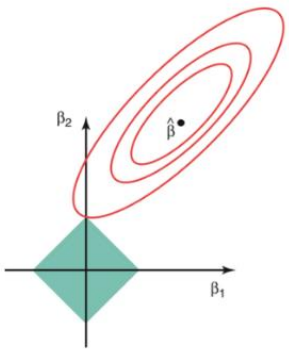
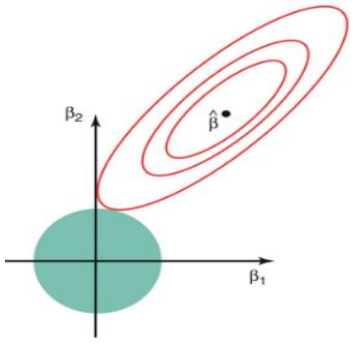
Question 2: List at least four differences in detail between L1 and L2 regularisation in regression.

L1 regularisation or in other words Lasso regularisation stands for Least Absolute Shrinkage and Selection Operator while L2 regularisation stands for Ridge regularisation technique.

Both these regularisation methods are used to make a regression model simpler, balancing the bias-variance trade off.

The significant differences between the both regression techniques are below -



L1 regularisation (LASSO)	L2 regularisation (RIDGE)
Key difference between both is the penalty term – LASSO regression has "absolute value of magnitude" of coefficient as penalty term to the loss function.	RIDGE regression has "squared magnitude" of coefficient as penalty term to the loss function.
Cost Function : $\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \lvert\beta_j\rvert$	Cost Function : $\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$
Effective in feature selection.(Lesser important features coefficients will be shrinked to zero hence the name LASSO)	No feature selection
Computationally intense. LASSO requires iterations to get to the final solution.	Less intense. RIDGE regression almost always has a matrix representation for the solution.
Sparse outputs. L1 has property of making many coefficients zero or low values.	Non sparse outputs
Contours – 	Contours – 

Question 3: Consider two linear models:

L1: y = 39.76x + 32.648628 AndL2: 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?

We are given here two models which does equally well with the test dataset. If we observe the models we can see that both the models seems to simpler models and don't seem to be like complex polynomial model or others. The only visible difference we can note here is the coefficient terms for both the models.

We can see L1 model coefficients to be little complex than the L2 model since more the number of bits in the coefficient term more number of bits needed to represent the term hence increasing the complexity. Observing the decimal terms of both the models we can clearly see that the number of digits after the decimal point are less in L2 model compared to L1 model hence making L1 model little more complex than L2. So, L2 model is preferred among these two.

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?

There is a need to make model robust and generalisable so as to perform equally well in all the given dataset and yet explaining the major variance of the data. In other terms, striking the right balance of bias variance trade off can be achieved with technique called as regularisation. Our model is supposed to be not too complex ending up over fitting the train data and performing poor in other sets and also not too simple ending up underfitting that it generalises everything. Complexity of the model can be controlled with the help of tuning the hyperparamater along with the cross validation techniques. There are a number of regularized regression techniques such as Lasso and Ridge that can be followed as well hence making the model robust and generalisable. The models hence created can be also evaluated with the selection criterions such as Akaike Information criterion or Bayesian information criterion or even Adjusted R2 which penalizes the models for more number of features hence making the model simple.

But the implication of the same sure would be for sure decreasing the accuracy on the model. We know that over fitted models would be explaining high variance in train dataset with a couple number of predictor variables hence making the complex. As part of generalizing, we will end up removing less important variables which might end up reducing the accuracy, but still the final model produced hence would be a good model in terms of performance over a varied

datasets. So For eg: A model with 10 parameters explaining a variance of 80-85% in train-test would be far better model than the one with 20+ parameters explaining a very high variance of 95%+ in train.

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?

We have seen the difference between both the regularisation techniques – Ridge and Lasso.

We know the very idea on regularized regression which is to decrease the model complexity. Ridge can be used when we wanted to decrease the model complexity at the same time wanted to retain all the variables of the model. Hence Ridge instead of forcing the coefficient of the features to be zero, just penalizes them if they are too far from zero, hence making them to be small.

Lasso can be used when with penalizing the model while reducing the model complexity, feature selection also needs to be done then Lasso can be used. Lasso shrinks the insignificant feature coefficients to zero hence doing feature selection.

The idea here is to understand the pricing dynamics of a new market. We wanted to know the significant features which help in predicting the price of the house. Hence even though there are 100s or 200s of features that are predictors for the price of a house, we just need to know the significant predictors. Hence we will choose Lasso over Ridge, since it does feature selection and provides us with important features.