Advanced Regression
– Questions & Answers

Assignment Subjective Questions

Author	**Rahul Kumar**
Date	**04-July-2019**
Course	**PGDDS**
Batch	**Dec 2018**

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?

→ The model built by Rahul seems having problem of overfitting. We could see that the model is having high training accuracy i.e. 97% and low test accuracy of 48%.

The reason could be

- The hypothesis function using may be too complex that model perfectly fits the training data but fails to do on test/validation data.
- The number of learning parameters in model is way too big that instead of generalizing the examples, your model learns those examples and hence the model performs badly on test/validation data.

To solve the above problems a number of solutions can be tried depending on dataset:

- Using a simple cost and loss function.
- Using regulation which helps in reducing over-fitting
- Reducing the number of learning parameters in model

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

➔ Regularization is helpful to discourage the model from becoming too complex even if the model explains the (training) observations better. There are two common methods of regularisation in regression, namely ridge regression and lasso regression
The Lasso regression uses L1 norm for regularization and The Ridge regression uses L2 norm for regularization
Let's find out some basic difference between L1 and L2 regularization in regression:

1. One of difference is that L1 is just the sum of the weights while L2 is the sum of the square of the weights (see the below highlighted box)

L1 regularization on least squares

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2 + \lambda \sum_{i=1}^{k} |w_i|$$

L2 regularization on least squares

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \sum_j \left(t(\mathbf{x}_j) - \sum_i w_i h_i(\mathbf{x}_j) \right)^2 + \lambda \sum_{i=1}^{k} w_i^2$$

2. Another difference is on feature selection i.e. L1 is having built-in feature selection while L2 does not. Let's suppose the model have 100 coefficients but only 10 of them have non-zero coefficients, this is shows that "the other 90 predictors are useless in predicting the target values". L2-norm produces non-sparse coefficients, so does not have this property. Lasso trims down the coefficients of redundant variables to zero and, hence, indirectly performs variable selection. While Ridge, on the other hand, reduces the coefficients to arbitrarily low values, but not zero.

3. L1 has the property of producing many coefficients with zero values or very small values with few large coefficients. Sparsity refers to that only very few entries in a matrix/vector is non-zero. On other hand L2 does not have spare output property.

4. L1-norm does not have an analytical solution, but L2-norm does. 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

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?

➔ Comparing between these linear models, I would consider model L2.
It is simpler comparatively to L1 and hence would be easy to generalize and having more robust feature. But with said this we need to look for Cp, AIC, BIC and adjusted R2 too. Lower the value of Cp, AIC and BIC, better is the fit of the model. Higher the Adjusted R2, better is the fit of the model.

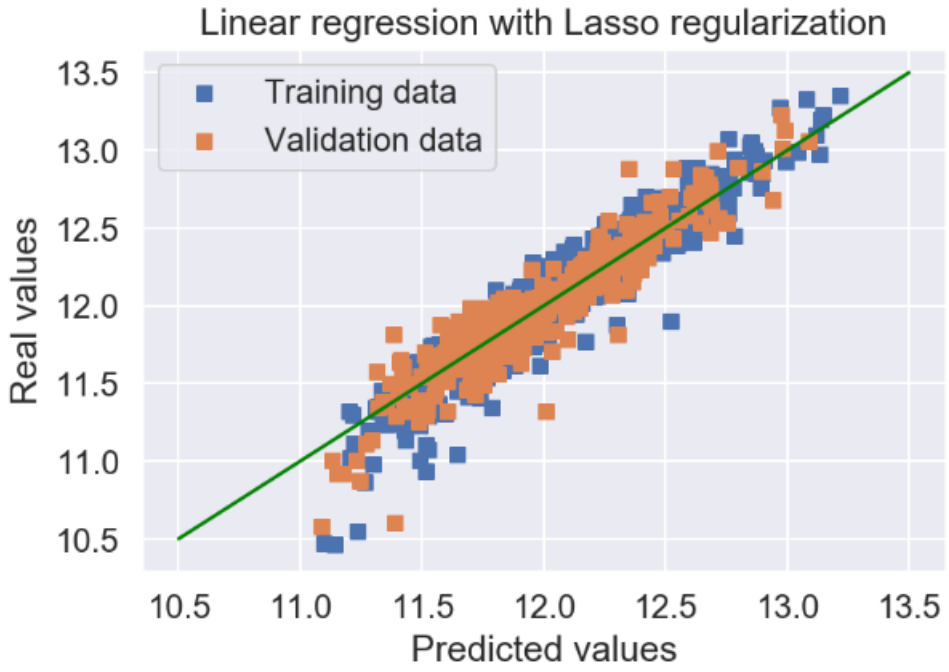
4. How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

➔ The model should be as simple as possible. This might lead to a decrease in training accuracy but this will be make the model more robust and easy to generalize. This can also be understood using the Bias variance tradeoff.
The regression does not account for model complexity - it only tries to minimize the error (e.g. MSE), although if it may result in arbitrarily complex coefficients. On the other hand, in regularized regression, the objective function has two parts - the error term and the

regularization term. With regularization we can achieve bias-variance trade off resulting the accuracy for both train and test dataset mostly remains nearby

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?

➔ I would be selecting Lasso Model. As we saw some parameters in analysis.
Alpha value: When alpha is too large the regularization is too strong and the model cannot capture all the complexities in the data. If however we let the model be too flexible (alpha small) the model begins to overfit. A value of alpha = 0.01 is right based on the Lasso plot.



- Complexity and Error: Lasso model seems to have same training error with only 48 features selected and for test data RMSE also lower compare to the Ridge model which make the Lasso model more simple and generic(Occam's Razor)

---------- End of File ----------