Advance Regression Interview Questions

We believe that you have learned both theoritical and practical knowledge on Naive Bayes classification algorithm through your assignment.

So let's test your knowledge here. This will help you to be prepared for interviews too!

Best with Quest

1. What's regularization, and what's the difference between L1 and L2 regularization?

Regularization in machine learning is the process of regularizing the parameters that constrain, regularizes, or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning of a more complex or flexible model, avoiding the risk of overfitting. Regularization basically adds the penalty as model complexity increases, which can help avoid overfitting.

L1 effectively removes features that are unimportant, and doing this too aggressively can lead to underfitting. L2 weighs each feature instead of removing them entirely, which can lead to better accuracy. Briefly, L1 removes features while L2 doesn't, and L2 regulates their weights instead.

2. How to identify a high variance model? How do you fix it?

A High Variance model is due to a complex model and can be easily identified when you see:

Low training error
High validation error or high test error
To fix a high variance model, you can:

Get more training data
Reduce input features
Increase the regularization term

3. How to choose the value of the parameter learning rate (α)?

Selecting the value of learning rate is a tricky business. If the value is too small, the gradient descent algorithm takes ages to converge to the optimal solution. On the other hand, if the value of the learning rate is high, the gradient descent will overshoot the optimal solution and most likely never converge to the optimal solution.
To overcome this problem, you can try different values of alpha over a range of values and plot the cost vs the number of iterations. Then, based on the graphs, the value corresponding to the graph showing the rapid decrease can be chosen.

4. How to choose the value of the regularisation parameter (λ)?

Selecting the regularisation parameter is a tricky business. If the value of λ is too high, it will lead to extremely small values of the regression coefficient β, which will lead to the model underfitting (high bias – low variance). On the other hand, if the value of λ is 0 (very small), the model will tend to overfit the training data (low bias – high variance).
There is no proper way to select the value of λ. What you can do is have a sub-sample of data and run the algorithm multiple times on different sets. Here, the person has to decide how much variance can be tolerated. Once the user is satisfied with the variance, that value of λ can be chosen for the full dataset.
One thing to be noted is that the value of λ selected here was optimal for that subset, not for the entire training data.

5. Why do we need Polynomial Regression?

- If we apply a linear model on a linear dataset, then it provides us a good result as we have seen in Simple Linear Regression, but if we apply the same model without any modification on a non-linear dataset, then it will produce a drastic output. Due to which loss function will increase, the error rate will be high, and accuracy will be decreased.

-So for such cases, where data points are arranged in a non-linear fashion, we need the Polynomial Regression model. We can understand it in a better way using the below comparison diagram of the linear dataset and non-linear dataset.
ML Polynomial Regression

-In the above image, we have taken a dataset which is arranged non-linearly. So if we try to cover it with a linear model, then we can clearly see that it hardly covers any data point. On the other hand, a curve is suitable to cover most of the data points, which is of the Polynomial model.

-Hence, if the datasets are arranged in a non-linear fashion, then we should use the Polynomial Regression model instead of Simple Linear Regression.

6. Which is more important to you: model accuracy or model performance?

Such machine learning interview questions tests your grasp of the nuances of machine learning model performance! Machine learning interview questions often look towards the details. There are models with higher accuracy that can perform worse in predictive power—how does that make sense?

Well, it has everything to do with how model accuracy is only a subset of model performance, and at that, a sometimes misleading one. For example, if you wanted to detect fraud in a massive dataset with a sample of millions, a more accurate model would most likely predict no fraud at all if only a vast minority of cases were fraud. However, this would be useless for a predictive model—a model designed to find fraud that asserted there was no fraud at all! Questions like this help you demonstrate that you understand model accuracy isn't the be-all and end-all of model performance.

7. How do you handle missing or corrupted data in a dataset?

You could find missing/corrupted data in a dataset and either drop those rows or columns, or decide to replace them with another value.

In Pandas, there are two very useful methods: isnull() and dropna() that will help you find columns of data with missing or corrupted data and drop those values. If you want to fill the invalid values with a placeholder value (for example, 0), you could use the fillna() method.

8. Difference between ridge regression and lasso and its effect?

Yes…Ridge and Lasso regression uses two different penalty functions. Ridge uses l2 where as lasso go with l1. In ridge regression, the penalty is the sum of the squares of the coefficients and for the Lasso, it's the sum of the absolute values of the coefficients. It's a shrinkage towards zero using an absolute value (l1 penalty) rather than a sum of squares(l2 penalty).

As we know that ridge regression can't zero coefficients. Here, you either select all the coefficients or none of them whereas LASSO does both parameter shrinkage and variable selection automatically because it zero out the co-efficients of collinear variables. Here it helps to select the variable(s) out of given n variables while performing lasso regression.

Another type of regularization method is ElasticNet, it is hybrid of lasso and ridge regression both. It is trained with L1 and L2 prior as regularizer. A practical advantage of trading-off between Lasso and Ridge is that, it allows Elastic-Net to inherit some of Ridge's stability under rotation.

9. What are some limitations of the Lasso and Ridge model. Why ElasticNet is used?

Lasso does not work well with multicollinearity. If you are unfamiliar, multicollinearity occurs when some of the dependent variables are correlated with each other. Why? Lasso might randomly choose one of the multicollinear variables without understanding the context. Such an action might eliminate relevant independent variables.

Ridge does not eliminate coefficients in your model even if the variables are irrelevant. This can be negative if you have more features than observations.

Elastic Net combines feature elimination from Lasso and feature coefficient reduction from the Ridge model to improve your model's predictions.

10. Why would a line with high variance be a bad thing?

After all, wouldn't a line that captures every point be the ideal line? Sadly no. This line captures the peculiar nuances of the sample data well but it will not necessarily perform as well on unseen data. These nuances of the sample data are the outliers and other unique characteristics of the sample data, characteristics that might not be very true on unseen data.

Now Rest with this Quest :)