

MACHINE LEARNING

- 1) Which of the following method we use to find the best fit line for data in Liner Regression?
= Option (A) Least Square Error
- 2) Which of the following statement is true about outliers in Liner Regression?
= Option (A) Liner Regression is sensitive to outliers.
- 3) A line falls from left to right if a slope is.....?
= Option (A) Positive
- 4) Which of the following have symmetric relation between dependent variable and independent?
= Option (A) Regression
- 5) Which of the following is the reason for over fitting condition?
= Option (A) High bias and high variance
- 6) If output involves label, then that model is called as:
= Option (B) Predictive modal
- 7) Lasso and Ridge regression techniques belong to?
= Option (B) Removing Outliers
- 8) To overcome with the imbalance dataset which techniques can be used?
= Option (D) SMOTE
- 9) The AUC receiver operator characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses....to make graphs?
= Option (A) TPR and FPR
- 10) In AUC receiver operator characteristics (AUCROC) curve for the better model area under the curve should be less.
= Option (B) False
- 11) Pick the feature extraction from below:
= Option (B) Apply PCA to project high dimensional data
- 12) Which of the following is true about Normal Equation used to compute the coefficient of the Liner Regression?
= Option (B) & (C) It becomes slow when number of features is very less.
AND We need to iterate. (Respectively)

13) Explain the term Regularization?

Regularization in Machine Learning is a process which prevents overfitting by not adding extra information into it. In other words, regularization is a technique to discourage the complexity of the model.

Example: We want to predict blood sugar using factors such as age, alcohol consumption and history of smoking. Generally, higher age, high/routine alcohol consumption and smoking are correlated with higher BP. So, there is a pattern that a model can pick up to work upon.

Sometimes, it happens that there can be some individuals with higher age, routine/high consumption of alcohol but then also they have below average BP (maybe they are able to maintain high physical workout sessions daily). This deviation in pattern is not measures by predictors. These are known as noise.

This noise somehow leads to overfitting (overfitting is basically modelling error which occurs when a function is too closely fit into a limited set of data points).

Therefore, regularization describes a technique to prevent overfitting. Various models are prone to picking up random noise from particular data set which might obscure the patterns found in the data. Regularizations helps in reducing the influence of noise on the model's performance.

14) Which particular algorithms are used for Regularizations?

There are basically three major algorithms used for regularization:

- L1(LASSO) Least Absolute Shrinkage and Selection Operator

= LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model. LASSO regression adds a penalty (*L1 penalty*) to the loss function that is equivalent to the magnitude of the coefficients.

In LASSO regression, the penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the regularization

parameter λ is sufficiently large. To sum up, LASSO regression converts coefficients of less important features to zero, which indeed helps in feature selection, and it shrinks the coefficients of remaining features to reduce the model complexity, hence avoiding overfitting.

- L2 (Ridge Regression)

=Lasso Regression adds “*absolute value of magnitude*” of coefficient as penalty term to the loss function(L). Ridge regression adds “*squared magnitude*” of coefficient as penalty term to the loss function(L). Ridge regression is a method for analysing data that suffer from multi-collinearity. Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.

- Dropout

=Dropout is a technique where randomly selected neurons are ignored during training. They are “dropped-out” randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass.

15) Explain the term error present in liner regression equation?

The error in linear regression is the rejection of the null hypothesis when the null hypothesis is true. If you reject the null hypothesis, then the linear regression equation has a big error.