MACHINE LEARNING

1. Least Square Error
2. Linear regression is sensitive to outliers
3. Negative
4. Correlation
5. Low bias and high variance
6. Predictive model
7. Regularization
8. Cross validation
9. Sensitivity and Specificity
10. False
11. Apply PCA to project high dimensional data
12. A) We don't have to choose the learning rate.
    B) It becomes slow when number of features is very large.

---

13. Explain the term regularization?

    Regularization is a technique which is used to solve the overfitting problem of the machine learning models.

    Overfitting- Overfitting is a phenomenon which occurs when a model learns the detail and noise in the training data to an extent that it negatively impacts the performance of the model on new data. So the overfitting is a major problem as it negatively impacts the performance. Regularization technique to the rescue. Generally, a good model does not give more weight to a particular feature. The weights are evenly distributed. This can be achieved by doing regularization.

    There are two types of regularization as follows:

- L1 Regularization or Lasso Regularization
- L2 Regularization or Ridge Regularization

➢ **L1 Regularization or Lasso Regularization**

L1 Regularization or Lasso Regularization adds a penalty to the error function. The penalty is the sum of the absolute values of weights.

➢ **L2 Regularization or Ridge Regularization**

➢ L2 Regularization or Ridge Regularization also adds a penalty to the error function. But the penalty here is the sum of the squared values of weights.

---

14. Which particular algorithms are used for regularization?

    The following algorithms are used for regularization:

Working of Ridge, LASSO, and Elastic-Net Regression

The working of all these algorithms is quite similar to that of Linear Regression, it's just the loss function that keeps on changing!

$$Loss = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2$$

Loss Function for Linear Regression

Ridge Regression

Ridge regression is a method for analyzing data that suffer from multi-collinearity.

$$Loss = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^{n} w_i^2$$

Loss Function for Ridge Regression

Ridge regression adds a penalty ***(L2 penalty)*** to the loss function that is equivalent to the square of the magnitude of the coefficients.

The regularization parameter *(λ)* regularizes the coefficients such that if the coefficients take large values, the loss function is penalized.

- *λ → 0*, the penalty term has no effect, and the estimates produced by ridge regression will be equal to least-squares i.e. the loss function resembles the loss function of the Linear Regression algorithm. Hence, a lower value of λ will resemble a model close to the Linear regression model.
- ***λ → ∞,*** the impact of the shrinkage penalty grows, and the ridge regression coefficient estimates will **approach zero** (coefficients are close to zero, but not zero).

*Note: Ridge regression is also known as the **L2 Regularization**.*

To sum up, **Ridge regression shrinks the coefficients as it helps to reduce the model complexity and multi-collinearity.**

LASSO Regression

LASSO is a regression analysis method that performs both feature selection and regularization in order to enhance the prediction accuracy of the model.

$$Loss = \sum_{i=1}^{n} (y_i - (w_i x_i + c))^2 + \lambda \sum_{i=1}^{n} |w_i|$$

Loss Function for LASSO Regression

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.

*Note: LASSO regression is also known as the **L1 Regularization (L1 penalty).***

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.**

Elastic-Net Regression

Elastic-Net is a regularized regression method that linearly combines the L1 and L2 penalties of the LASSO and Ridge methods respectively.

$$Loss = \sum_{i=0}^{n} (y_i - (w_i x_i + c))^2 + \lambda_1 \sum_{i=0}^{n} |w_i| + \lambda_2 \sum_{i=0}^{n} w_i^2$$

Loss Function for Elastic-Net Regression

What does Regularization achieve?

A standard least-squares model tends to have some variance in it i.e. the model won't generalize well for a data set different than its training data. ***Regularization, significantly reduces the variance of the model, without a substantial increase in its bias***.

So the regularization parameter λ, used in the techniques described above, controls the impact on bias and variance. As the value of λ rises, it reduces the value of coefficients and thus reducing the variance. ***This increase in λ is beneficial as it is only reducing the variance (hence avoiding overfitting), without losing any important properties in the data.*** But after a certain value, the model starts losing important properties, giving rise to bias in the model and thus underfits the data. Therefore, the value of λ should be carefully selected.

It is a useful technique that can help in improving the accuracy of your regression models.

---

15. Explain the term error present in linear regression equation?

An error term is a residual variable produced by a statistical or mathematical model, which is created when the model does not fully represent the actual relationship between the independent variables and the dependent variables. As a result of this incomplete relationship, the error term is the amount at which the equation may differ during empirical analysis. The error term is also known as the residual, disturbance, or remainder term, and is variously represented in models by the letters *e,* ε, or u.

- An error term appears in a statistical model, like a regression model, to indicate the uncertainty in the model.
- The error term is a residual variable that accounts for a lack of perfect goodness of fit.
- Heteroskedastic refers to a condition in which the variance of the residual term, or error term, in a regression model varies widely.

An error term represents the margin of error within a statistical model; it refers to the sum of the deviations within the regression line, which provides an explanation for the difference between the theoretical value of the model and the actual observed results. The regression line is used as a point of analysis when attempting to determine the correlation between one independent variable and one dependent variable.

An error term essentially means that the model is not completely accurate and results in differing results during real-world applications. For example, assume there is a multiple linear regression function that takes the following form:

$Y=\alpha X+\beta\rho+\epsilon$

where: α,β=Constant parameters X,ρ=Independent variablesϵ=Error term

When the actual Y differs from the expected or predicted Y in the model during an empirical test, then the error term does not equal 0, which means there are other factors that influence Y**.**
