

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

Both R-squared and residual sum of squares (RSS) are measure of goodness of fit in regression analysis but they capture different aspects of the model's performance

2. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.

What Is SST in Statistics?

The sum of squares total (SST) or the total sum of squares (TSS) is the sum of squared differences between the observed dependent variables and the overall mean. Think of it as the dispersion of the observed variables around the mean—similar to the variance in descriptive statistics. But SST measures the total variability of a dataset, commonly used in regression analysis and ANOVA.

What is SSR in statistics?

The **sum of squares due to regression (SSR)** or **explained sum of squares (ESS)** is the sum of the differences between the *predicted value* and the **mean** of the *dependent variable*. In other words, it describes how well our line fits the data.

What Is SSE in Statistics?

The **sum of squares error (SSE)** or **residual sum of squares (RSS, where** residual means remaining or unexplained**)** is the difference between the *observed* and *predicted* values.

3. What is the need of regularization in machine learning?

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

4. What is Gini–impurity index?

A measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree

5. Are unregularized decision-trees prone to overfitting? If yes, why?

in the case of trees, adding a node to a leave based on one feature should be done only if the feature really brings information at this level. The feature could be random though and this would deteriorate greatly the fit

6. What is an ensemble technique in machine learning?

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. To better understand this definition lets take a step back into ultimate goal of machine learning and model building.

7. What is the different between bagging and boosting techniques?

S.NO	Bagging	Boosting
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Each model is built independently.	New models are influenced by the performance of previously built models.
5.	Different training data subsets are selected using row sampling with replacement and random sampling methods from the entire training dataset.	Every new subset contains the elements that were misclassified by previous models.
6.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.
7.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
8.	In this base classifiers are trained parallelly.	In this base classifiers are trained sequentially.
9	Example: The Random forest model uses Bagging.	Example: The AdaBoost uses Boosting techniques

8. What is out-of-bag error in random forests?

OOB (out-of-bag) errors are an estimate of the performance of a random forest classifier or regressor on unseen data. In scikit-learn, the OOB error can be obtained using the `oob_score_` attribute of the random forest classifier or regressor.

9. What is K-fold cross-validation?

To evaluate the performance of some model on a dataset, we need to measure how well the predictions made by the model match the observed data.

The most common way to measure this is by using the mean squared error (MSE), which is calculated as

$$\text{MSE} = (1/n) * \sum (y_i - f(x_i))^2$$

10. What is hyper parameter tuning in machine learning and why it is done?

parameters are estimated from the dataset. They are part of the model equation. The equation below is a logistic regression model. Theta is the vector containing the parameters of the model.

Hyperparameters are set manually to help in the estimation of the model parameters. They are not part of the final model equation.

11. What issues can occur if we have a large learning rate in Gradient Descent?

learning rate is one of The most important thing to consider in whole of machine learning. But choosing the correct learning rate is pretty much impossible all the time.

But having the knowledge of what it can do is useful in some cases where we can at least guess what could be the appropriate learning rate We will see how even the slightest change in learning_rate can improve the speed drastically or could break your model. I mean it can overshoot the minima and it might never return .

12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary

13. Differentiate between Adaboost and Gradient Boosting ?

AdaBoost

AdaBoost or Adaptive Boosting is the first Boosting ensemble model. The method automatically adjusts its parameters to the data based on the actual performance in the current iteration. Meaning, both the weights for re-weighting the data and the weights for the final aggregation are re-computed iteratively.

In practice, this boosting technique is used with simple classification trees or stumps as base-learners, which resulted in improved performance compared to the classification by one tree or other single base-learner.

Gradient Boosting

Gradient Boost is a robust machine learning algorithm made up of Gradient descent and Boosting. The word 'gradient' implies that you can have two or more derivatives of the same function. Gradient Boosting has three main components: additive model, loss function and a weak learner.

The technique yields a direct interpretation of boosting methods from the perspective of numerical optimisation in a function space and generalises them by allowing optimisation of an arbitrary loss function.

14. What is bias-variance trade off in machine learning?

The **bias-variance tradeoff** refers to the tradeoff that takes place when we choose to lower bias which typically increases variance, or lower variance which typically increases bias.

15. Give short description each of Linear, RBF, Polynomial kernels used in SVM

Linear Kernel

A linear kernel is a type of kernel function used in machine learning, including in SVMs (Support Vector Machines). It is the simplest and most commonly used kernel function, and it defines the dot product between the input vectors in the original feature space

The linear kernel can be defined as $K(x, y) = x \cdot y$