# MACHINE LEARNING

Q1. 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?

**Ans:** Both R-squared and Residual Sum of Squares (RSS) are measures of goodness of fit in regression model, but they capture different aspect of model performance. R-squared is often used when you want to understand the proportion of variance explained by the model especially in the context of comparing different model. RSS is useful when you want to evaluate the absolute goodness of fit, focusing on the magnitude of the prediction errors. In general, a good model should have both a high R-Squared value and a low RSS value.

Q2. 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.

**Ans:** TSS= ∑(Yi−¯Y)2, where Yi is the actual value of the response variable for observation i, and ¯Y is the mean of the response variable.

ESS= ∑(^Yi−¯Y)2, where ^Yi is the predicted value of the response variable for observation i.

RSS= is the sum of squared difference between the actual and predicted values of the response variable.

Relating equation= Var (Y)=∑(Yi−¯Y)2.

Now, Yi=ˆY− ei. ∑ni =1(Yi−¯Y)2= ∑ni =1(^Yi−¯Y)2+∑ni =1e2i.

Q3. What is the need of regularization in machine learning?

**Ans:** While training a machine learning model, the model can easily be overfitted or under fitted. To avoid this, we use regularization in machine learning to properly fit a model into our test set. Regularization techniques help reduce the chance of overfitting and help us get an optimal model.

Q4. What is Gini–impurity index?

**Ans:** Gini Impurity tell us what is the probability of misclassifying of an observation. Lower the Gini, better the split. In other word, lower the likelihood of misclassification.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

**Ans:** Yes, unregularized Decision trees are prone to overfitting because they can keep growing deeper and more complex until they perfectly classify the training data. This can lead to the tree capturing noise in the data, rather than the underlying relationships, and thus performing poorly on new, unseen data. Regularization techniques such as pruning, setting a minimum number of

samples required to split a node, or limiting the maximum depth of the tree can help mitigate overfitting in decision trees.

Q6. What is an ensemble technique in machine learning?

**Ans:** Ensemble learning is a machine learning technique that enhances accuracy and resilience in forecasting by merging predictions from multiple models. It aims to mitigate errors or biases that may exist in individual models by leveraging the collective intelligence of the ensemble.

Q7. What is the difference between Bagging and Boosting techniques?

**Ans:** Bagging involves training multiple models independently on random subsets of data and then combining their predictions through a majority vote. Boosting focuses on correcting the errors made by previous weak models in a sequence to create a stronger model.

Q8. What is out-of-bag error in random forests?

**Ans:** Out-of-bag error, is a method of measuring the prediction errors of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging).

Q9. What is K-fold cross-validation?

**Ans:** K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds. The model is trained and evaluated k times, using a different fold as the validation set each time.

Q10. What is hyper parameter tuning in machine learning and why it is done?

**Ans:** When you are training machine learning models, each dataset and model needs a different set of hyper parameters, which are a kind of variable. The only way to determine these is through multiple experiments, where you pick a set of hyper parameters and run them through your model. This is called *hyper parameter tuning*. It is done to directly control model structure, function, and performance. Hyper parameter tuning allows data scientists to tweak model performance for optimal results. This process is an essential part of machine learning, and choosing appropriate hyper parameter values is crucial for success.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

**Ans:** When the learning rate is too large, gradient descent can suffer from divergence. This means that weights increase exponentially, resulting in exploding gradients which can cause problems such as instabilities and overly high loss values.

Q12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

**Ans:** We cannot use Logistic Regression for classification of non Linear Data because Logistic Regression is considered a generalized linear model. It will perform poorly on the dataset if the target label has no linear correlation with the feature.

Q13. Differentiate between Adaboost and Gradient Boosting.

**Ans:**

| | Adaboost | Gradient Boosting |
|---|---|---|
| 1 | An additive model where shortcomings of previous models are identified by high-weight data points. | An additive model where shortcomings of previous models are identified by the gradient. |
| 2 | The trees are usually grown as decision stumps | The trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes. |
| 3 | Each classifier has different weights assigned to the final prediction based on its performance. | All classifiers are weighed equally and their predictive capacity is restricted with learning rate to increase accuracy. |
| 4 | It gives weights to both classifiers and observations thus capturing maximum variance within data. | It builds trees on previous classifier's residuals thus capturing variance in data. |

Q14. What is bias-variance trade off in machine learning?

**Ans:** In machine learning, as you try to minimize one component of the error (e.g., bias), the other component (e.g., variance) tends to increase, and vice versa. Finding the right balance of bias and variance is key to creating an effective and accurate model. This is called the bias-variance tradeoff.

Q15. Give short description each of Linear, RBF, Polynomial kernels used in SVM

**Ans:** Linear- A linear essentially fits a straight line to the data, allowing it to make prediction based on a linear relationship between the input features and the output variable.

RBF- Are the real valued functions that use supervised machine learning to perform as a nonlinear classifier.

Polynomial kernels used in SVM- In machine learning, the polynomial kernel is a kernel function commonly used with support vector machines (SVMs) and other kernelized models, that represents the similarity of vectors in a feature space over polynomials of the original variables, allowing learning of non-linear models.