MACHINE LEARNING

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?

Ans:

R-squared is a better measure of goodness of fit model in regression. R-squared gives the degree of variability in the target variable that is explained by the independent variables. It is a scale invariant statistics. Higher the R-Squared value indicated smaller differences between the observed data and the fitted values.

On the contrary Residual sum of squares or RSS determine the sum of the squared difference between the actual and predicted value, the value depends on the scale of the target variable. It is scale variant statistic.

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.

Ans:

Total Sum of Squares is the squared differences between the observed dependent variable and its mean. It is a measure of the total variability of the dataset.

Explained Sum of Squares tells how much of the variation in the dependent variable the model explained.

Residual Sum of Squares is the sum of the squared differences between the actual and predicted value.

Metrics related to each other:

TSS= RSS + ESS

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}e_i^2$$

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

Ans:

Regularization refers to techniques used to calibrate machine learning models to minimize the adjusted loss function and avoid overfitting or underfitting. Using Regularization, we can fit our machine learning model on a given test set and hence reduce the errors in it.

4. What is Gini–impurity index?

Ans:

Gini-impurity is a measurement used to build Decision Trees to determine how the features of the dataset should split nodes to form the tree. It has a bias toward selecting splits that result in a more balanced distribution of classes. It is a linear measure. The Range of the Gini Index is [0,1], where o indicates perfect purity and 1 indicates maximum impurity. Gini Index is typically used in CART (Classification and Regression Trees) algorithms.

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

Ans:

Yes, unregularized decision-trees are prone to overfitting. Unregularized Decision trees tend to overfit and do not generalize well to new data. Using Regularization we can fit our model on a given test set and hence reduce the errors in it.

6. What is an ensemble technique in machine learning?

Ans:

Simple Ensemble methods:

- Mode: It is the number or value that most often appears in a dataset of numbers or values. In this ensemble technique, machine learning professionals use a number of models for making predictions about each data point. The predictions made by different models are taken as separate votes. Subsequently, the prediction made by most models is treated as the ultimate prediction.

- Mean/Average: In the mean/average ensemble technique, data analysts take the average predictions made by all models into account when making the ultimate prediction.

Advanced Ensemble Methods

- Bagging: This ensemble method is to minimize variance errors in decision trees.
- Boosting: This ensemble method adjusts an observation's weight based on its last classification.

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

Ans:

Bagging	**Boosting**
Simplest way to combining predictions that belong to the same type	Way of combining predictions that belong to the different types
Aim to decrease variance and not bias	Aim to decrease bias and not variance
Each model receives equal weight	Models are weighted according to their performance
Each model is built independently	New models are influenced by the performance of previously built models
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
Bagging tries to solve over-fitting problem	Boosting tries to reduce bias.
If the classifier is unstable (high variance),then apply bagging	If the classifier is stable and simple (high bias) then apply boosting
Base classifiers are trained parallel.	Base classifiers are trained subsequently
Ex: Random forest model uses bagging	Ex: AdaBoost uses boosting techniques

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

Ans:

The RandomForestClassifier is trained using bootstrap *aggregation*, where each new tree is fit from a bootstrap sample of the training observations z=(x,y). The out-of-bag error is the average error for each z calculated using predictions from the tress that do not contain z in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained.

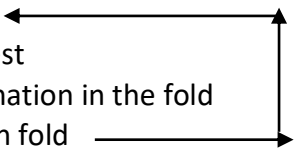
9. What is K-fold cross-validation?

Ans:

Cross-Validation is a technique for evaluating a machine learning model and testing its performance.K-fold Cross validation is one of the method of cross validation that minimizes the disadvantage of the hold-out method.K-fold introduces a new way of splitting the dataset which helps to overcome the "test only once bottleneck".

Life cycle of K-Fold Cross-Validation:

- Scuffle dataset
- Split the dataset into training and test
- Split training dataset into k-folds
- Use (k-1) fold for training
- Always leave 1 fold for test
- Take care of all transformation in the fold
- Find the accuracy on each fold



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

Ans:

Hyper parameter tuning is used for controlling the behavior of the machine learning model.

Hyper parameter is specific to the algorithm. We can find the model parameters using hyper parameters.

It consists of finding a set of optimal hyper parameter values for a learning algorithm while applying the optimized algorithm to any dataset. The combination of hyper parameters maximize the model performance, minimizing a predefined loss function to produce better results with fewer errors.

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

Ans:

If we have large learning rate in Gradient Descent then the model converge quickly to a sub optimal solution. Learning rate in Gradient Descent determines how fast or slow the model will move towards the optimal weights.

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

Ans:

No we cannot use Logistic Regression for classification of Non-Linear Data. Because Logistic Regression has a linear decision surface.Linearly separable data is rarely found in real-world. It is used to predict the discrete function.

13. Differentiate between Adaboost and Gradient Boosting.

Ans:

S.no.	Adaboost	Gradient Boost
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	The trees are grown to a

	grown as decision stumps.	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.

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

Bias is the difference between predicted value and expected value.
High bias: little attention to training data,high error on training and test data

Variance is when the model takes into account the fluctuations in the data.
High variance: lot of attention to training data and does not generalize on the data,high error on test data.

Bias–Variance Tradeoff helps optimize the error in the model and keep it as low as possible.
Both the variance and bias should be low as to prevent overfitting and underfitting.

- Low bias and low variance: underfitting
- High bias and high variance: overfitting

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

- Linear kernel: It is one-dimensional function.
- Polynomial Kernel: It represents the similarity of vectors in the training data in a feature space over polynomials of the original variables used in the kernel. It is less efficient and accurate.
- RBF kernel: It is mostly preferred and used kernel function in SVM. It is used for non-linear data. It helps to make proper separation when there is no prior knowledge of data. It improve the transformation by adding radial basis method.