Batch ID 2404\1023

Q1 to Q15 are subjective answer type questions, Answer them briefly.

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 : a) R-squared, as it represents the proportion of variance explained by the model.
b) RSS, as it measures the overall model fit by summing up the squared residuals.
c) Both R-squared and RSS are equally valid measures.
d) It depends on the specific characteristics of the dataset.

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: The begin by defining the three terms: TSS = $\sum(Yi - \bar{Y})^2$, where Yi is the actual value of the response variable for observation i, and $\bar{Y}$ is the mean of the response variable. ESS = $\sum(\hat{Y}i - \bar{Y})^2$, where $\hat{Y}i$ is the predicted value of the response variable for observation i.

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

ANS

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

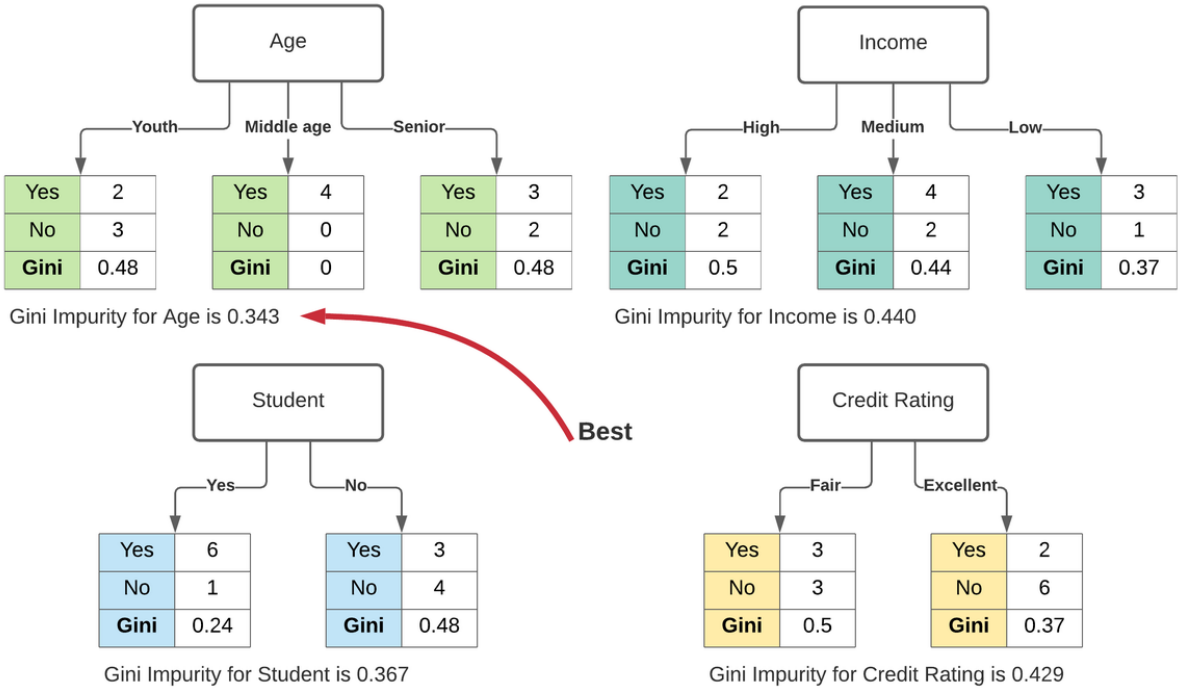
:

4. What is Gini–impurity index?

ANS:

Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree. More precisely, the Gini Impurity of a dataset is a

number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.



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

ANS: Some cons of using decision trees: prone to overfitting. requires a way to turn numeric data into a single decision rule.

6. What is an ensemble technique in machine learning?

ANS:

Ensemble techniques in machine learning combine multiple models to improve the accuracy of results. The three most popular ensemble learning techniques are bagging, boosting, and stacking:

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

ANS:

Bagging and boosting are both ensemble learning techniques in machine learning that combine multiple models to improve the performance of a predictive model. The main difference between the two is how they train the models:

- Bagging

  Trains models independently in parallel on different random subsets of data. Bagging reduces variance by averaging predictions from the models. It's best for models with high variance and low bias. Bagging can be faster than boosting because the models are trained independently.

- Boosting
  Trains models sequentially, with each model learning from the errors of the previous one. Boosting reduces bias by focusing on the errors made by the previous model. It's effective when the model must be adaptive to errors, and is suitable for bias and variance errors. Boosting can be computationally expensive because the models are trained sequentially. However, optimized boosting versions, such as XGBoost and LightGBM, are designed to be faster and more memory-efficient.

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

ANS:

Out-of-bag (OOB) error is a performance metric that estimates how well a random forest model performs on unseen data. It's calculated by averaging the prediction errors of individual trees that don't contain a particular training sample in their bootstrap sample.

9. What is K-fold cross-validation?

ANS:

K-fold cross-validation is a machine learning technique that evaluates the performance of a model by dividing a dataset into k subsets, or folds, of roughly equal size. The model is then trained and evaluated k times, each time using a different fold as the validation set

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

ANS:

Hyperparameter tuning is the process of finding the best values for a machine learning model's hyperparameters to maximize its performance. Hyperparameters are model parameters that

control the learning process, like the learning rate, number of neurons, or kernel size. They are specific to the algorithm and can't be calculated from the data, so they are configured before the model learning process begins

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

ANS:

A large learning rate in gradient descent can cause the following issues:

- Overshooting: The algorithm can "jump over" the minima it's trying to reach, leading to oscillations or divergence
- Loss value doesn't converge: The loss function may not decrease on every iteration, or it may not converge at all
- Learning process fails: The algorithm may fail to learn
- Overfitting: The model may overfit the training data, leading to poor performance on new data
- Converges too quickly: The model may converge too quickly to a suboptimal solution

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

ANS:

No, logistic regression (LR) is not suitable for classifying non-linear data because it's a linear model that assumes a linear relationship between input features and output. This means that LR can't capture the complexity and non-linearity of the data. LR is also only able to predict discrete functions, and has a linear decision surface.

13. Differentiate between Adaboost and Gradient Boosting.

ANS:

AdaBoost and Gradient Boosting are both boosting algorithms that combine multiple weak learners to create strong learners. However, they differ in how they identify model shortcomings and build trees:

- AdaBoost

Uses a single-level decision tree to identify model shortcomings using high-weight data points. AdaBoost builds one-depth regression trees, or decision stumps, and requires multiple iterations to build a sequence of models.

- Gradient Boosting
  Builds trees with higher depth, and identifies model shortcomings using the gradient. Gradient Boosting combines weak learners to form a strong learner, and uses intuition to determine the next model to combine with prior models to minimize cumulative errors

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

Ans:

Bias-variance tradeoff is a theoretical concept in machine learning that describes the balance between a model's bias and variance. Bias is the model's simplifying assumptions to make the target function easier to approximate, while variance is the degree to which the target function estimate varies when using different training data

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

ANS:

In Support Vector Machines (SVMs), kernel functions map datasets into higher-dimensional spaces to make them linear. The three most common kernel functions in SVMs are linear, polynomial, and radial basis function (RBF):

- Linear

  The simplest and most computationally efficient kernel function, it's best for low-dimensional datasets with many features. It's also often used for text classification.

- Polynomial

  Used when data has polynomial features or interactions between features. Polynomial kernels map input data to higher-dimensional spaces using polynomial functions of the original features. They work well for problems where all training data is normalized.

- RBF
  Used when data can't be separated well by a linear or polynomial decision boundary. RBF kernels are a type of Gaussian kernel that project high-dimensional data and search for a linear separation. They're often used when boundaries are expected to be curve-shaped. RBF kernels can lead to higher accuracies and robustness than traditional linear models