1. Using a goodness of fit,we can assess whether a set of obtained frequencies differ from a set of frequencies.
a) Mean
b) Actual
c) Predicted
d) Expected

Ans: d) Expected

2. Chisquare is used to analyse
a) Score
b) Rank
c) Frequencies
d) All of these

Ans: c) Frequencies

3. What is the mean of a Chi Square distribution with 6 degrees of freedom?
a) 4
b) 12
c) 6
d) 8

Ans: c) 6

4. Which of these distributions is used for a goodness of fit testing?
a) Normal distribution
b) Chisqared distribution
c) Gamma distribution
d) Poission distribution

Ans: b) Chisqared distribution

5. Which of the following distributions is Continuous
a) Binomial Distribution
b) Hypergeometric Distribution
c) F Distribution
d) Poisson Distribution

Ans: c) F Distribution

6. A statement made about a population for testing purpose is called?
a) Statistic
b) Hypothesis
c) Level of Significance
d) TestStatistic

Ans: b) Hypothesis

7. If the assumed hypothesis is tested for rejection considering it to be true is called?
a) Null Hypothesis
b) Statistical Hypothesis
c) Simple Hypothesis
d) Composite Hypothesis

Ans: a) Null Hypothesis

8. If the Critical region is evenly distributed then the test is referred as?

a) Two tailed
b) One tailed
c) Three tailed
d) Zero tailed

Ans: a) Two tailed

9. Alternative Hypothesis is also called as?
a) Composite hypothesis
b) Research Hypothesis
c) Simple Hypothesis
d) Null Hypothesis

Ans: b) Research Hypothesis

10. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by
a) np
b) n

Ans: a) np

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 used to measure the data fit in regression model by taking the values between 0 to 1 always, the values we find from R-squared, the values more closer to the 1 is the best or the perfect fit for the goodness of the model whereas,

RSS is the sum of squared distance between the actual values and the predicted values from the model that is clearly showing that using the RSS only isn't the good thing to find the goodness of fit for the model in regression.

So we can say the R-squared is the best fit for goodness for fit in the regression model.

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 : *TSS* represents the variation in the dependent variable. It measures the total variability in the observed data points around the mean of the dependent variable. mathematically It can be calculated as the sum of the square differences between each observed data points and the mean of the dependent variable.

ESS quantifies how well the regression model fits the data by measuring the variability in the dependent variable that is explained by the independent variable in the regression model. mathematically it is the sum of the square difference between the predicted values and the mean of the dependent variable.

RSS measures the unexplained variability in the dependent variable that is not accounted by the regression model. It represents the squared difference between the observed values and the predicted values.

The equation relating these three matrix is known as the (FRR) Fundamental Relationship of Regression.

TSS = ESS + RSS

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

Ans: Regularization is a technique used in machine learning to prevent overfitting and improve the generalisation of the model. In machine learning, we need regularization for discouraging the model from assigning too much importance to individual or a coefficient .
We need the regularization in Machine Learning for more roles like Complexity Control, Preventing Overfitting, Balancing Bias and Variance, Features Selection, Handling multicollinearity and Generalization.

4. What is Gini–impurity index?

Ans: The Gini impurity index is measure of how often a randomly chosen element from the set would be incorrectly labelled if it was randomly labelled according to the distribution of levels in the subset.
It is used in decision tree algorithm, especially for binary classification Problem to evaluate the quality of a split
The impurity index ranges from 0 to 0.5 impurity of 0 indicates the set contain only one class while a impurity of 0.5 indicates maximum impurity, meaning an equal distribution of classes.

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

Ans: Yes Unregularized Decision Tree are prone to overfitting. Overfitting occurs when a model learns to capture Noise and specific detail of training data to such an extend that it performs poorly on unseen data, Decision tree especially when grew deep, have a high capacity to capture intricate details and noise in the training data, which can lead to overfitting. Several reasons for the presence of unregular decision tree to overfitting like - High Variance, Memorisation of Noise, Large tree size, no Pruning.

6. What is an ensemble technique in machine learning?

Ans: Ensemble techniques in machine learning involve combining multiple individual model to improve the overall performance of the systems. Instead of relying on a single model prediction ensemble method leverage the collective wisdom of multiple models to make more accurate prediction or classification.
Ensemble techniques are widely used across various machine learning and have proven to be highly effective in practice. Ensemble technique are valuable because they can reduce overfitting, improve generalisation and enhance robustness by levering the complimentary strength of different models.
There are multiple types of ensemble techniques like- Bagging (Boot Strap Aggregating), Boosting, Stacking(Stack Generalisation), Voting.

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

Ans: Difference Between Bagging and Boosting: Bagging vs Boosting are as follows -

	Bagging	**Boosting**
Basic Concept	Combines multiple models trained on different subsets of data.	Train models sequentially, focusing on the error made by the previous model.
Objective	To reduce variance by averaging out individual model error.	Reduces both bias and variance by correcting misclassifications of the previous model.
Data Sampling	Use Bootstrap to create subsets of the data.	Re-weights the data based on the error from the previous model, making the next models focus on misclassified instances.
Model Weight	Each model serves equal weight in the final decision.	Models are weighted based on accuracy, i.e., better-accuracy models will have a higher weight.
Error Handling	Each model has an equal error rate.	It gives more weight to instances with higher error, making subsequent model focus on them.
Overfitting	Less prone to overfitting due to average mechanism.	Generally not prone to overfitting, but it can be if the number of the model or the iteration is high.
Performance	Improves accuracy by reducing variance.	Achieves higher accuracy by reducing both bias and variance.
Common Algorithms	Random Forest	AdaBoost, XG Boost, Gradient Boosting Mechanism
Use Cases	Best for high variance, and low bias models.	Effective when the model needs to be adaptive to errors, suitable for both bias and variance errors.

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

Ans: OOB (out-of-bag) errors are an estimate of the performance of a random forest classifier or regressor on unseen data. The OOB error is computed using the samples that were not included in the training of the individual trees. This is different from the error computed using the usual training and validation sets, which are used to tune the hyperparameters of the random forest.

The OOB error can be useful for evaluating the performance of the random forest on unseen data. It is not always a reliable estimate of the generalization error of the model, but it can provide a useful indication of how well the model is performing.

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?**

Ans: K-fold cross-validation is a popular technique used in machine learning for assessing the performance and generalization ability of a predictive model. It involves partitioning the dataset into k equal-sized subsets (or "folds"), using k-1 subsets for training the model and the remaining subset for evaluating its performance. This process is repeated k times, with each subset used once as the validation data exactly once. The performance measures from each iteration are then averaged to provide a more robust estimate of the model's performance.

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

Ans: Hyperparameter tuning is the process of selecting the optimal values for a machine learning model's hyperparameters. Hyperparameters are settings that control the learning process of the model, such as the learning rate, the number of neurons in a neural network, or the kernel size in a support vector machine. The goal of hyperparameter tuning is to find the values that lead to the best performance on a given task.

A Machine Learning model is defined as a mathematical model with several parameters that need to be learned from the data. By training a model with existing data, we can fit the model parameters.

However, there is another kind of parameter, known as **Hyperparameters**, that cannot be directly learned from the regular training process. They are usually fixed before the actual training process begins. These parameters express important properties of the model such as its complexity or how fast it should learn. This article aims to explore various strategies to tune hyperparameters for Machine learning models.

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

Ans: Gradient Descent is a powerful optimization algorithm, it can also present some challenges that can affect its performance. Some of these challenges include:

1. Divergence: When the learning rate is too large, the algorithm may overshoot the minimum of the loss function, causing it to oscillate or diverge rather than converging to the optimal solution. This can result in unstable behavior and failure to find a good solution.

2. Unstable Updates: Large learning rates can result in large updates to the model parameters at each iteration. These large updates can lead to erratic behavior, causing the algorithm to bounce around the minimum of the loss function rather than converging smoothly towards it.

3. Missed Minima: With a large learning rate, the algorithm may jump over or completely miss local minima in the loss function. This can prevent the algorithm from finding the optimal solution or settling in a suboptimal region of the parameter space.

4. Oscillations: A large learning rate can lead to oscillations in the parameter updates, where the algorithm alternates between overshooting and undershooting the minimum of the loss function. This oscillatory behavior can slow down convergence and prevent the algorithm from reaching a stable solution.

5. Numerical Instability: Large learning rates can cause numerical instability, especially in models with high-dimensional parameter spaces or ill-conditioned optimization problems. This can manifest as numerical overflow or underflow, NaN values, or other computational errors.

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

Ans: Logistic Regression may struggle to accurately model complex non-linear relationships present in the data. For example, if the decision boundary between classes is highly non-linear or involves interactions between features that cannot be captured by simple linear relationships, Logistic Regression may not perform well. Logistic Regression may struggle to accurately model complex non-linear relationships present in the data. For example, if the decision boundary between classes is highly non-linear or involves interactions between features that cannot be captured by simple linear relationships, Logistic Regression may not perform well.
and also,
Non-linear problems can't be solved with logistic regression because it has a linear decision surface. Linearly separable data is rarely found in real-world scenarios.

13. Differentiate between Adaboost and Gradient Boosting.

Ans: 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.
While,
Gradient Boosting 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.

In summary, while both AdaBoost and Gradient Boosting involve training multiple weak learners sequentially and combining their predictions, AdaBoost focuses on correcting errors by assigning higher weights to misclassified examples, while Gradient Boosting fits each new model to the residuals (or gradients) of the previous model to minimize the overall loss function. Additionally, Gradient Boosting introduces a shrinkage parameter to control the contribution of each weak learner, which can improve model robustness and generalization.

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

Ans: The bias-variance trade off is a fundamental concept in machine learning that describes the relationship between the bias of a model and its variance, and how they influence the model's predictive performance.

The **Bias** is known as the difference between the prediction of the values by the Machine Learning model and the correct value. Being high in biasing gives a large error in training as well as testing data. It recommended that an algorithm should always be low-biased to avoid the problem of underfitting. By high bias, the data predicted is in a straight line format, thus not fitting accurately in the data in the data set. Such fitting is known as the Underfitting of Data. This happens when the hypothesis is too simple or linear in nature.

The variability of model prediction for a given data point which tells us the spread of our data is called the **Variance** of the model. The model with high variance has a very complex fit to the training data and thus is not able to fit accurately on the data which it hasn't seen before. As a result, such models perform very well on training data but have high error rates on test data. When a model is high on variance, it is then said to as Overfitting of Data. Overfitting is fitting the training set accurately via complex curve and high order hypothesis but is not the solution as the error with unseen data is high. While training a data model variance should be kept low.

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

Ans: Linear Kernel- 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.
When using a Linear Kernel in an SVM, the decision boundary is a linear hyperplane that separates the different classes in the feature space. This linear boundary can be useful when the data is already separable by a linear decision boundary or when dealing with high-dimensional data, where the use of more complex kernel functions may lead to overfitting.

Polynomial kernels- A particular kind of kernel function utilised in machine learning, such as in SVMs, is a **Polynomial Kernel** (Support Vector Machines). It is a nonlinear kernel function that employs polynomial functions to transfer the input data into a higher-dimensional feature space.

The polynomial kernel has the benefit of being able to detect both linear and nonlinear correlations in the data. It can be difficult to select the proper degree of the polynomial, though, as a larger degree can result in overfitting while a lower degree cannot adequately represent the underlying relationships in the data.

In general, the polynomial kernel is an effective tool for converting the input data into a higher-dimensional feature space in order to capture nonlinear correlations between the input characteristics.

RBF kernels- The Gaussian kernel, also known as the radial basis function (RBF) kernel, is a popular kernel function used in machine learning, particularly in SVMs (Support Vector Machines). It is a nonlinear kernel function that maps the input data into a higher-dimensional feature space using a Gaussian function.

One advantage of the Gaussian kernel is its ability to capture complex relationships in the data without the need for explicit feature engineering. However, the choice of the gamma parameter can be challenging, as a smaller value may result in under fitting, while a larger value may result in over fitting.