1. What is Machine Learning? State any two types of machine learning.

Machine learning is a field of artificial intelligence (AI) that enables computers to learn and improve from experience without being explicitly programmed. Two types of machine learning are:

Supervised Learning: The algorithm learns from labeled data, where the input and output are provided.

Unsupervised Learning: The algorithm learns from unlabeled data, finding hidden patterns or intrinsic structures.

2.Explain the type of supervised Machine Learning models with example.

Supervised learning models learn from labeled data, making predictions or decisions based on that data. Examples include:

Linear Regression: Predicting house prices based on features like size and location.

Support Vector Machines (SVM): Classifying emails as spam or not spam based on their content.

3.State few applications of Machine Learning.

Recommendation systems (e.g., Netflix recommendations).

Image and speech recognition.

Fraud detection in finance.

Medical diagnosis.

Autonomous vehicles.

4.Explain some issues with Machine Learning.

Overfitting: When a model is too complex and fits the training data too closely, leading to poor generalization.

Underfitting: When a model is too simple to capture the underlying patterns in the data.

Bias-variance tradeoff: Balancing model complexity to minimize both bias and variance.

5.How can you handle overfitting and underfitting?

To handle overfitting, you can use simpler models, add regularization, or use more training data.

To handle underfitting, you can use more complex models, add more features, or reduce regularization.

6.Explain data pre-processing steps.

Data preprocessing involves cleaning and preparing the data for analysis. Steps include:

Handling missing values.

Encoding categorical variables.

Feature scaling.

Splitting the data into training and testing sets.

7.What is Linear Regression in Machine Learning?

Linear regression is a supervised learning algorithm used to predict a continuous target variable based on one or more input features. It assumes a linear relationship between the features and the target variable.

8.Explain the assumptions of Linear Regression.

The assumptions of linear regression include:

Linearity: The relationship between the features and the target variable is linear.

Independence of errors: The errors are independent of each other.

Homoscedasticity: The variance of the errors is constant.

Normality of errors: The errors are normally distributed.

9. What are the benefits of using Linear Regression?

Linear regression is simple to implement, easy to interpret, and computationally efficient. It is also useful for understanding the relationships between variables.

10.Write the expression for the cost function, which needs to be minimized in Linear regression with RIDGE regularization.

The cost function for linear regression with ridge regularization is:

$$J(\theta) = \frac{1}{2m} \left(\sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^{n} \theta_j^2 \right)$$

Where λ is the regularization parameter and θ_j are the regression coefficients.

11.What is the difference between Classification and Regression problems?

Classification: Predicting a categorical label or class (e.g., spam or not spam).

Regression: Predicting a continuous value (e.g., house prices).

12.What is Multicollinearity? How to detect the presence of multicollinearity and which variables are involved in it?

Multicollinearity occurs when two or more independent variables in a regression model are highly correlated. It can be detected by calculating the variance inflation factor (VIF) for each variable, where VIF values greater than 10 indicate multicollinearity.

13.Write about different variance measures involved in the operation of Linear Regression.

In linear regression, the variance measures include:

Residual standard error (RSE): Measures the average distance that the observed values fall from the regression line.

R-squared: Measures the proportion of variance in the dependent variable that is predictable from the independent variables.

Adjusted R-squared: Adjusts the R-squared value for the number of predictors in the model.

14.How can the problem of overfitting be reduced in Linear Regression? What is bias-variance tradeoff?

To reduce overfitting in linear regression, you can:

Use simpler models.

Add regularization.

Use cross-validation.

The bias-variance tradeoff is the balance between a model's ability to capture the underlying patterns in the data (low bias) and its sensitivity to fluctuations in the training data (low variance).

15. **If $y = 2x_1 + 12x_2 + 3x_3 + 5$ is the linear regression equation, then explain how the coefficients of x_1 and x_2 affect the value of y.**
    - The coefficient of x_1 (2) indicates that a one-unit increase in x_1 results in a 2-unit increase in y.
    - The coefficient of x_2 (12) indicates that a one-unit increase in x_2 results in a 12-unit increase in y.

16.Discuss the use of Rsquared and adjusted Rsquared in a Linear Regression model?

R-squared (R^2) measures the proportion of variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1, where 1 indicates a perfect fit.

Adjusted R-squared adjusts the R-squared value for the number of predictors in the model, penalizing the addition of unnecessary variables.

17.How is adjusted R-square different from R-square? Brief the role of adjusted R-square in the feature selection process.

Adjusted R-squared penalizes the addition of unnecessary variables, unlike R-squared, which may increase even with the addition of irrelevant variables. Adjusted R-squared is used in the feature selection process to select the most relevant variables for the model.

18.What is k-fold cross-validation? Write briefly about the procedure.

K-fold cross-validation is a technique used to evaluate the performance of a machine learning model. It involves splitting the data into k subsets (folds), training the model on k-1 folds, and testing it on the remaining fold. This process is repeated k times, with each fold used as the test set once.

19.Explain Reciprocal Transformation Technique.

The reciprocal transformation involves taking the reciprocal (1/x) of each data point. This transformation is useful when the data is skewed and a linear relationship can be better modeled with the reciprocal values.

20.Explain the procedure involved in Forward Feature Selection.

Forward feature selection is a method. Here's a step-by-step explanation of the procedure:

Start with an empty set of features: Initially, the model does not contain any features.

Evaluate performance with each feature: Add one feature at a time to the model and evaluate its performance using a performance metric (e.g., accuracy, R-squared, etc.). This is typically done using cross-validation to ensure the results are robust.

Select the best-performing feature: Select the feature that improves the model's performance the most when added. This could be based on an increase in the performance metric or a decrease in the error metric.

Iterate: Repeat steps 2 and 3, adding one feature at a time and selecting the best-performing feature, until a stopping criterion is met. This criterion could be a predefined number of features or a threshold improvement in the performance metric.

Finalize the feature set: Once the stopping criterion is met, the selected subset of features is considered the final feature set for the model.

21. Explain Gradient Descent in brief

Gradient Descent: Gradient descent is an iterative optimization algorithm used to minimize a cost function by adjusting the parameters of a model. It works by calculating the gradient of the cost function with respect to each parameter and moving in the direction of the steepest decrease in the cost function. This process is repeated until the algorithm converges to a minimum of the cost function, at which point the optimal parameters are found.

22. What is Lasso Regularization?

Lasso Regularization: Lasso (Least Absolute Shrinkage and Selection Operator) regularization is a method used to prevent overfitting in linear regression models by adding a penalty term to the cost function. The penalty is proportional to the absolute value of the coefficients, which encourages the model to select a subset of features and shrink the coefficients of less important features to zero. This helps in feature selection and reduces model complexity.

23. A linear regression model is build with three independent variable price, advertisement cost and promotion cost to predict unit sales of mobile phone. Say the p value for the t-test of the variable 'advertisement cost' is 0.02. What is your inference on this?

Interpretation of p-value: A p-value of 0.02 for the t-test of the variable 'advertisement cost' indicates that there is a statistically significant relationship between the 'advertisement cost' and the 'unit sales of mobile phone'. In other words, the 'advertisement cost' is likely to have a significant impact on the 'unit sales of mobile phone'.

24. The RMSE of the regression model which predicting the CTC salary is 12324 and the RMSE of the other regression model which predicting the age of the person is 55. Comment on the performance of these two models. [output column is not scaled or transformed]

For the regression model predicting CTC salary with an RMSE of 12324, the model's predictions are off by an average of 12324 units from the actual values. This indicates that the model's performance is moderate, and there is room for improvement.

For the regression model predicting age with an RMSE of 55, the model's predictions are off by an average of 55 units from the actual values. This indicates that the model's performance is relatively good, with predictions being close to the actual values.

25. If we increase the value of lambda, what will happen to the estimated coefficients in RIDGE model?

Effect of increasing lambda in Ridge Regression: Increasing the value of lambda in a Ridge regression model increases the amount of regularization applied to the model. This leads to the estimated coefficients being shrunk towards zero, effectively reducing the impact of each feature on the model's predictions. As lambda increases, the model becomes more biased but less prone to overfitting, ultimately leading to a simpler model with potentially better generalization performance on unseen data.