1. Which type of machine learning is used when the output variable is continuous?  
    a) Classification  
    b) Clustering  
    c) Regression  
    d) Reinforcement Learning

Ans: we use **machine learning regression** for continuous prediction

1. In which scenario would you use classification instead of regression?  
    a) Predicting house prices based on area  
    b) Predicting whether an email is spam or not  
    c) Predicting the temperature of a city  
    d) Predicting the total sales of a store

Ans: I use classification instead of regression when I want to predict categorical value like **“Predicting whether an email is spam or not”**

1. Which of the following is an unsupervised learning algorithm?  
    a) Decision Tree  
    b) K-Means  
    c) Linear Regression  
    d) Logistic Regression

Ans: **K-means is a clustering and unsupervised algorithm** because it works for finding groups not continuous values

1. What does the term "overfitting" refer to in machine learning?  
    a) When the model does not learn from the data  
    b) When the model fits the training data too well but fails on new data  
    c) When the model generalizes well to unseen data  
    d) When the model cannot capture complex patterns

**Ans: The model works better for training data and fails on new data (unseen data)**

1. Which evaluation metric is most commonly used for classification problems?  
    a) Mean Squared Error (MSE)  
    b) R-squared  
    c) Accuracy  
    d) Root Mean Squared Error (RMSE)

**Ans:** We use **confusion matrix (Accuracy is one of parameter in CM)** as evaluation metrics for classification problems.

1. What is the main assumption of linear regression?  
    a) The relationship between dependent and independent variables is linear  
    b) The data is always normally distributed  
    c) There is no correlation between independent variables  
    d) It can only be used for categorical data

**Ans: The relationship between dependent and independent variables is linear**

1. Which clustering algorithm is best for detecting outliers?  
    a) K-Means  
    b) DBSCAN  
    c) Hierarchical Clustering d) Linear Regression

**Ans: we use DBDCAN for detecting outliers or noise.**

1. Which regularization technique adds both L1 and L2 penalties to a regression model?  
    a) Ridge Regression  
    b) Lasso Regression  
    c) Elastic Net  
    d) Polynomial Regression

**Ans: Elastic Net**

1. What is the purpose of the kernel trick in Support Vector Machines (SVM)?  
    a) To reduce training time  
    b) To map data into a higher-dimensional space for better classification  
    c) To normalize data  
    d) To make the model interpret results more easily

**Ans: To map data into a higher-dimensional space for better classification**

1. Which clustering algorithm does not require specifying the number of clusters in advance?  
    a) K-Means  
    b) DBSCAN  
    c) KNN  
    d) Decision Tree

**Ans: DBSCAN (It’s a density based algorithm so no need to specify number of clusters in advance)**

1. What type of function does logistic regression use to convert values into probabilities?  
    a) Linear function  
    b) Sigmoid function  
    c) Exponential function  
    d) Softmax function

**Ans: Sigmoid function because it convert values into probabilities by using “s curve” and thrushhold**

1. Which metric is commonly used to evaluate a clustering algorithm?  
    a) F1-score  
    b) Sum of Squared Errors (SSE)  
    c) Mean Squared Error  
    d) Log-Loss

**Ans: Sum of squared errors (SSE) and silhouette score**

1. If your classification model has a high recall but low precision, what does it mean?  
    a) The model is predicting too many false positives  
    b) The model is predicting too many false negatives  
    c) The model has low accuracy  
    d) The model is not generalizing well

**Ans: The model is predicting too many false positives (while calculating this two metrics TP not changed)**

1. Which technique can be used to reduce the dimensionality of a dataset before applying clustering?  
    a) Support Vector Machine  
    b) Principal Component Analysis (PCA)  
    c) Logistic Regression  
    d) Random Forest

**Ans: Principal Component Analysis (PCA)**

1. What is the main limitation of the K-Means clustering algorithm?  
    a) It does not scale well for large datasets  
    b) **It requires the number of clusters to be specified in advance**  
    c) It cannot handle numerical data  
    d) It always produces overlapping clusters