**1(A)** Train **SVM classifier** using sklearn digits dataset and then,

1. Measure accuracy of your model using different kernels such as rbf and linear.
2. Tune your model further using regularization and gamma parameters and try to come up with highest accuracy score.
3. Use 70% of samples as training data size.

**1(B)** Write a python Program to implement **K-Fold Cross Validation** Technique.

**1(C)**

1. What is an SVM and how does it work for classification tasks?
2. What are the differences between linear and RBF kernels in SVM?
3. How does the gamma parameter affect the RBF kernel in SVM?
4. What is the role of the regularization parameter C in SVM?
5. What does the value of k represent in K-Fold CV?

**2(A)** Use iris flower dataset from sklearn.datasets to predict flower species using **random forest classifier**.

1. Measure prediction score using default n\_estimator’s (10)
2. Now fine tune your model by changing number of trees in your classifier and write down what best score you can get using how many trees.

2(B) Write a python Program to implement **Stacking.**

2(C)

1. What is AdaBoost and how does it differ from Random Forest?How do you evaluate the performance of a Random Forest model?
2. How do you determine the best value for n\_estimators?
3. What does the parameter n\_estimators mean in Random Forest?
4. What is a Random Forest classifier and how does it work?
5. How does AdaBoost adjust weights for misclassified samples?

**3(A)** Use iris flower dataset from sklearn library and try to form **clusters o**f flowers using petal width and length features. Drop other two features for simplicity.

1. Figure out if any preprocessing such as scaling is required.
2. Draw elbow plot and from that figure out optimal value of k.

**3(B)** Write a python program to analyze and compare the performance of different classification algorithms on an Iris dataset.

**3(C)**

1. What is the purpose of the elbow plot in clustering?
2. Which classification algorithms did you use for comparing performance on the Iris dataset?
3. How do you evaluate and compare the performance of different classifiers?
4. Why is it important to split the dataset into training and testing sets before evaluation?
5. How does K-Nearest Neighbors (KNN) classify a new sample?

**4(A)**  Implementation of **Decision tree** using sklearn and its parameter tuning

**4(B**) From sklearn.datasets load digits dataset and do following

1. Classify digits (0 to 9) using **KNN classifier**. You can use different values for k neighbors and need to figure out a value of K that gives you a maximum score.
2. Plot confusion matrix
3. Plot classification report

**4 (C)**

1. How does the K-Nearest Neighbors (KNN) algorithm work for classification?
2. How did you determine the best value of k for the KNN classifier?
3. What is the purpose of a confusion matrix in classification tasks?
4. What insights can you derive from the classification report?
5. For Decision Trees, what are the important hyperparameters ??

**5(A)** Use sklearn.datasets iris flower dataset to train your model using **logistic regression**. You need to figure out accuracy of your model and use that to predict different samples in your test dataset.

1. Train the model and measure **accuracy**
2. Plot confusion matrix
3. Plot classification report

**5(B)** Using a dataset that includes individuals' age and whether they purchased life insurance (yes/no), answer the following:

**Data Preparation**

1. Load the dataset containing Age and Purchased columns.
2. Perform basic data exploration (e.g., shape, null values, data types).
3. Plot the distribution of data to understand class balance.

**Train a Logistic Regression Model**

1. Use binary logistic regression to model the probability of purchasing life insurance based on age.
2. Split the data into training (70%) and testing (30%) sets.
3. Train the model using LogisticRegression from sklearn.linear\_model.

**Evaluation**

1. Predict the labels on the test set.
2. Measure model performance using:
   * Accuracy score
   * Confusion matrix
   * Classification report (Precision, Recall, F1-score)

**5(C)**

1. What is logistic regression? How is it different from linear regression?
2. Why is logistic regression suitable for predicting life insurance purchase?
3. What does the sigmoid function output and what is its formula?
4. What is the purpose of the threshold value in logistic regression?
5. Why is it important to visualize the distribution of data before training?