Practical 1: PCA - Feature Transformation

PCA (Principal Component Analysis) is used for dimensionality reduction. It transforms high-dimensional data into a smaller number of principal components which capture maximum variance.

Steps:

1. Load dataset
2. Scale features using StandardScaler
3. Apply PCA
4. Visualize PC1 vs PC2 to distinguish classes

Viva Questions & Answers:

Q: What is PCA?
Ans: PCA reduces dimensions by extracting components capturing maximum variance.

Q: Why do we scale data before PCA?
Ans: To give equal weight to all features since PCA is affected by magnitude differences.

Q: How do we decide number of principal components?
Ans: Using explained variance ratio or elbow plot.

Practical 2: Uber Fare Prediction

Regression predicts continuous values like price. We applied Linear, Ridge, and Lasso.
Steps:
1. Data cleaning and preprocessing
2. Feature extraction (distance)
3. Outlier detection
4. Train regression models
5. Compare R2, MSE, RMSE

Viva Questions & Answers:

Q: Difference between Ridge and Lasso?
Ans: Ridge uses L2 penalty, Lasso uses L1 and performs feature selection.

Q: What is RMSE?
Ans: Root Mean Squared Error measures average model prediction error.

Q: Why remove outliers?
Ans: Because they affect regression line and reduce accuracy.

Practical 3: SVM Classification on Digits

SVM classifies digits by finding the best hyperplane that separates classes.
We use RBF kernel for non-linear data.
Steps:
1. Load dataset
2. Scale features
3. Train SVM
4. Evaluate with accuracy, confusion matrix

Viva Questions & Answers:

Q: What is SVM?
Ans: SVM finds optimal hyperplane to separate classes with maximum margin.

Q: Why use RBF kernel?
Ans: To handle non-linear decision boundaries.

Q: What are support vectors?
Ans: Points closest to the hyperplane that influence boundary.

Practical 4: K-Means Clustering on Iris

K-Means clusters data into K groups.
Elbow method is used to find optimal K.
Steps:
1. Load dataset
2. Scale data
3. Run K-Means for different K
4. Plot elbow graph

Viva Questions & Answers:

Q: What is K-Means?
Ans: Unsupervised algorithm to cluster data into K groups.

Q: What is Elbow method?
Ans: Plot inertia vs K to find optimal cluster count.

Q: Why scaling is required in clustering?
Ans: Distance based algorithm; large values dominate.

Practical 5: Random Forest Classifier

Random Forest is an ensemble of Decision Trees.
It prevents overfitting by averaging results.
Steps:
1. Load dataset
2. Encode categorical features
3. Train RandomForestClassifier
4. Evaluate accuracy, F1

Viva Questions & Answers:

Q: What is Random Forest?
Ans: Ensemble of decision trees reducing overfitting.

Q: Difference between Bagging & Boosting?
Ans: Bagging builds parallel models; Boosting sequentially improves weak models.

Q: What is feature importance?
Ans: Indicates contribution of each feature in prediction.

Practical 6: Reinforcement Learning - Maze

RL trains agent to learn from rewards & punishment.
Q-learning is used to update Q-values.
Steps:
1. Define states, actions, rewards
2. Initialize Q-table
3. Train agent over episodes

Viva Questions & Answers:

Q: What is Reinforcement Learning?
Ans: Agent learns by interacting and receiving rewards.

Q: What is Q-learning?
Ans: Updates Q-values based on reward and future value.

Q: What is exploration vs exploitation?
Ans: Choosing new actions vs using known best action.

Practical 7: Emotion Detection NLP

Emotion detection classifies text into emotions.
Steps:
1. Preprocessing (tokenization, lemma)
2. Convert to TF-IDF
3. Train Logistic Regression / Random Forest
4. Evaluate with F1 score

Viva Questions & Answers:

Q: What is NLP?
Ans: Field of AI that processes human language.

Q: What is TF-IDF?
Ans: Numerical value showing importance of a word in document.

Q: What is F1-score?
Ans: Harmonic mean of precision and recall.