Random Forest Assignments

1. Classification on the Wine Dataset

Assignment Task:

- Use the Wine dataset for a multi-class classification problem.
- Implement a Random Forest Classifier to predict the wine category based on various chemical properties.
- Evaluate the model's performance using cross-validation and provide a detailed report on the classification accuracy, precision, recall, and F1-score.

Additional Challenges:

- Compare the performance of Random Forest with other classifiers (e.g., Logistic Regression, SVM).
- Visualize feature importance and discuss which features have the most impact on the classification.

Dataset: Wine Dataset from UCI Machine Learning Repository

---

2. Regression on the California Housing Dataset

Assignment Task:

- Use the California Housing dataset to predict housing prices based on various factors like population, median income, and latitude/longitude.
- Implement a Random Forest Regressor and perform a full train-test split on the dataset.
- Evaluate the model performance using Mean Squared Error (MSE) and R-squared metrics. Plot learning curves and discuss how well the model generalizes.

Additional Challenges:

- Compare Random Forest Regressor's performance with other models like Decision Tree Regressor and Gradient Boosting Regressor.
- Discuss the impact of feature scaling on model performance.

Dataset: California Housing Dataset (Scikit-learn)

---

3. Classification and Regression on the Heart Disease Dataset

Assignment Task:

- Use the Heart Disease dataset to perform both classification and regression tasks.
    - Classification: Predict whether a patient has heart disease (binary classification) using Random Forest Classifier.
    - Regression: Predict the patient's maximum heart rate achieved during exercise using a Random Forest Regressor.
- Perform feature engineering to handle missing data, and try techniques like feature scaling and one-hot encoding for categorical variables.
-

Additional Challenges:

- Visualize the decision boundaries for the classifier.
- Compare the model’s performance against a baseline classifier (like k-NN) and regressor (like linear regression).