

## ML -Assignments – 22April 2025

### 1. Decision Tree Classification

Use `make_classification(n_samples=100, n_features=2)` to generate a binary classification dataset.

Train a `DecisionTreeClassifier`, predict, and print the **classification report**.

---

### 2. Decision Tree Regression

Generate data using `make_regression(n_samples=100, n_features=1, noise=10)`.

Train a `DecisionTreeRegressor` and evaluate using **MAE**, **RMSE**, and **R<sup>2</sup> score**.

---

### 3. Plot Decision Boundaries of a Decision Tree

Use a 2D dataset (`make_classification` with 2 features), fit a `DecisionTreeClassifier`, and plot the **decision boundary**.

---

### 4. Max Depth Impact in Decision Trees

Use the Iris dataset.

Train `DecisionTreeClassifier` with different `max_depth` values and compare **accuracy** on the test set.

---

### 5. Visualize Tree Structure

Train a decision tree on the Iris dataset and use `plot_tree` from `sklearn.tree` to visualize the structure of the tree.

---

## ◆ Random Forest

### 6. Random Forest Classification

Use the Breast Cancer dataset from `sklearn.datasets.load_breast_cancer`.

Train a `RandomForestClassifier`, compute **accuracy**, and display **feature importance**.

---

### 7. Random Forest Regression

Create data using `make_regression`.

Train a `RandomForestRegressor`, evaluate on test data using **R<sup>2</sup> and RMSE**, and visualize predicted vs actual values.

---

## 8. Compare Random Forest vs Decision Tree

On a classification dataset (`make_classification`), compare **accuracy, F1-score**, and **confusion matrix** between a `DecisionTreeClassifier` and a `RandomForestClassifier`.

---

### ◆ Gradient Boosting Machines (GBM)

## 9. Gradient Boosting Classifier

Use the Wine dataset (`load_wine()` from `sklearn`).

Train a `GradientBoostingClassifier` and compare accuracy with `RandomForestClassifier`.

---

## 10. Gradient Boosting Regressor

Generate regression data and train `GradientBoostingRegressor`.

Tune parameters like `n_estimators`, `learning_rate` and observe impact on **R<sup>2</sup> score**.

---

### ◆ Model Stacking

## 11. Stacking Classifier

On a classification dataset, stack `LogisticRegression`, `DecisionTreeClassifier`, and `RandomForestClassifier` using `sklearn.ensemble.StackingClassifier`.

Compare accuracy with individual base models.

---

## 12. Stacking Regressor

On regression data, implement stacking using `RandomForestRegressor`, `GradientBoostingRegressor`, and `LinearRegression`.

Measure RMSE.

---

### ◆ CatBoost

### 13. CatBoost Classifier

Use a dataset with categorical features (e.g., Titanic dataset or simulate one with `pandas`).  
Train a `CatBoostClassifier` and evaluate accuracy and F1-score.

---

#### ◆ LightGBM

### 14. LightGBM Classifier

Generate a dataset with 10+ features using `make_classification`.  
Train a `LGBMClassifier`, plot feature importance, and compare it to `RandomForestClassifier`.

---

#### ◆ XGBoost

### 15. XGBoost Regressor

Use `make_regression()` data.  
Train an `XGBRegressor`, tune parameters (`n_estimators`, `max_depth`, `learning_rate`), and evaluate using  $R^2$  and RMSE.

### 16. Compare Hierarchical Clustering with K-Means

**Dataset:** Use `make_blobs(n_samples=300, centers=4)`.

- Cluster the dataset using both `AgglomerativeClustering` and `KMeans`.
- Compare the cluster assignments visually.

### 17. Cluster Real-World Dataset (e.g., Mall Customer Segmentation)

**Dataset:** Use the `Mall_Customers.csv` dataset (Age, Income, Spending Score).

- Normalize features using `StandardScaler`.
- Apply `AgglomerativeClustering`.
- Visualize clusters using scatter plots.

### 18. Determine Optimal Number of Clusters Using Dendrogram

**Dataset:** `make_blobs(n_samples=100, centers=4, random_state=42)`

- Plot the dendrogram and cut it at different heights to determine the number of clusters.

## 19. Perform Agglomerative Clustering and Visualize Dendrogram

**Dataset:** Generate data using `make_blobs(n_samples=150, centers=3, cluster_std=1.0)`.

- Apply `AgglomerativeClustering` from `sklearn`.
- Use `scipy.cluster.hierarchy.dendrogram` to visualize the hierarchical tree.
- Plot clusters using different colors.

20. Use Breast cancer dataset and perform the above .