Decision Trees & Ensemble Methods

Topics Covered

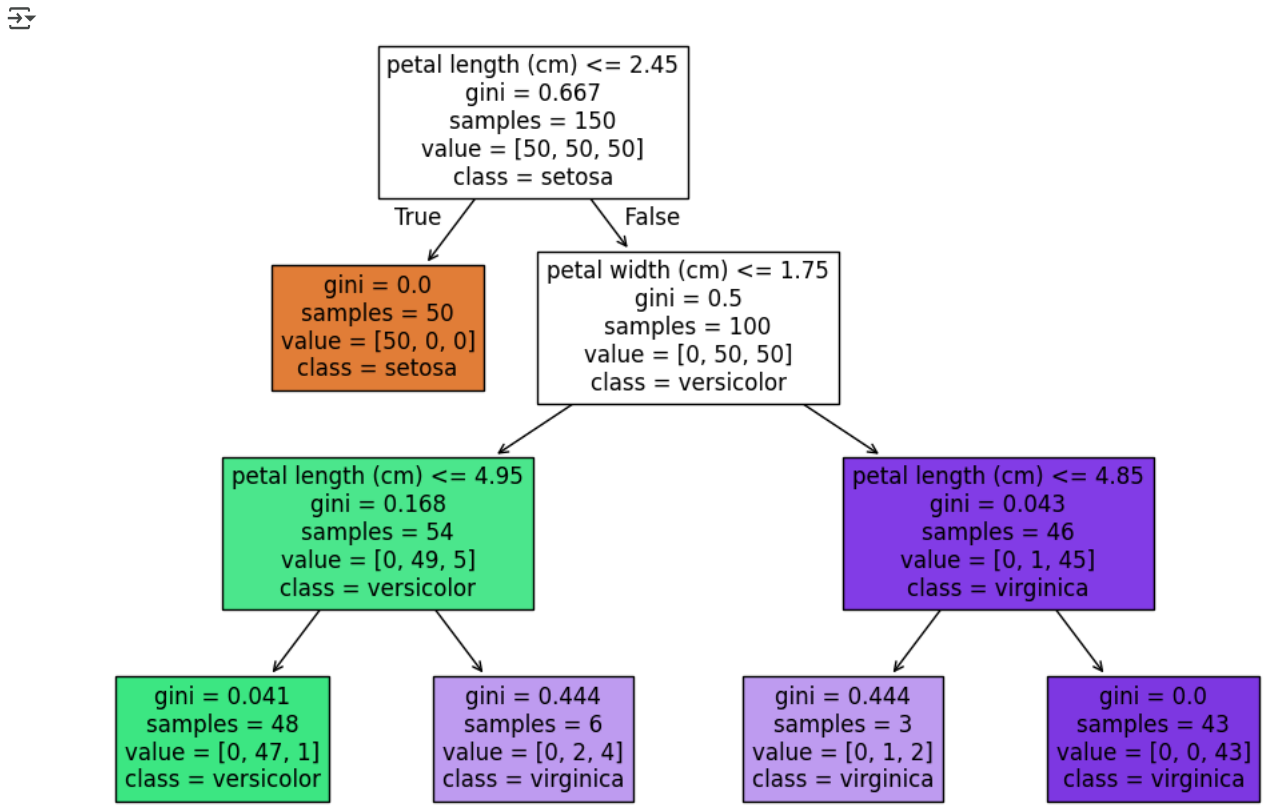
1. **Decision Trees**
   - Structure and intuition
   - Information Gain and Entropy
   - Gini Impurity
   - Overfitting and tree pruning
2. **Ensemble Methods**
   - Bagging
   - Random Forests
   - Boosting (e.g., AdaBoost)

```
# Decision Tree Classification Example
from sklearn.datasets import load_iris
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Load dataset
iris = load_iris()
X, y = iris.data, iris.target

# Fit Decision Tree
clf = DecisionTreeClassifier(max_depth=3, random_state=42)
clf.fit(X, y)

# Plot the tree
plt.figure(figsize=(12, 8))
plot_tree(clf, feature_names=iris.feature_names, class_names=iris.target_names, filled=True)
plt.show()
```




```
# Random Forest Example
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
```

```
# Split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train Random Forest
rf_clf = RandomForestClassifier(n_estimators=100, random_state=42)
rf_clf.fit(X_train, y_train)

# Evaluate
y_pred = rf_clf.predict(X_test)
print("Random Forest Accuracy:", accuracy_score(y_test, y_pred))
```

```
Random Forest Accuracy: 1.0
```

```
# AdaBoost Example
from sklearn.ensemble import AdaBoostClassifier

# Train AdaBoost
ada_clf = AdaBoostClassifier(n_estimators=50, random_state=42)
ada_clf.fit(X_train, y_train)

# Evaluate
y_pred_ada = ada_clf.predict(X_test)
print("AdaBoost Accuracy:", accuracy_score(y_test, y_pred_ada))
```

```
AdaBoost Accuracy: 1.0
```