Homework 5

Problem 1

(a) Partition the data into a 60/20/20 split for training/validation/testing and apply feature standardization (min-max standardization) to both the training and holdout data.

In [213…
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler

data = pd.read_csv('winequality-white.csv')

X = data.drop(columns=['quality'])
y = data['quality']

# 60/20/20
X_train, X_temp, y_train, y_temp = train_test_split(X, y, test_size=0.4, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)

# Apply feature standardization
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_val = scaler.transform(X_val)
X_test = scaler.transform(X_test)
```

(b) Start by fitting a single decision tree (as a weak learner) to the training data and report the prediction accuracy on the test data. Use `sklearn.tree.DecisionTreeClassifier` with default parameters.

In [214…
```
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score

# Train a decision tree
tree = DecisionTreeClassifier()
tree.fit(X_train, y_train)

y_pred = tree.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Decision Tree Test Accuracy: {accuracy:.2f}")
```

```
Decision Tree Test Accuracy: 0.56
```

(c) Now train several AdaBoost classifiers with an increasing number of estimators (i.e., increasing number of weak learners). Use `sklearn.ensemble.AdaBoostClassifier`. Search this parameter between 1 and 100 estimators. Plot the accuracy as a function of the number of estimators.

In [215…
```
from sklearn.ensemble import AdaBoostClassifier
import matplotlib.pyplot as plt

n_estimators_list = range(1, 101)
accuracies = []

for n in n_estimators_list:
    ada = AdaBoostClassifier(n_estimators=n)
    ada.fit(X_train, y_train)
    y_pred = ada.predict(X_test)
    accuracies.append(accuracy_score(y_test, y_pred))

plt.plot(n_estimators_list, accuracies)
plt.xlabel('Number of Estimators')
plt.ylabel('Accuracy')
plt.show()
```

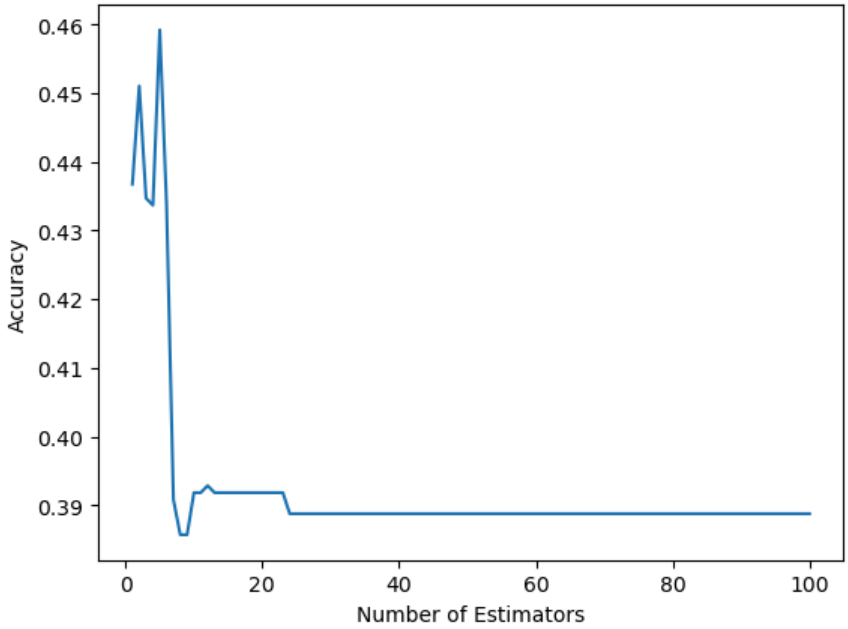
```
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
```

```
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
```

```
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
```

```
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
```

```
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
d:\ProgramData\Anaconda3\Lib\site-packages\sklearn\ensemble\_weight_boosting.py:519: FutureWarning: The SAMME.R a
lgorithm (the default) is deprecated and will be removed in 1.6. Use the SAMME algorithm to circumvent this warni
ng.
  warnings.warn(
```



(d) AdaBoostClassifier allows you to configure other parameters for the training algorithm, including learning rate and algorithm. Choose two additional parameters. Read the documentation and explain what they do. Try to improve on the model found in part (c) by searching for better values for those parameters

In [216…
```
ada = AdaBoostClassifier(n_estimators=50, learning_rate=0.1, algorithm='SAMME')
ada.fit(X_train, y_train)
y_pred = ada.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)

print(f"Improved AdaBoost Accuracy: {accuracy:.2f}")
```

```
Improved AdaBoost Accuracy: 0.51
```

(e) For the best AdaBoost model you found in part (d), report its accuracy and confusion matrix on the test set.

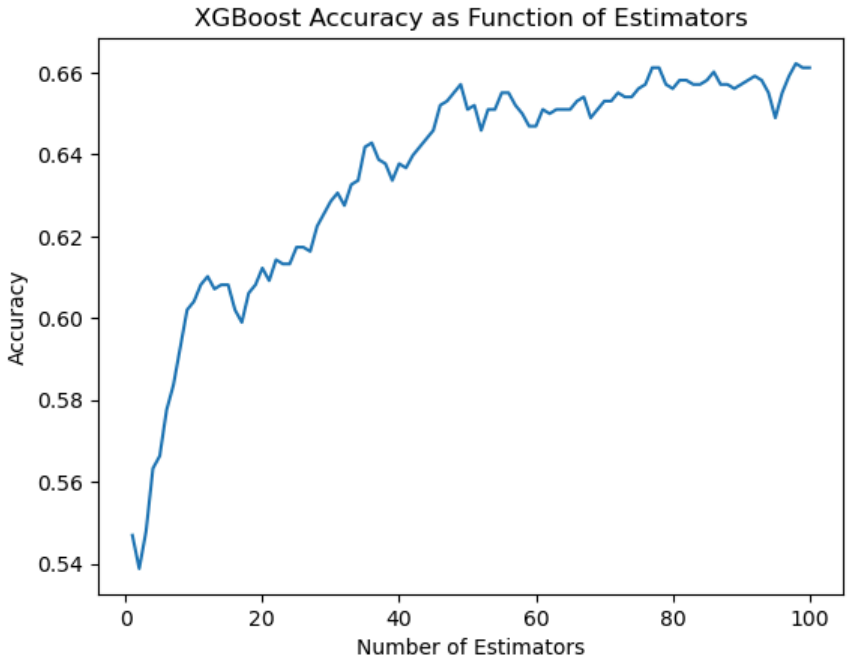
In [217…
```
from sklearn.metrics import confusion_matrix

conf_matrix = confusion_matrix(y_test, y_pred)
print(f"Confusion Matrix:\n{conf_matrix}")
```

```
Confusion Matrix:
[[  0   0   3   1   0   0]
 [  0   0  11   9   1   0]
 [  0   0 160 120   4   0]
 [  0   0 103 266  72   0]
 [  0   0  10 118  69   0]
 [  0   0   0  11  22   0]]
```

(f) Now, let's introduce XGBoost, another powerful boosting algorithm. Repeat steps (c), (d), and (e) using xgboost.XGBClassifier instead of AdaBoostClassifier. For XG-Boost, consider parameters such as max depth, learning rate, n estimators, min child weight, and subsample.

In [233…
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import accuracy_score, confusion_matrix
import matplotlib.pyplot as plt
from xgboost import XGBClassifier

data = pd.read_csv('winequality-white.csv')

X = data.drop(columns=['quality'])
y = data['quality']

# mapping label
label_mapping = {3: 0, 4: 1, 5: 2, 6: 3, 7: 4, 8: 5, 9: 6}
inverse_label_mapping = {v: k for k, v in label_mapping.items()}
y_mapped = y.map(label_mapping)

# 60/20/20
X_train, X_temp, y_train, y_temp = train_test_split(X, y_mapped, test_size=0.4, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, random_state=42)

# MinMaxScaler
```

```
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_val = scaler.transform(X_val)
X_test = scaler.transform(X_test)

# Train Multiple Models
n_estimators_list = range(1, 101)
accuracies = []

for n in n_estimators_list:
    xgb = XGBClassifier(n_estimators=n, eval_metric='mlogloss')
    xgb.fit(X_train, y_train)

    y_pred_mapped = xgb.predict(X_test)

    y_pred = pd.Series(y_pred_mapped).map(inverse_label_mapping)
    y_test_original = pd.Series(y_test).map(inverse_label_mapping)

    accuracies.append(accuracy_score(y_test_original, y_pred))

plt.plot(n_estimators_list, accuracies)
plt.xlabel('Number of Estimators')
plt.ylabel('Accuracy')
plt.title('XGBoost Accuracy as Function of Estimators')
plt.show()

conf_matrix = confusion_matrix(y_test_original, y_pred)
print(f"Confusion Matrix:\n{conf_matrix}")

xgb = XGBClassifier(n_estimators=50, eval_metric='mlogloss')
xgb.fit(X_train, y_train)
y_pred = xgb.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)

print(f"Improved XgBoost Accuracy: {accuracy:.2f}")
```




```
Confusion Matrix:
[[  0   1   2   1   0   0   0]
 [  0   5   7   9   0   0   0]
 [  0   3 193  83   5   0   0]
 [  0   1  75 326  39   0   0]
 [  0   1   2  74 112   8   0]
 [  0   0   0  10  10  12   1]
 [  0   0   0   0   0   0   0]]
Improved XgBoost Accuracy: 0.65
```

(g)

Xgboost has higher accuracy than Adaboost, shorter training time, and requires fewer estimators to achieve the best output.

(h)

XGBoost can often achieve better performance with fewer estimators for a given complex problem. Therefore, XGBoost is usually a better choice when you want to achieve high performance quickly.

Problem 2

In [130…

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score

data = pd.read_csv('Xray.csv')

X = data.iloc[:, :-1]
y = data.iloc[:, -1]

# 70/15/15
X_train, X_temp, y_train, y_temp = train_test_split(X, y, test_size=0.3, stratify=y, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, stratify=y_temp, random_state=42)

rf = RandomForestClassifier(n_estimators=100)
rf.fit(X_train, y_train)

y_val_pred = rf.predict(X_val)
y_test_pred = rf.predict(X_test)
val_accuracy = accuracy_score(y_val, y_val_pred)
test_accuracy = accuracy_score(y_test, y_test_pred)

conf_matrix = confusion_matrix(y_test, y_test_pred)

print(f"Validation Accuracy: {val_accuracy:.2f}")
print(f"Test Accuracy: {test_accuracy:.2f}")
print(f"Confusion Matrix:\n{conf_matrix}")
```

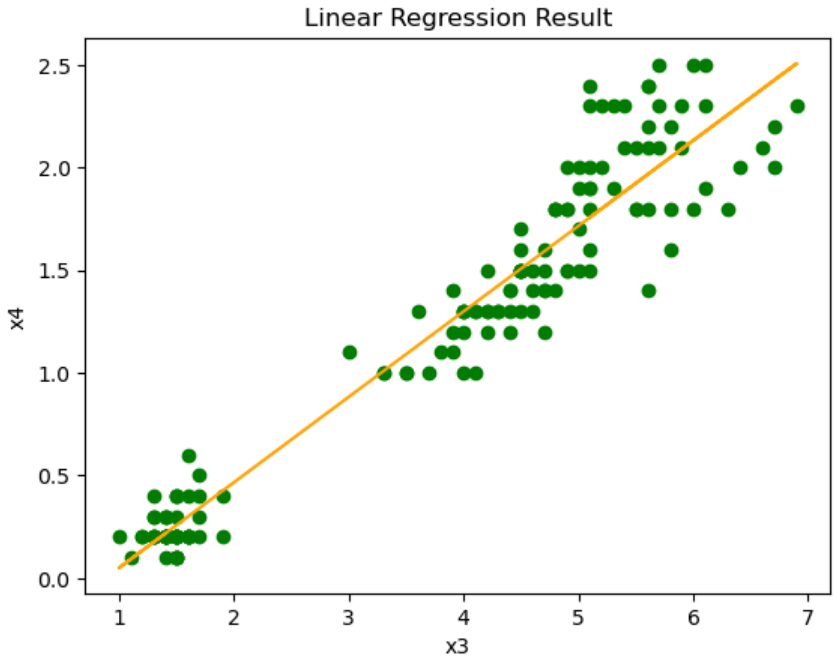
```
Validation Accuracy: 0.98
Test Accuracy: 0.96
Confusion Matrix:
[[58  2]
 [ 3 57]]
```

Problem 3

(a)

In [131…

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

data = pd.read_csv('iris.csv')
# print([key for key in data])

x3 = data.iloc[:,2].values.reshape(-1, 1)
x4 = data.iloc[:,3].values.reshape(-1, 1)

X = x3
y = x4

lr = LinearRegression()
lr.fit(X, y)

# 绘制模型
plt.scatter(X, y, color='green')
plt.plot(X, lr.predict(X), color='orange')
plt.xlabel('x3')
plt.ylabel('x4')
plt.title('Linear Regression Result')
plt.show()
```

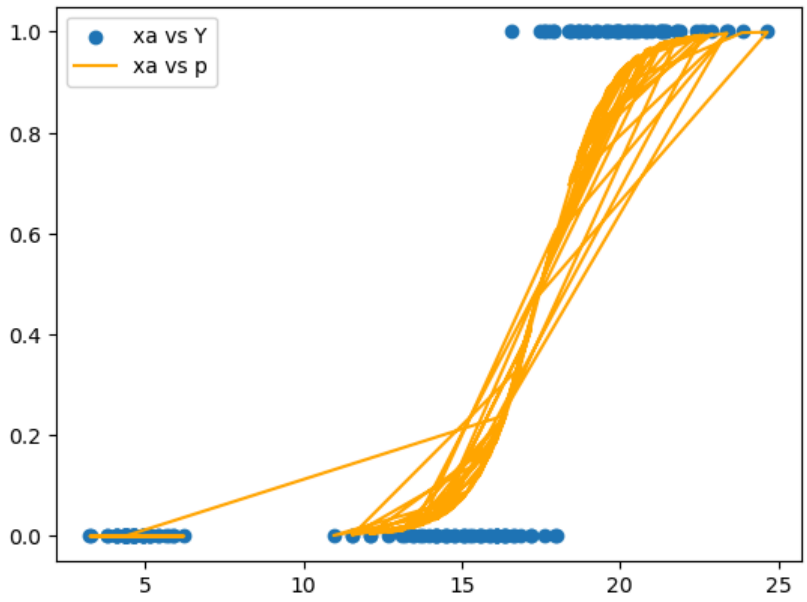


(b)

In [143…

```
from sklearn.linear_model import LogisticRegression

# x3, x4 Features
X_features = data.iloc[:, 2:4].values.reshape(-1, 2)
Y = (data.iloc[:, 4] == 3).astype(int)

print(X_features.shape)

log_reg = LogisticRegression()
log_reg.fit(X_features, Y)

w0 = log_reg.intercept_[0]
w1, w2 = log_reg.coef_[0]
print(f"Intercept: {w0}, Coefficients: w1={w1}, w2={w2}")

xa = w1 * data.iloc[:, 2] + w2 * data.iloc[:, 3]
p = log_reg.predict_proba(X_features)[:, 1]

plt.scatter(xa, y, label='xa vs Y')
plt.plot(xa, p, color='orange', label='xa vs p')
plt.legend()
plt.show()
```

```
(149, 2)
Intercept: -17.547204171639002, Coefficients: w1=2.780284083635839, w2=2.3772719716507127
```



Problem 4

(a)

In [154…

```
data = pd.read_csv('re_dat.csv')

X = data.iloc[:, 0].values.reshape(-1, 1)
y = data.iloc[:, 1].values.reshape(-1, 1)

# 90/10
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=42)
```

(b)

In [155…

```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import cross_val_score

for P in range(2, 16):
    poly = PolynomialFeatures(degree=P)
    X_poly = poly.fit_transform(X_train)
    model = LinearRegression()
    scores = cross_val_score(model, X_poly, y_train, cv=5)
    print(f"Degree {P}, Cross-Validation Score: {scores.mean():.2f}")
```

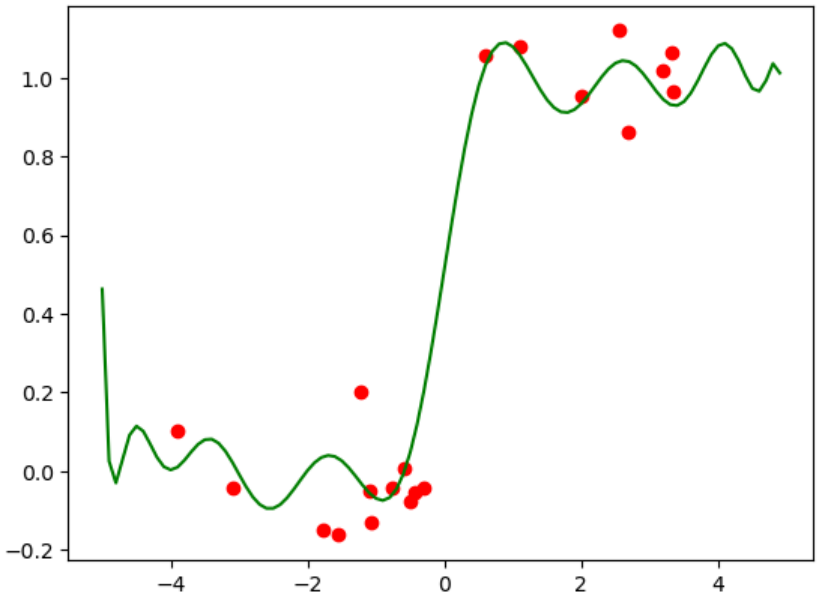
```
Degree 2, Cross-Validation Score: 0.70
Degree 3, Cross-Validation Score: 0.82
Degree 4, Cross-Validation Score: 0.82
Degree 5, Cross-Validation Score: 0.86
Degree 6, Cross-Validation Score: 0.86
Degree 7, Cross-Validation Score: 0.88
Degree 8, Cross-Validation Score: 0.88
Degree 9, Cross-Validation Score: 0.89
Degree 10, Cross-Validation Score: 0.88
Degree 11, Cross-Validation Score: 0.89
Degree 12, Cross-Validation Score: 0.89
Degree 13, Cross-Validation Score: 0.90
Degree 14, Cross-Validation Score: 0.89
Degree 15, Cross-Validation Score: 0.91
```

(c)

In [162…

```
poly = PolynomialFeatures(degree=15)
X_poly_train = poly.fit_transform(X_train)
X_poly_test = poly.transform(X_test)

model = LinearRegression()
model.fit(X_poly_train, y_train)

r2_score = model.score(X_poly_test, y_test)
print(f"R^2 Score: {r2_score:.2f}")

X_grid = np.arange(-5, 5, 0.1).reshape(-1, 1)
X_grid_poly = poly.transform(X_grid)
plt.plot(X_grid, model.predict(X_grid_poly), color='green')
plt.scatter(X_test, y_test, color='red')
plt.show()
```

```
R^2 Score: 0.94
```


1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−4
−2
0
2
4