SENTONGO HAMZA

MASTER’S IN DATA SCIENCE

QUESTION 1

First, in order to have a good K-means classifier we need to have time series data in a 2d format(Shape: **(n\_samples, n\_timestamps)**).Instead of random x and y data points. We don’t include the z because we are dealing with unsupervised learning.

Precautions

-same length of time series

-Data should be normalized or standardized to avoid scale bias.

-We should use feature extraction (mean, variance, frequency components) should be applied before clustering.

We have already practiced the above in the previous exercises

QUESTION 2

Recognition Accuracy Results:

SVC with linear kernel: 0.80

LinearSVC (linear kernel): 0.78

SVC with RBF kernel: 0.80

SVC with polynomial (degree 3) kernel: 0.78

A diagram of different sizes and colors

AI-generated content may be incorrect.

**CODE**

# -\*- coding: utf-8 -\*-

"""

Created on Tue Aug  6 10:14:14 2024

@author: turunenj

"""

#https://scikit-learn.org/stable/auto\_examples/svm/plot\_iris\_svc.html#sphx-glr-auto-examples-svm-plot-iris-svc-py

import matplotlib.pyplot as plt

from matplotlib import colormaps

from sklearn import datasets, svm, metrics

from sklearn.model\_selection import train\_test\_split

from sklearn.inspection import DecisionBoundaryDisplay

# Load Iris dataset

iris = datasets.load\_iris()

X = iris['data'][:, :2]   # Use only first two features

y = iris['target']

# Split into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

C = 1.0  # SVM regularization parameter

models = [

    ("SVC with linear kernel", svm.SVC(kernel="linear", C=C)),

    ("LinearSVC (linear kernel)", svm.LinearSVC(C=C, max\_iter=10000)),

    ("SVC with RBF kernel", svm.SVC(kernel="rbf", gamma=0.7, C=C)),

    ("SVC with polynomial (degree 3) kernel", svm.SVC(kernel="poly", degree=3, gamma="auto", C=C)),

]

# Train, test, and display accuracy for each model

print("Recognition Accuracy Results:")

for name, clf in models:

    clf.fit(X\_train, y\_train)

    y\_pred = clf.predict(X\_test)

    acc = metrics.accuracy\_score(y\_test, y\_pred)

    print(f"{name}: {acc:.2f}")

# Plot decision boundaries

fig, sub = plt.subplots(2, 2, figsize=(8, 8))

plt.subplots\_adjust(wspace=0.4, hspace=0.4)

X0, X1 = X[:, 0], X[:, 1]

for (name, clf), ax in zip(models, sub.flatten()):

    disp = DecisionBoundaryDisplay.from\_estimator(

        clf,

        X,

        response\_method="predict",

        cmap=colormaps['coolwarm'],

        alpha=0.8,

        ax=ax,

        xlabel=iris.feature\_names[0],

        ylabel=iris.feature\_names[1],

    )

    ax.scatter(X0, X1, c=y, cmap=colormaps['coolwarm'], s=20, edgecolors="k")

    ax.set\_title(name)

plt.show()

Question 3

A blue squares with white text

AI-generated content may be incorrect.

**Code**

# -\*- coding: utf-8 -\*-

"""

Modified SVM\_example.py to include 3x3 confusion matrix

"""

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn import datasets, svm, metrics

from sklearn.model\_selection import train\_test\_split

# Load Iris dataset

iris = datasets.load\_iris()

X = iris.data[:, :2]

y = iris.target

# Split data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=42)

# Train SVM (RBF kernel example)

clf = svm.SVC(kernel="rbf", gamma=0.7, C=1.0)

clf.fit(X\_train, y\_train)

y\_pred = clf.predict(X\_test)

# Create and plot 3x3 confusion matrix

cm = metrics.confusion\_matrix(y\_test, y\_pred)

plt.figure(figsize=(5, 4))

sns.heatmap(cm, annot=True, fmt="d", cmap="Blues",

            xticklabels=iris.target\_names, yticklabels=iris.target\_names)

plt.title("3x3 Confusion Matrix for RBF SVM")

plt.xlabel("Predicted")

plt.ylabel("Actual")

plt.tight\_layout()

plt.show()

Question 4