Лабораторна робота №6

Завдання 1:

Лістинг програми:

from sklearn.datasets import \_samples\_generator  
from sklearn.feature\_selection import SelectKBest, f\_regression  
from sklearn.pipeline import Pipeline  
from sklearn.ensemble import ExtraTreesClassifier  
  
X, y = \_samples\_generator.make\_classification(n\_samples=150,  
 n\_features=25,  
 n\_classes=3,  
 n\_informative=6,  
 n\_redundant=0,  
 random\_state=7  
 )  
k\_best\_selector = SelectKBest(f\_regression, k=9)  
classifier = ExtraTreesClassifier(n\_estimators=60, max\_depth=4)  
process\_pipeline = Pipeline([('selector', k\_best\_selector), ('erf', classifier)])  
  
  
process\_pipeline.set\_params(selector\_\_k=7, erf\_\_n\_estimators=30)  
process\_pipeline.fit(X, y)  
output = process\_pipeline.predict(X)  
print("\nPredicted output:\n", output)  
  
print("\nScore:", process\_pipeline.score(X, y))  
status = process\_pipeline.named\_steps['selector'].get\_support()  
selected = [i for i, x in enumerate(status) if x]  
print("\nIndices of selected features: ", ", ".join([str(x) for x in selected]))

A screenshot of a computer

Description automatically generated with medium confidence

Завдання 2:

*Лістинг програми:*

#Task 2  
import numpy as np  
import matplotlib.pyplot as plt  
from sklearn.neighbors import NearestNeighbors  
  
X = np.array([[2.1, 1.3], [1.3, 3.2], [2.9, 2.5], [2.7, 5.4],  
 [3.8, 0.9], [7.3, 2.1], [4.2, 6.5], [3.8, 3.7],  
 [2.5, 4.1], [3.4, 1.9], [5.7, 3.5], [6.1, 4.3],  
 [5.1, 2.2], [6.2, 1.1]])  
k = 5  
  
test\_datapoint = [4.3, 2.7]  
  
plt.figure()  
plt.title('Input data')  
plt.scatter(X[:, 0], X[:, 1], marker='o', s=75, color='black')  
  
  
knn\_model = NearestNeighbors(n\_neighbors=k, algorithm='ball\_tree').fit(X)  
  
distances, indices = knn\_model.kneighbors([test\_datapoint])  
  
print("\nK Nearest Neighbors: ")  
for rank, index in enumerate(indices[0][:k], start=1):  
 print(str(rank) + " ==>", X[index])  
plt.figure()  
plt.title('Nearest Neighbors')  
plt.scatter(X[:, 0], X[:, 1], marker='o', s=75, color='k')  
plt.scatter(X[indices][0][:][:,0], X[indices][0][:][:,1],  
 marker='o', s=250, color='k', facecolors='none')  
plt.scatter(test\_datapoint[0], test\_datapoint[1], marker='x', s=75, color='k')  
plt.show()

Chart, scatter chart

Description automatically generated

Завдання 3:

*Лістинг програми:*

#Task 3  
import numpy as np  
import matplotlib.pyplot as plt  
import matplotlib.cm as cm  
from sklearn import neighbors, datasets  
  
input\_file = 'data.txt'  
data = np.loadtxt(input\_file, delimiter=',')  
X, y = data[:, :-1], data[:, -1].astype(np.int)  
  
plt.figure()  
plt.title("InputData")  
marker\_shapes = 'v^os'  
mapper = [marker\_shapes[i] for i in y]  
for i in range(X.shape[0]):  
 plt.scatter(X[i, 0],  
 X[i, 1],  
 marker=mapper[i],  
 s=75,  
 edgecolors='black',  
 facecolors='none')  
  
num\_neighbors = 12  
step\_size = 0.01  
  
#-----  
Classifier = neighbors.KNeighborsClassifier(num\_neighbors,  
 weights='distance')  
Classifier.fit(X, y)  
  
x\_min, x\_max = X[:, 0].min() - 1, X[:, 0].max() + 1  
y\_min, y\_max = X[:, 1].min() - 1, X[:, 1].max() + 1  
x\_values, y\_values = np.meshgrid(np.arange(x\_min, x\_max, step\_size),  
 np.arange(y\_min, y\_max, step\_size))  
  
output = Classifier.predict(np.c\_[x\_values.ravel(),  
 y\_values.ravel()])  
  
output = output.reshape(x\_values.shape)  
plt.figure()  
plt.pcolormesh(x\_values, y\_values, output, cmap=cm.Paired)  
  
for i in range(X.shape[0]):  
 plt.scatter(X[i, 0], X[i, 1],  
 marker=mapper[i],  
 s=50,  
 edgecolors='black',  
 facecolors='none')  
  
plt.xlim(x\_values.min(), x\_values.max())  
plt.ylim(y\_values.min(), y\_values.max())  
plt.title('Edges model classifier on base K nearest neighbors')  
  
test\_datapoint = [5.1, 3.6]  
plt.figure()  
plt.title('Test data point')  
for i in range(X.shape[0]):  
 plt.scatter(X[i, 0], X[i, 1], marker=mapper[i],  
 s=75, edgecolors='black', facecolors='none')  
  
plt.scatter(test\_datapoint[0], test\_datapoint[1], marker='x', linewidths=6,  
 s=200, facecolors='black')  
  
\_, indices = Classifier.kneighbors([test\_datapoint])  
indices = indices.astype(np.int)[0]  
  
plt.figure()  
plt.title('K nearest neighbors')  
for i in indices:  
 plt.scatter(X[i, 0],  
 X[i, 1],  
 marker=mapper[y[i]],  
 linewidths=3,  
 s=100,  
 facecolors='black')  
plt.scatter(test\_datapoint[0], test\_datapoint[1], marker='x',  
 linewidths=6, s=200, facecolors='black')  
  
for i in range(X.shape[0]):  
 plt.scatter(X[i, 0], X[i, 1], marker=mapper[i], s=75,  
 edgecolors='black', facecolors='none')  
  
print('Predicted output:', Classifier.predict([test\_datapoint])[0])  
plt.show()

Scatter chart

Description automatically generated

Завдання 4

Лістинг програми:

#Task 4  
  
import argparse  
import json  
import numpy as np  
  
def build\_arg\_parser() :  
 parser = argparse.ArgumentParser(description='Compute similarity score')  
 parser.add\_argument('--user1', dest='user1', required=True, help='First user')  
 parser.add\_argument('--user2', dest='user2', required=True, help='Second user')  
 parser.add\_argument('--score-type',  
 dest='score\_type',  
 required=True,  
 choices=['Euclidean', 'Pearson'],  
 help='Similarity metric to be user')  
 return parser  
  
  
def euclidean\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 if len(common\_movies) == 0:  
 return 0  
  
 squared\_diff = []  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 squared\_diff.append(np.square(dataset[user1][item]  
 - dataset[user2][item]))  
 return 1 / (1 + np.sqrt(np.sum(squared\_diff)))  
  
def pearson\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 num\_ratings = len(common\_movies)  
 if num\_ratings == 0:  
 return 0  
  
 user1\_sum = np.sum([dataset[user1][item] for item in common\_movies])  
 user2\_sum = np.sum([dataset[user2][item] for item in common\_movies])  
  
 user1\_squared\_sum = np.sum([np.square(dataset[user1][item])  
 for item in common\_movies])  
 user2\_squared\_sum = np.sum([np.square(dataset[user2][item])  
 for item in common\_movies])  
  
 sum\_of\_products = np.sum([dataset[user1][item] \*  
 dataset[user2][item] for item in common\_movies])  
 Sxy = sum\_of\_products - (user1\_sum \* user2\_sum / num\_ratings)  
 Sxx = user1\_squared\_sum - np.square(user1\_sum) / num\_ratings  
 Syy = user2\_squared\_sum - np.square(user2\_sum) / num\_ratings  
  
 if Sxx \* Syy == 0:  
 return 0  
  
 return Sxy / np.sqrt(Sxx \* Syy)  
  
if \_\_name\_\_=='\_\_main\_\_':  
 args = build\_arg\_parser().parse\_args()  
 user1 = args.user1  
 user2 = args.user2  
 score\_type = args.score\_type  
  
ratings\_file = 'ratings.json'  
  
with open(ratings\_file, 'r') as f:  
 data = json.loads(f.read())  
  
if score\_type == 'Euclidean':  
 print("\nEuclidean score:")  
 print(euclidean\_score(data, user1, user2))  
else:  
 print("\nPearson score:")  
 print(pearson\_score(data, user1, user2))

Text

Description automatically generated with medium confidenceText

Description automatically generated

Text

Description automatically generated

Завдання 5:

Лістинг програми:

#Task 5  
import argparse  
import json  
import numpy as np  
  
  
def euclidean\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 if len(common\_movies) == 0:  
 return 0  
  
 squared\_diff = []  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 squared\_diff.append(np.square(dataset[user1][item]  
 - dataset[user2][item]))  
 return 1 / (1 + np.sqrt(np.sum(squared\_diff)))  
  
  
def pearson\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 num\_ratings = len(common\_movies)  
 if num\_ratings == 0:  
 return 0  
  
 user1\_sum = np.sum([dataset[user1][item] for item in common\_movies])  
 user2\_sum = np.sum([dataset[user2][item] for item in common\_movies])  
  
 user1\_squared\_sum = np.sum([np.square(dataset[user1][item])  
 for item in common\_movies])  
 user2\_squared\_sum = np.sum([np.square(dataset[user2][item])  
 for item in common\_movies])  
  
 sum\_of\_products = np.sum([dataset[user1][item] \*  
 dataset[user2][item] for item in common\_movies])  
 Sxy = sum\_of\_products - (user1\_sum \* user2\_sum / num\_ratings)  
 Sxx = user1\_squared\_sum - np.square(user1\_sum) / num\_ratings  
 Syy = user2\_squared\_sum - np.square(user2\_sum) / num\_ratings  
  
 if Sxx \* Syy == 0:  
 return 0  
  
 return Sxy / np.sqrt(Sxx \* Syy)  
  
  
def pearson\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 num\_ratings = len(common\_movies)  
 if num\_ratings == 0:  
 return 0  
  
 user1\_sum = np.sum([dataset[user1][item] for item in common\_movies])  
 user2\_sum = np.sum([dataset[user2][item] for item in common\_movies])  
  
 user1\_squared\_sum = np.sum([np.square(dataset[user1][item])  
 for item in common\_movies])  
 user2\_squared\_sum = np.sum([np.square(dataset[user2][item])  
 for item in common\_movies])  
  
 sum\_of\_products = np.sum([dataset[user1][item] \*  
 dataset[user2][item] for item in common\_movies])  
 Sxy = sum\_of\_products - (user1\_sum \* user2\_sum / num\_ratings)  
 Sxx = user1\_squared\_sum - np.square(user1\_sum) / num\_ratings  
 Syy = user2\_squared\_sum - np.square(user2\_sum) / num\_ratings  
  
 if Sxx \* Syy == 0:  
 return 0  
  
 return Sxy / np.sqrt(Sxx \* Syy)  
  
  
def find\_similar\_users(dataset, user, num\_users):  
 if user not in dataset:  
 raise TypeError('Cannot find ' + user + ' in the dataset')  
 scores = np.array([[x, pearson\_score(dataset, user, x)] for x in dataset if x != user])  
 scores\_sorted = np.argsort(scores[:, 1])[::-1]  
 top\_users = scores\_sorted[:num\_users]  
 return scores[top\_users]  
  
  
#user = "Bill Duffy"  
user = "Clarissa Jackson"  
ratings\_file = 'ratings.json'  
  
with open(ratings\_file, 'r') as f:  
 data = json.loads(f.read())  
  
print('\nUsers similar to ' + user + ':\n')  
similar\_users = find\_similar\_users(data, user, 3)  
print('User\t\t\tSimilarity score')  
print('-'\*41)  
  
for item in similar\_users:  
 print(item[0], '\t\t', round(float(item[1]), 2))

Text

Description automatically generatedText

Description automatically generated with medium confidence

Завдання 6:

Лістинг програми

#Task 6  
import argparse  
import json  
import numpy as np  
  
  
def pearson\_score(dataset, user1, user2):  
 if user1 not in dataset:  
 raise TypeError('Cannot find ' + user1 + ' in the dataset')  
 if user2 not in dataset:  
 raise TypeError('Cannot find ' + user2 + ' in the dataset')  
  
 common\_movies = {}  
  
 for item in dataset[user1]:  
 if item in dataset[user2]:  
 common\_movies[item] = 1  
  
 num\_ratings = len(common\_movies)  
 if num\_ratings == 0:  
 return 0  
  
 user1\_sum = np.sum([dataset[user1][item] for item in common\_movies])  
 user2\_sum = np.sum([dataset[user2][item] for item in common\_movies])  
  
 user1\_squared\_sum = np.sum([np.square(dataset[user1][item])  
 for item in common\_movies])  
 user2\_squared\_sum = np.sum([np.square(dataset[user2][item])  
 for item in common\_movies])  
  
 sum\_of\_products = np.sum([dataset[user1][item] \*  
 dataset[user2][item] for item in common\_movies])  
 Sxy = sum\_of\_products - (user1\_sum \* user2\_sum / num\_ratings)  
 Sxx = user1\_squared\_sum - np.square(user1\_sum) / num\_ratings  
 Syy = user2\_squared\_sum - np.square(user2\_sum) / num\_ratings  
  
 if Sxx \* Syy == 0:  
 return 0  
  
 return Sxy / np.sqrt(Sxx \* Syy)  
  
  
def build\_arg\_parser():  
 parser = argparse.ArgumentParser(description=  
 'Find users who are similar to the input user ')  
 parser.add\_argument('--user', dest='user', required=True, help='Input user')  
 return parser  
  
  
def find\_similar\_users(dataset, user, num\_users):  
 if user not in dataset:  
 raise TypeError('Cannot find ' + user + ' in the dataset')  
 scores = np.array([[x, pearson\_score(dataset, user, x)] for x in dataset if x != user])  
 scores\_sorted = np.argsort(scores[:, 1])[::-1]  
 top\_users = scores\_sorted[:num\_users]  
 return scores[top\_users]  
  
  
def get\_recommendations(dataset, input\_user):  
 if input\_user not in dataset:  
 raise TypeError('Cannot find ' + input\_user + ' in the dataset')  
  
 overall\_scores = {}  
 similarity\_scores = {}  
  
 for user in [x for x in dataset if x != input\_user]:  
 similarity\_score = pearson\_score(dataset, input\_user, user)  
 if similarity\_score <= 0:  
 continue  
 filtered\_list = [x for x in dataset[user] if x not in \  
 dataset[input\_user] or dataset[input\_user][x] == 0]  
  
 for item in filtered\_list:  
 overall\_scores.update({item: dataset[user][item] \* similarity\_score})  
 similarity\_scores.update({item: similarity\_score})  
  
 if len(overall\_scores) == 0:  
 return ["No recommendations possible"]  
  
 movie\_scores = np.array([[score/similarity\_scores[item],  
 item] for item, score in overall\_scores.items()])  
  
 movie\_scores = movie\_scores[np.argsort(movie\_scores[:, 0])[::-1]]  
  
 movie\_recommendations = [movie for \_, movie in movie\_scores]  
 return movie\_recommendations  
  
  
user = "Chris Duncan"  
#user = "Julie Hammel"  
  
  
ratings\_file = 'ratings.json'  
with open(ratings\_file, 'r') as f:  
 data = json.loads(f.read())  
  
print("\n Movie recommendations for " + user + ":")  
movies = get\_recommendations(data, user)  
for i, movie in enumerate(movies):  
 print(str(i+1) + '. ' + movie)

Text

Description automatically generatedGraphical user interface, text

Description automatically generated