

ЛАБОРАТОРНА РОБОТА № 3

ДОСЛІДЖЕННЯ МЕТОДІВ РЕГРЕСІЇ ТА НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

Посилання на GitHub: <https://github.com/FrancIwanicki/OAI.git>

Мета роботи: використовуючи спеціалізовані бібліотеки і мову програмування Python дослідити методи регресії та неконтрольованої класифікації даних у машинному навчанні.

2. ЗАВДАННЯ НА ЛАБОРАТОРНУ РОБОТУ ТА МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО ЙОГО ВИКОНАННЯ

Завдання 2.1. Створення регресора однієї змінної

Побудувати регресійну модель на основі однієї змінної.
Використовувати файл вхідних даних: data_singlevar_regr.txt.

Лістинг програми:

```
import pickle
import numpy as np
from sklearn import linear_model
import sklearn.metrics as sm
import matplotlib.pyplot as plt

# Вхідний файл, який містить дані
input_file = 'data_singlevar_regr.txt'

# Завантаження даних
data = np.loadtxt(input_file, delimiter=',')
X, y = data[:, :-1], data[:, -1]

num_training = int(0.8 * len(X))
num_test = len(X) - num_training

X_train, y_train = X[:num_training], y[:num_training]
X_test, y_test = X[num_training:], y[num_training:]

regressor = linear_model.LinearRegression()
regressor.fit(X_train, y_train)

y_test_pred = regressor.predict(X_test)

# Побудова графіка
plt.scatter(X_test, y_test, color='green')
plt.plot(X_test, y_test_pred, color='black', linewidth=4)
plt.xticks(())
plt.yticks(())
plt.show()

print("Linear regressor performance:")
print("Mean absolute error =",
      round(sm.mean_absolute_error(y_test, y_test_pred), 2))
print("Mean squared error =",
      round(sm.mean_squared_error(y_test, y_test_pred), 2))
```

					ДУ «Житомирська політехніка».23.121.8.000 – Лр3			
Змн.	Арк.	№ докум.	Підпис	Дата				
Розроб.		Іваницький Ф.А.			Звіт з лабораторної роботи		Літ.	Арк.
Перевір.		Голенко М. Ю.						1
Керівник							ФІКТ Гр. ІПЗ-20-3	
Н. контр.								
Зав. каф.								

```

print("Median absolute error =",
      round(sm.median_absolute_error(y_test, y_test_pred), 2))
print("Explain variance score =",
      round(sm.explained_variance_score(y_test, y_test_pred), 2))
print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))

# Файл для збереження моделі
output_model_file = 'model.pkl'
with open(output_model_file, 'wb') as f:
    pickle.dump(regressor, f)

with open(output_model_file, 'rb') as f:
    regressor_model = pickle.load(f)
y_test_pred_new = regressor_model.predict(X_test)
print("\nNew mean absolute error =",
      round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))

```

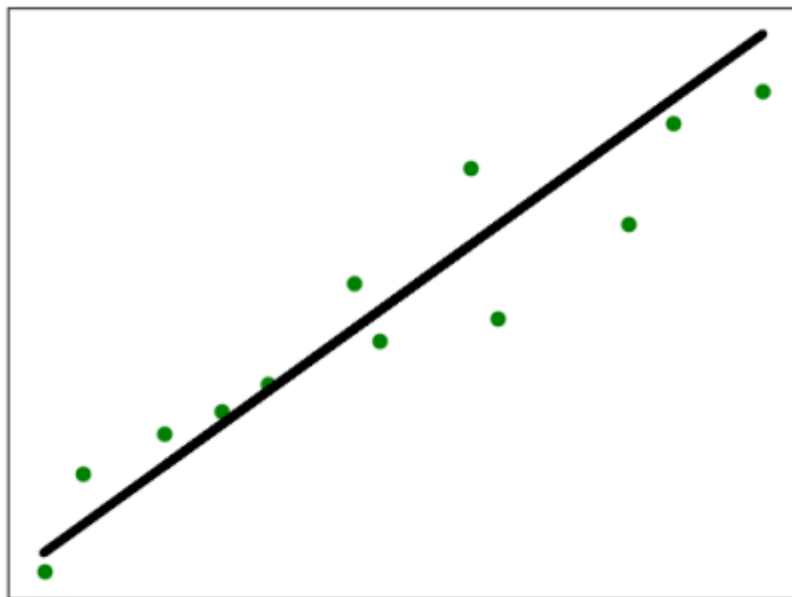
Результат виконання:

```

C:\Users\franc\OneDrive\Desktop\labs0AI\lab1\lab\Scripts\python.exe C:/Users/franc/OneDrive/Desktop/labs0AI/lab3/task1.py
Linear regressor performance:
Mean absolute error = 0.59
Mean squared error = 0.49
Median absolute error = 0.51
Explain variance score = 0.86
R2 score = 0.86

New mean absolute error = 0.59

```



Регресійний аналіз за допомогою лінійної регресії

Висновок: Модель лінійної регресії була піддана тренуванню на 80% даних і оцінена на залишкових 20%. Середня абсолютна помилка – 0.59, середня квадратична різниця – 0.49, медіана абсолютних помилок – 0.51, оцінка поясненої дисперсії – 0.86. При відновленні моделі з використанням збереженого файлу нова абсолютна помилка також дорівнює 0.59, це означає, що збережена та завантажена модель демонструють == результати.

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		2

Завдання 2.2. Передбачення за допомогою регресії однієї змінної

Побудувати регресійну модель на основі однієї змінної. Використовувати вхідні дані відповідно свого варіанту, що визначається за списком групи у журналі (таблиця 2.1).

Таблиця 2.1

№ за списком	1	2	3	4	5	6	7	8
№ варіанту	1	2	3	4	5	1	2	3

Варіант 3 файл: data_regr_3.txt

Лістинг програми:

```
import pickle
import numpy as np
from sklearn import linear_model
import sklearn.metrics as sm
import matplotlib.pyplot as plt

input_file = 'data_regr_3.txt'

data = np.loadtxt(input_file, delimiter=',')
X, y = data[:, :-1], data[:, -1]

num_training = int(0.8 * len(X))
num_test = len(X) - num_training

X_train, y_train = X[:num_training], y[:num_training]
X_test, y_test = X[num_training:], y[num_training:]

regressor = linear_model.LinearRegression()
regressor.fit(X_train, y_train)

y_test_pred = regressor.predict(X_test)

plt.scatter(X_test, y_test, color='green')
plt.plot(X_test, y_test_pred, color='black', linewidth=4)
plt.xticks(())
plt.yticks(())
plt.show()

print("Linear regressor performance:")
print("Mean absolute error =",
      round(sm.mean_absolute_error(y_test, y_test_pred), 2))
print("Mean squared error =",
      round(sm.mean_squared_error(y_test, y_test_pred), 2))
print("Median absolute error =",
      round(sm.median_absolute_error(y_test, y_test_pred), 2))
print("Explain variance score =",
      round(sm.explained_variance_score(y_test, y_test_pred), 2))
print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))

output_model_file = 'model2.pkl'
with open(output_model_file, 'wb') as f:
    pickle.dump(regressor, f)

with open(output_model_file, 'rb') as f:
    regressor_model = pickle.load(f)
y_test_pred_new = regressor_model.predict(X_test)
print("\nNew mean absolute error =",
      round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))
```

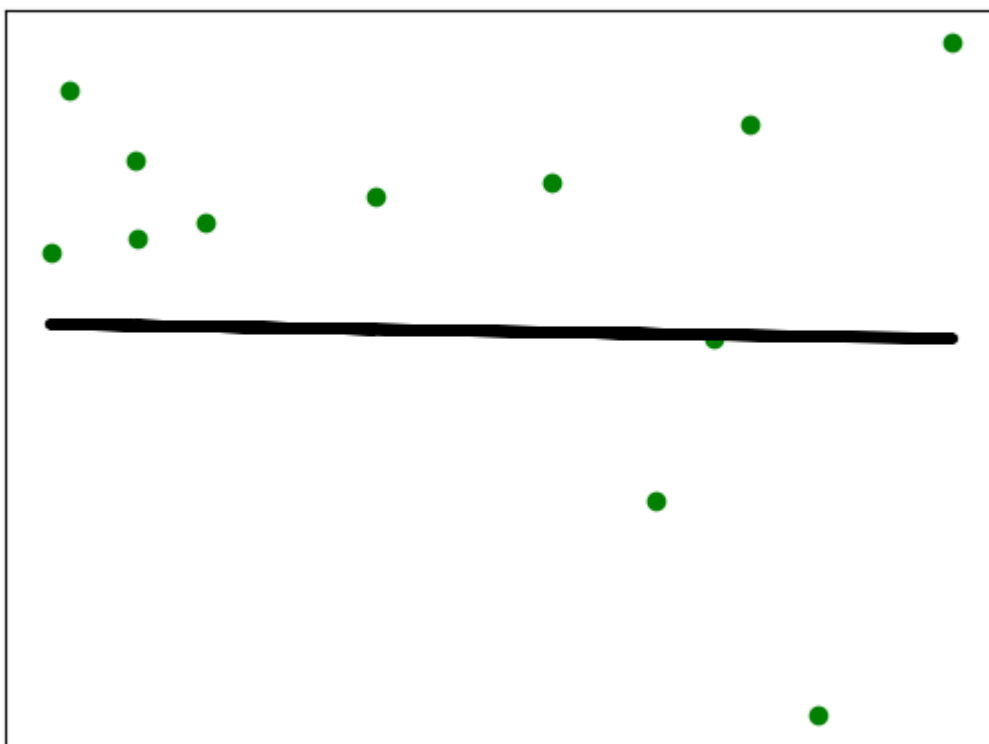
		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				3
Змн.	Арк.	№ докум.	Підпис	Дата		

Результат виконання:

```
sys.path.extend(['C:\\Users\\franc\\OneDrive\\Desktop\\labs0AI\\lab3', 'C:/Users/franc/OneDrive/Desktop/labs0AI/lab3'])

Python 3.10.1 (tags/v3.10.1:2cd268a, Dec 6 2021, 19:10:37) [MSC v.1929 64 bit (AMD64)]
Linear regressor performance:
Mean absolute error = 3.59
Mean squared error = 17.39
Median absolute error = 3.39
Explain variance score = 0.02
R2 score = -0.16

New mean absolute error = 3.59
```



Висновок: Модель лінійної регресії була піддана тренуванню на 80% даних і оцінена на залишкових 20%. Середня абсолютна помилка – 3.59, середня квадратична різниця – 17.39 (свідчить про великі відхилення у прогнозах), медіана абсолютних помилок – 3.39, оцінка поясненої дисперсії – 0.02 (прогнози значно відрізняються від фактичних значень). Нова абсолютна помилка для моделі, відновленої з використанням збереженого файлу, також дорівнює – 3.59, що підтверджує, що збережена і завантажена модель повертає ті ж самі результати.

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		4

Завдання 2.3. Створення багатовимірного регресора

Використовувати файл вхідних даних: data_multivar_regr.txt, побудувати регресійну модель на основі багатьох змінних.

Лістинг програми:

```
from math import degrees
import numpy as np
from sklearn import linear_model
import sklearn.metrics as sm
from sklearn.preprocessing import PolynomialFeatures

input_file = 'data_multivar_regr.txt'
data = np.loadtxt(input_file, delimiter=',')
X, y = data[:, :-1], data[:, -1]

num_training = int(0.8 * len(X))
num_test = len(X) - num_training
X_train, y_train = X[:num_training], y[:num_training]
X_test, y_test = X[num_training:], y[num_training:]

linear_regressor = linear_model.LinearRegression()
linear_regressor.fit(X_train, y_train)

y_test_pred = linear_regressor.predict(X_test)

print("Linear regressor performance:")
print("Mean absolute error =",
      round(sm.mean_absolute_error(y_test, y_test_pred), 2))
print("Mean squared error =",
      round(sm.mean_squared_error(y_test, y_test_pred), 2))
print("Median absolute error =",
      round(sm.median_absolute_error(y_test, y_test_pred), 2))
print("Explain variance score =",
      round(sm.explained_variance_score(y_test, y_test_pred), 2))
print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))

polynomial = PolynomialFeatures(degree=10)
X_train_transformed = polynomial.fit_transform(X_train)

datapoint = [[7.75, 6.35, 5.56]]
poly_datapoint = polynomial.fit_transform(datapoint)

poly_linear_model = linear_model.LinearRegression()
poly_linear_model.fit(X_train_transformed, y_train)
print("\nLinear regression:\n",
      linear_regressor.predict(datapoint))
print("\nPolynomial regression:\n",
      poly_linear_model.predict(poly_datapoint))
```

Результат виконання:

```
sys.path.extend(['C:\\Users\\franc\\OneDrive\\Desktop\\labs0AI\\lab3', 'C:/Users/franc/OneDrive/Desktop/labs0AI/lab3'])

Python 3.10.1 (tags/v3.10.1:2cd268a, Dec 6 2021, 19:10:37) [MSC v.1929 64 bit (AMD64)]
Linear regressor performance:
Mean absolute error = 3.58
Mean squared error = 20.31
Median absolute error = 2.99
Explain variance score = 0.86
R2 score = 0.86

Linear regression:
[36.05286276]

Polynomial regression:
[41.45561819]
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				5
Змн.	Арк.	№ докум.	Підпис	Дата		

Висновок: Середня абсолютна помилка становить – 3.58, середня квадратична різниця дорівнює – 20.31 (значення вище ніж попереднє, що свідчить про великій відхиління у прогнозах), медіана абсолютних помилок дорівнює – 2.99, оцінка поясненої дисперсії становить – 0.86 (прогнози схожі до фактичних значень). Для поліноміальної регресії використовується поліном 10-го ступеня, для вхідного значення [7.75, 6.35, 5.56], поліноміальна регресійна модель прогнозує значення – 41.46, в той час як лінійна регресія прогнозує значення – 36.05. Обидві моделі мають схожі значення показників, що свідчить про те, що обидві моделі надають приблизно однакові результати.

Завдання 2.4. Регресія багатьох змінних

Лістинг програми:

```
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, linear_model
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.model_selection import train_test_split

diabetes = datasets.load_diabetes()
X = diabetes.data
y = diabetes.target

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=0)

regr = linear_model.LinearRegression()
regr.fit(X_train, y_train)
y_pred = regr.predict(X_test)

print("Linear regressor performance:")
print("regr.coef_ =", regr.coef_)
print("regr.intercept_ =", regr.intercept_)
print("R2 score =", round(r2_score(y_test, y_pred), 2))
print("Mean absolute error =", round(mean_absolute_error(y_test, y_pred), 2))
print("Mean squared error =", round(mean_squared_error(y_test, y_pred), 2))

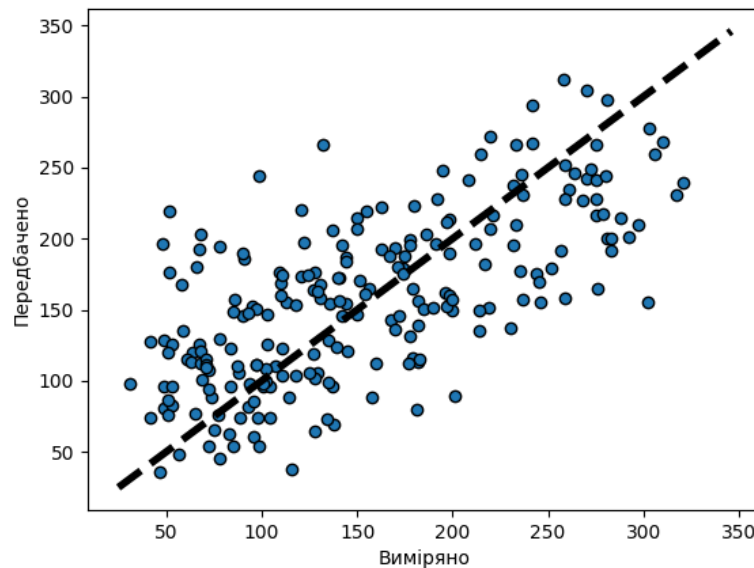
fig, ax = plt.subplots()
ax.scatter(y_test, y_pred, edgecolors=(0, 0, 0))
ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
ax.set_xlabel('Виміряно')
ax.set_ylabel('Передбачено')
plt.show()
```

Результат виконання:

```
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\franc\\OneDrive\\Desktop\\labs0AI\\lab3', 'C:/Users/franc/OneDrive/Desktop/labs0AI/lab3'])

Python 3.10.1 (tags/v3.10.1:2cd268a, Dec 6 2021, 19:10:37) [MSC v.1929 64 bit (AMD64)]
Linear regressor performance:
regr.coef_ = [ -20.4047621  -265.88518066  564.65086437  325.56226865 -692.16120333
  395.55720874   23.49659361  116.36402337  843.94613929  12.71856131]
regr.intercept_ = 154.3589285280134
R2 score = 0.44
Mean absolute error = 44.8
Mean squared error = 3075.33
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				6
Змн.	Арк.	№ докум.	Підпис	Дата		



Висновок: Було виконано лінійну регресію для набору даних «Diabetes» та було отримано такі результати якості: **Коефіцієнти регресії** представляються як масив чисел, вони вказують на вагу кожної ознаки $[-20.4047621, -265.88518, 564.65086437, 325.56226865, 692.16120333, 395.55720874, 23.49659361, 116.36402337, 843.94613929, 12.71856131]$, **Перетин** рівний -154.36 і представляє відсоток, на який зміщується пряма регресії, **Оцінка R^2** дорівнює -0.44 (модель пояснює близько 44% варіації в цільовій змінній), Середня Абсолютна різниця становить 44.8, Середня Квадратична Оцінка дорівнює -3075.33 (показує різницю між прогнозованим і фактичним значеннями). Модель має обмежену ефективність, що підтверджує значення R^2)

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				7
Змн.	Арк.	№ докум.	Підпис	Дата		

Завдання 2.5. Самостійна побудова регресії

Згенеруйте свої випадкові дані обравши за списком відповідно свій варіант (згідно табл. 2.2) та виведіть їх на графік. Побудуйте по них модель лінійної регресії, виведіть на графік. Побудуйте по них модель поліноміальної регресії, виведіть на графік. Оцініть її якість.

Таблиця 2.2

Лістинг програми:

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

m = 100
X = 6 * np.random.rand(m, 1) - 5
y = 0.7 * X ** 2 + X + 3 + np.random.randn(m, 1)

fig, ax = plt.subplots()
ax.scatter(X, y, edgecolors=(0, 0, 0))
plt.show()

poly_features = PolynomialFeatures(degree=3, include_bias=False)
X_poly = poly_features.fit_transform(X)
print("X[0] = ",X[0])
print("X[1] = ",X[1])
print("Y[1] = ",y[1])
lin_reg = linear_model.LinearRegression()
lin_reg.fit(X_poly, y)
print("Перетин:",lin_reg.intercept_,"Коефіцієнти регресії", lin_reg.coef_)
y_pred = lin_reg.predict(X_poly)

print("X_poly = ",X_poly)

fig, ax = plt.subplots()
ax.scatter(X, y, edgecolors=(0, 0, 0))
plt.plot(X, y_pred, color='red', linewidth=4)
plt.show()
```

Результат виконання:

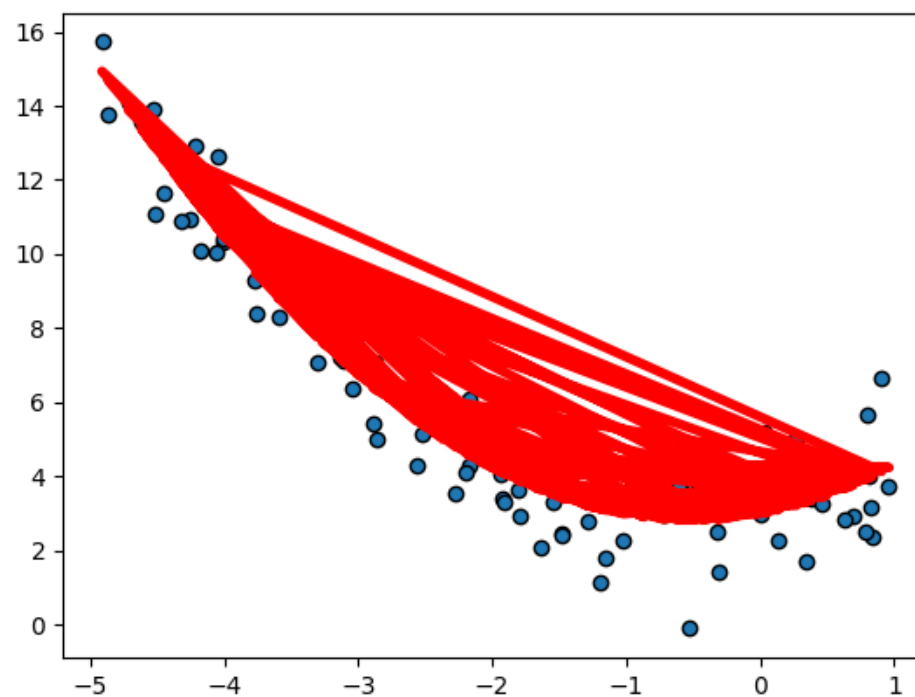
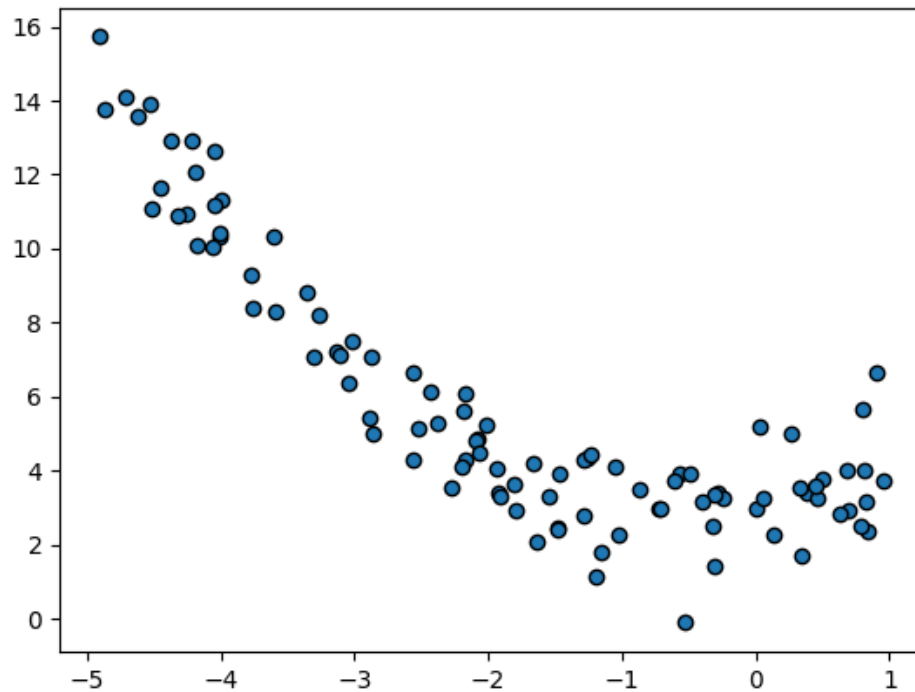
		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				8
Змн.	Арк.	№ докум.	Підпис	Дата		


```

sys.path.extend(['C:\\Users\\franc\\OneDrive\\Desktop\\labs0AI\\lab3', 'C:/Users
Python 3.10.1 (tags/v3.10.1:2cd268a, Dec 6 2021, 19:10:37) [MSC v.1929 64 bit
X[0] = [-4.01317516]
X[1] = [0.8012889]
Y[1] = [5.64220611]
Перетин: [2.99584012] Коефіцієнти регресії [[0.66414778 0.65211804 0.00454916]]
X_poly = [[-4.01317516e+00 1.61055749e+01 -6.46344929e+01]
[ 8.01288901e-01 6.42063903e-01 5.14478679e-01]
[-4.17333037e+00 1.74166864e+01 -7.26855863e+01]
[ 3.77865303e-01 1.42782187e-01 5.39524345e-02]
[-3.59913673e+00 1.29537852e+01 -4.66224440e+01]
[ 3.39637864e-01 1.15353879e-01 3.91785451e-02]
[-1.79217935e+00 3.21190684e+00 -5.75631312e+00]
[-7.29986110e-01 5.32879720e-01 -3.88994794e-01]
[-4.45354221e+00 1.98340383e+01 -8.83317267e+01]
[-2.86337375e+00 8.19890922e+00 -2.34765414e+01]
[-8.67990017e-01 7.53406669e-01 -6.53949467e-01]
[-2.88940630e+00 8.34866878e+00 -2.41226962e+01]
[ 4.59803246e-01 2.11419025e-01 9.72111538e-02]
[-1.47959782e+00 2.18920970e+00 -3.23914989e+00]
[-2.16539144e+00 4.68892008e+00 -1.01533474e+01]
[-4.71522086e+00 2.22333078e+01 -1.04834957e+02]
[-1.81138717e+00 3.28112347e+00 -5.94338493e+00]
[ 1.32494597e-01 1.75548182e-02 2.32591856e-03]
[-4.05682519e+00 1.64578306e+01 -6.67665419e+01]
[-2.56637758e+00 6.58629389e+00 -1.69029170e+01]
[-4.91225989e+00 2.41302972e+01 -1.18534291e+02]
[-1.54983116e+00 2.40197663e+00 -3.72265824e+00]
[-4.86426683e+00 2.36610918e+01 -1.15093864e+02]
[-3.04716379e+00 9.28520719e+00 -2.82935472e+01]
[-4.25578424e+00 1.81116995e+01 -7.70794853e+01]
[ 8.38651098e-01 7.03335665e-01 5.89853228e-01]
[-2.07368003e+00 4.30014885e+00 -8.91713278e+00]
[ 9.61121735e-01 9.23754989e-01 8.87840998e-01]
[-2.09226408e+00 4.37756899e+00 -9.15903036e+00]
[-3.99453563e+00 1.59563149e+01 -6.37380682e+01]
[-2.87719866e-01 8.27827214e-02 -2.38182335e-02]
[-3.13117742e+00 9.80427201e+00 -3.06989151e+01]
[-1.04787200e+00 1.09803572e+00 -1.15060089e+00]
[-2.19268283e+00 4.80785799e+00 -1.05421077e+01]
[-1.25271647e+00 1.56929855e+00 -1.96588614e+00]
[-2.48409777e-01 6.17074174e-02 -1.53287258e-02]
[ 6.84722636e-01 4.68845088e-01 3.21028845e-01]
[-3.76109166e+00 1.41458105e+01 -5.32036898e+01]
[-3.20070753e-01 1.02445287e-01 -3.27897400e-02]
[-3.02050361e-01 9.12344204e-02 -2.75573896e-02]
[-1.92831393e+00 3.71839460e+00 -7.17023210e+00]
[-5.29645949e-01 2.80524831e-01 -1.48578841e-01]
[-4.53285344e+00 2.05467603e+01 -9.31354530e+01]
[ 9.06533559e-01 8.21803094e-01 7.44992084e-01]
[-2.87083348e+00 8.24168489e+00 -2.36605049e+01]
[-1.47239632e+00 2.16795094e+00 -3.19208299e+00]
[ 3.33262573e-01 1.11063943e-01 3.70134554e-02]
[-4.21383558e+00 1.77564103e+01 -7.48225937e+01]
[-3.10985811e+00 9.67121743e+00 -3.00761139e+01]
[-5.67422063e-01 3.21967798e-01 -1.82691632e-01]
[-1.28219999e+00 1.64403681e+00 -2.10798399e+00]
[-4.37702295e+00 1.91583299e+01 -8.38564494e+01]
[-4.18963687e+00 1.75530571e+01 -7.35409354e+01]
[-3.30627754e+00 1.09314712e+01 -3.61424776e+01]
[-4.32627858e+00 1.87166863e+01 -8.09735992e+01]
[-1.23596221e+00 1.52760258e+00 -1.88805906e+00]
[-4.91058878e-01 2.41138821e-01 -1.18413359e-01]
[-4.04234687e+00 1.63405682e+01 -6.60542449e+01]
[-4.61851637e+00 2.13306935e+01 -9.85161570e+01]
[-4.51966769e+00 2.04273961e+01 -9.23250421e+01]
[-1.63951316e+00 2.68800339e+00 -4.40701691e+00]
[-2.56573411e+00 6.58299151e+00 -1.68902059e+01]
[-3.27049271e+00 1.06961225e+01 -3.49815907e+01]
[-2.18083281e+00 4.75603174e+00 -1.03721101e+01]
[-7.15646178e-01 5.12149453e-01 -3.66517798e-01]

```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – ЛрЗ	Арк.
		Голенко М. Ю.				9
Змн.	Арк.	№ докум.	Підпис	Дата		



Графіки моделі

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – ЛрЗ	Арк.
		Голенко М. Ю.				10
Змн.	Арк.	№ докум.	Підпис	Дата		

Висновок: Математичне рівняння для даної моделі:

$$y = 0.7x_1^2 + 1.0x_1 + 3 + \text{гауссов шум}$$

Отримана модель регресії з передбаченими коефіцієнтами:

$$y = 3.07618256 + 1.09246979x_1 + 0.61181343x_2 - 0.02138683x_3$$

Результати регресії:

$$X[0] = [-1.59591606]$$

$X[1] = [-0.48231179]$ – значення X для першого спостереження.

$Y[1] = [3.4939998]$ – відповідне значення Y для першого спостереження.

Перетин: $[2.68172353]$ Коефіцієнти регресії $[0.97095503, 0.80659551, 0.02275606]$

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				11
Змн.	Арк.	№ докум.	Підпис	Дата		

Завдання 2.6. Побудова кривих навчання

Побудуйте криві навчання для ваших даних у попередньому завданні.

Лістинг програми:

```
import matplotlib.pyplot as plt

import numpy as np

from sklearn import linear_model

from sklearn.metrics import mean_squared_error

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import PolynomialFeatures

from sklearn.pipeline import Pipeline


m = 100

X = 6 * np.random.rand(m, 1) - 5

y = 0.7 * X ** 2 + X + 3 + np.random.randn(m, 1)


def ivanytskyi_plot_learning_curves(model, X, y):

    X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2)

    train_errors, val_errors = [], []

    for m in range(1, len(X_train)):

        model.fit(X_train[:m], y_train[:m])

        y_train_predict = model.predict(X_train[:m])

        y_val_predict = model.predict(X_val)

        train_errors.append(mean_squared_error(y_train_predict, y_train[:m]))

        val_errors.append(mean_squared_error(y_val_predict, y_val))

    fig, ax = plt.subplots()

    plt.ylim(0, 2)

    ax.plot(np.sqrt(train_errors), "r-+", linewidth=2, label='train')

    ax.plot(np.sqrt(val_errors), "b-", linewidth=3, label='val')

    plt.show()
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. ІО.				12
Змн.	Арк.	№ докум.	Підпис	Дата		

```

lin_reg = linear_model.LinearRegression()

ivanytskyi_plot_learning_curves(lin_reg, X, y)

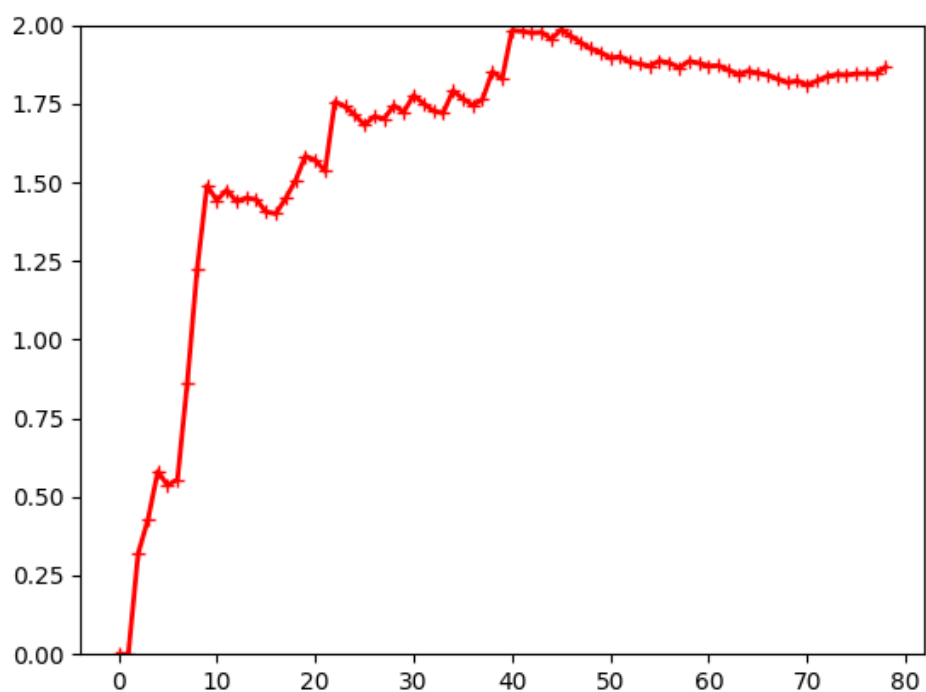
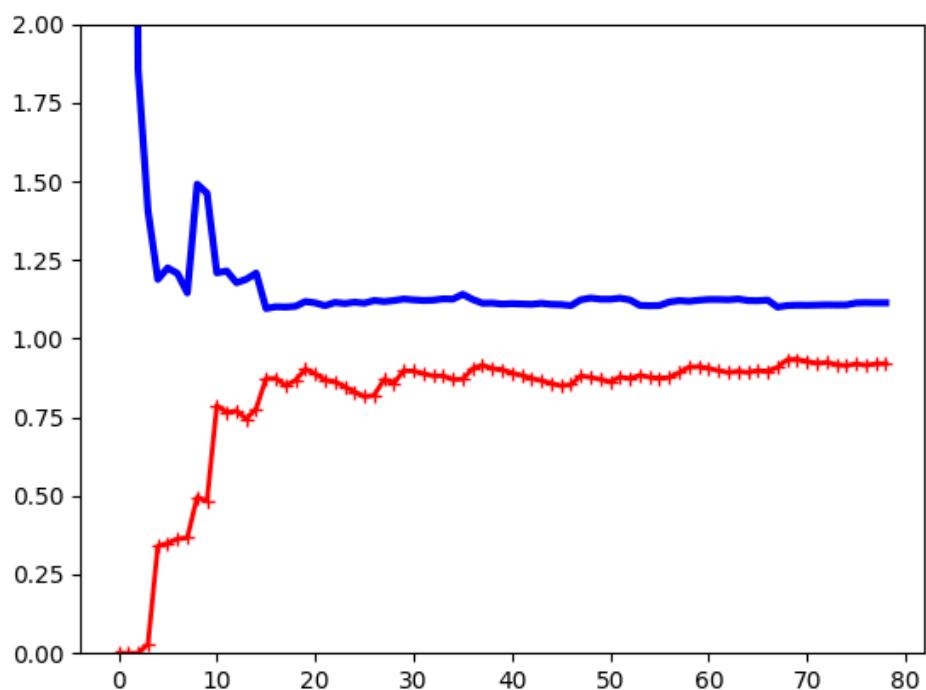
polynomial_regression = Pipeline([
    ('poly_features', PolynomialFeatures(degree=2, include_bias=False)),
    ('lin_reg', linear_model.LinearRegression()),
])

ivanytskyi_plot_learning_curves(polynomial_regression, np.array(X).reshape(-1, 1), y)

```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. ІО.				13
Змн.	Арк.	№ докум.	Підпис	Дата		

Результат виконання:



		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – ЛрЗ	Арк.
		Голенко М. Ю.				14
Змн.	Арк.	№ докум.	Підпис	Дата		

2. ЗАВДАННЯ НА ЛАБОРАТОРНУ РОБОТУ ТА МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО ЙОГО ВИКОНАННЯ

Завдання 2.7. Кластеризація даних за допомогою методу k-середніх

Провести кластеризацію даних методом k-середніх. Використовувати файл вхідних даних: data_clustering.txt.

Лістинг програми:

```
import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn import metrics


X = np.loadtxt('data_clustering.txt', delimiter=',')


num_clusters = 5


plt.figure()

plt.scatter(X[:, 0], X[:, 1], marker='o', facecolors='none', edgecolors='black',
s=80)

x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1

plt.title('Input data')

plt.xlim(x_min, x_max)

plt.ylim(y_min, y_max)

plt.xticks(())

plt.yticks(())

kmeans = KMeans(init='k-means++', n_clusters=num_clusters, n_init=10)

kmeans.fit(X)

step_size = 0.01

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_vals, y_vals = np.meshgrid(np.arange(x_min, x_max, step_size),
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		15

```

np.arange(y_min, y_max, step_size))

output = kmeans.predict(np.c_[x_vals.ravel(), y_vals.ravel()])

output = output.reshape(x_vals.shape)

plt.figure()

plt.clf()

plt.imshow(output, interpolation='nearest',

            extent=(x_vals.min(), x_vals.max(),

                    y_vals.min(), y_vals.max()),

            cmap=plt.cm.Paired,

            aspect='auto',

            origin='lower')

plt.scatter(X[:, 0], X[:, 1], marker='o', facecolors='none',

            edgecolors='black', s=80)

cluster_centers = kmeans.cluster_centers_

plt.scatter(cluster_centers[:, 0], cluster_centers[:, 1],

            marker='o', s=210, linewidths=4, color='black',

            zorder=12, facecolors='black')

x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1

y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1

plt.title('Межі кластерів')

plt.xlim(x_min, x_max)

plt.ylim(y_min, y_max)

plt.xticks(())

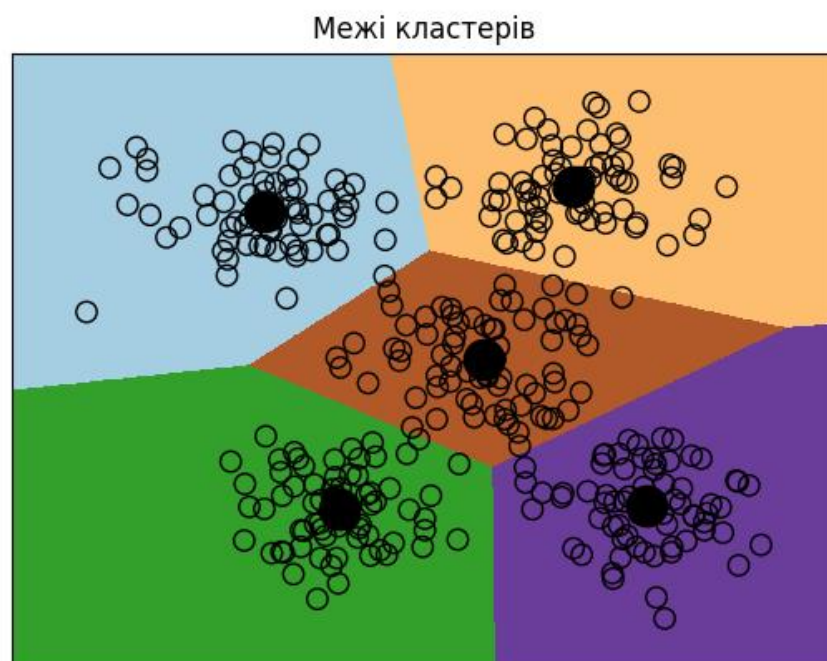
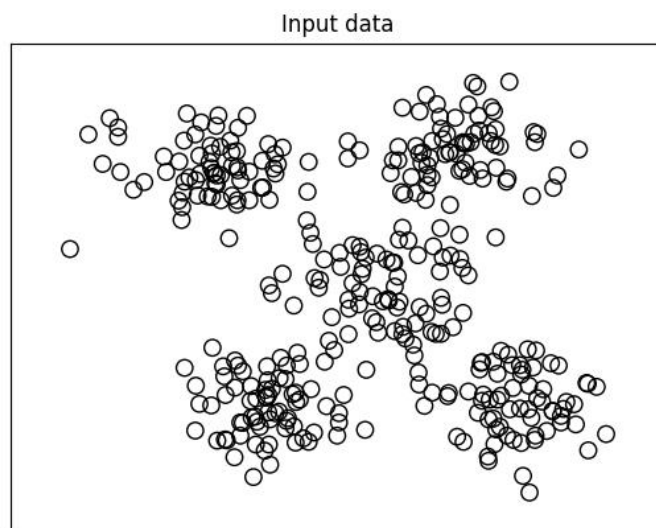
plt.yticks(())

plt.show()

```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		16

Результат виконання:



Висновок: Ми використали алгоритм K-Means для кластеризації даних, було використано 5 кластерів, на графіку показано розділення простору на кластери та центри кластерів.

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				17
Змн.	Арк.	№ докум.	Підпис	Дата		

Завдання 2.8. Кластеризація К-середніх для набору даних Iris

Виконайте кластеризацію К-середніх для набору даних Iris, який включає три типи (класи) квітів ірису (Setosa, Versicolour і Virginica) з чотирма атрибутами: довжина чашолистка, ширина чашолистка, довжина пелюстки та ширина пелюстки. У цьому завданні використовуйте `sklearn.cluster.KMeans` для пошуку кластерів набору даних Iris.

Лістинг програми:

```
import matplotlib.pyplot as plt

from sklearn import datasets

from sklearn.cluster import KMeans

from sklearn.metrics import pairwise_distances_argmin

import numpy as np

iris = datasets.load_iris()

X = iris.data[:, :2]

Y = iris.target

kmeans = KMeans(n_clusters=Y.max() + 1, init='k-means++', n_init=10, max_iter=300,
                 tol=0.0001, verbose=0, random_state=None, copy_x=True)

kmeans.fit(X)

y_pred = kmeans.predict(X)

print("n_clusters: 3, n_init: 10, max_iter: 300, tol: 0.0001, verbose: 0, random_state:
None, copy_x: True")

print(y_pred)

plt.figure()

plt.scatter(X[:, 0], X[:, 1], c=y_pred, s=50, cmap='viridis')

centers = kmeans.cluster_centers_

plt.scatter(centers[:, 0], centers[:, 1], c='black', s=200, alpha=0.5)

plt.show()

def ivanytskyi_find_clusters(X, n_clusters, rseed=2):

    rng = np.random.RandomState(rseed)

    i = rng.permutation(X.shape[0])[:n_clusters]
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				18
Змн.	Арк.	№ докум.	Підпис	Дата		

```

centers = X[i]

while True:

    labels = pairwise_distances_argmin(X, centers)

    new_centers = np.array([X[labels == i].mean(0) for i in range(n_clusters)])

    if np.all(centers == new_centers):

        break

    centers = new_centers

return centers, labels

print("using find_clusters():")
centers, labels = ivanytskyi_find_clusters(X, 3)
print("n_clusters: 3, rseed: 2")
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.show()

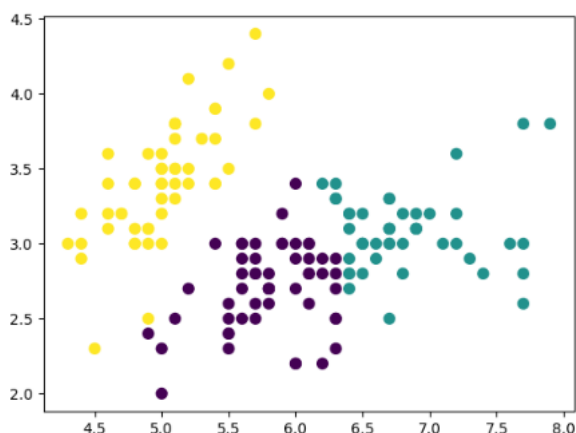
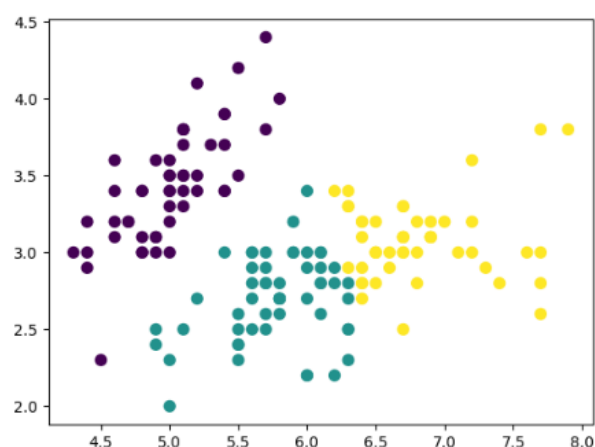
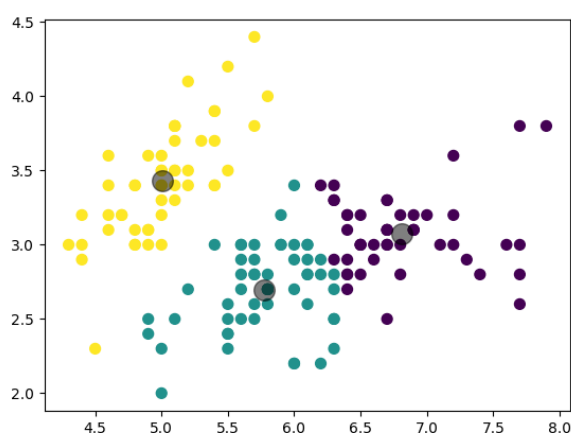
centers, labels = ivanytskyi_find_clusters(X, 3, rseed=0)
print("n_clusters: 3, rseed: 0")
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.show()

labels = KMeans(3, random_state=0).fit_predict(X)
print("n_clusters: 3, rseed: 0")
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.show()

```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				19
Змн.	Арк.	№ докум.	Підпис	Дата		

Результат виконання:

[illegible]

Висновок: K-Means алгоритм може бути використаний для кластеризації даних, і ви можете вибирати різні параметри та початкові умови для отримання різних результатів кластеризації. Точна кількість кластерів та початкові умови можуть бути важливими для правильного розділення даних на кластери

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				20
Змн.	Арк.	№ докум.	Підпис	Дата		

Завдання 2.9. Оцінка кількості кластерів з використанням методу зсуву середнього

Відповідно до рекомендацій, напишіть програму та оцініть максимальну кількість кластерів у заданому наборі даних за допомогою алгоритму зсуву середньою. Для аналізу використовуйте дані, які містяться у файлі data_clustering.txt.

Лістинг програми:

```
import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import MeanShift, estimate_bandwidth
from itertools import cycle

# Завантаження даних
X = np.loadtxt('data_clustering.txt', delimiter=',')

# Оцінка ширини вікна для X
bandwidth_X = estimate_bandwidth(X, quantile=0.1, n_samples=len(X))

# Кластеризація даних методом зсуву середнього
meanshift_model = MeanShift(bandwidth=bandwidth_X, bin_seeding=True)
meanshift_model.fit(X)

# Витягування центрів кластерів
cluster_centers = meanshift_model.cluster_centers_
print('\nCenters of clusters:\n', cluster_centers)

# Оцінка кількості кластерів
labels = meanshift_model.labels_
num_clusters = len(np.unique(labels))
print("\nNumber of clusters in input data =", num_clusters)

# Відображення на графіку точок та центрів кластерів
plt.figure()

markers = 'o*xvs'
```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – ЛрЗ	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		21

```

for i, marker in zip(range(num_clusters), markers):

    # Відображення на графіку точок, що належать поточному кластеру
    plt.scatter(X[labels == i, 0], X[labels == i, 1], marker=marker,
                color=np.random.rand(3,))

    # Відображення на графіку центру кластера
    cluster_center = cluster_centers[i]

    plt.plot(cluster_center[0], cluster_center[1], marker='o',
             markerfacecolor='black', markeredgecolor='red',
             markersize=15)

plt.title('Кластери')

plt.show()

```

Результат виконання:

```

import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['C:\\Users\\franc\\OneDrive\\Desktop\\labs0AI\\lab3', 'C:/Users/franc/OneDrive/Desktop/labs0AI/lab3'])

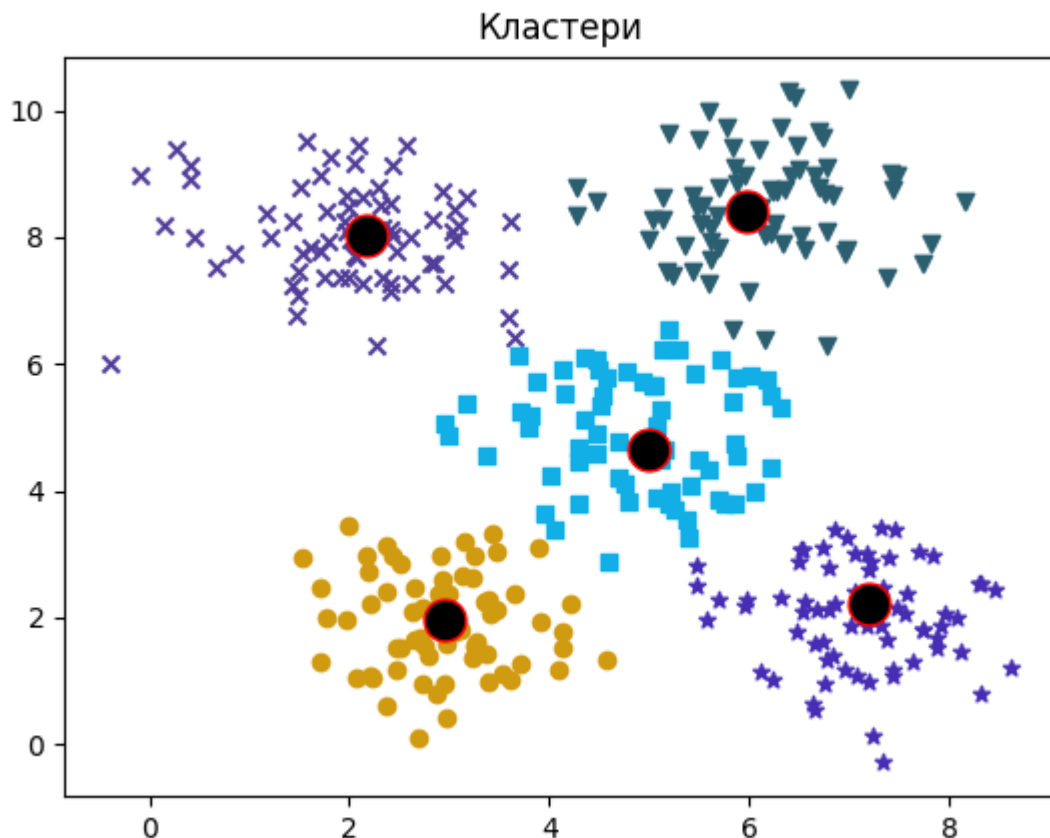
Python 3.10.1 (tags/v3.10.1:2cd268a, Dec 6 2021, 19:10:37) [MSC v.1929 64 bit (AMD64)]

centers of clusters:
[[2.95568966 1.95775862]
 [7.20690909 2.20836364]
 [2.17603774 8.03283019]
 [5.97960784 8.39078431]
 [4.99466667 4.65844444]]

Number of clusters in input data = 5

```

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				22
Змн.	Арк.	№ докум.	Підпис	Дата		



Висновок: Mean Shift алгоритм ефективно використовується для групування даних. В результаті його застосування було сформовано п'ять кластерів, а координати їх центрів були виведені на екран і зображені на графіку. Кожен кластер був відзначений окремим кольором для полегшення візуалізації та аналізу структури даних.

Висновок до лабораторної роботи: використовуючи спеціалізовані бібліотеки і мову програмування Python дослідив методи регресії та неконтрольованої класифікації даних у машинному навчанні.

		Іваницький Ф.А.			ДУ «Житомирська політехніка».23.121.8.000 – Лр3	Арк.
		Голенко М. Ю.				
Змн.	Арк.	№ докум.	Підпис	Дата		23