

ЛАБОРАТОРНА РОБОТА №3

ДОСЛІДЖЕННЯ МЕТОДІВ РЕГРЕСІЇ

Мета заняття: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити методи регресії даних у машинному навчанні.

Хід роботи

GitHub репозиторій: https://github.com/AlexanderHorielko/SAI_Horielko_PI-59

Завдання 2.1. Створення регресора однієї змінної

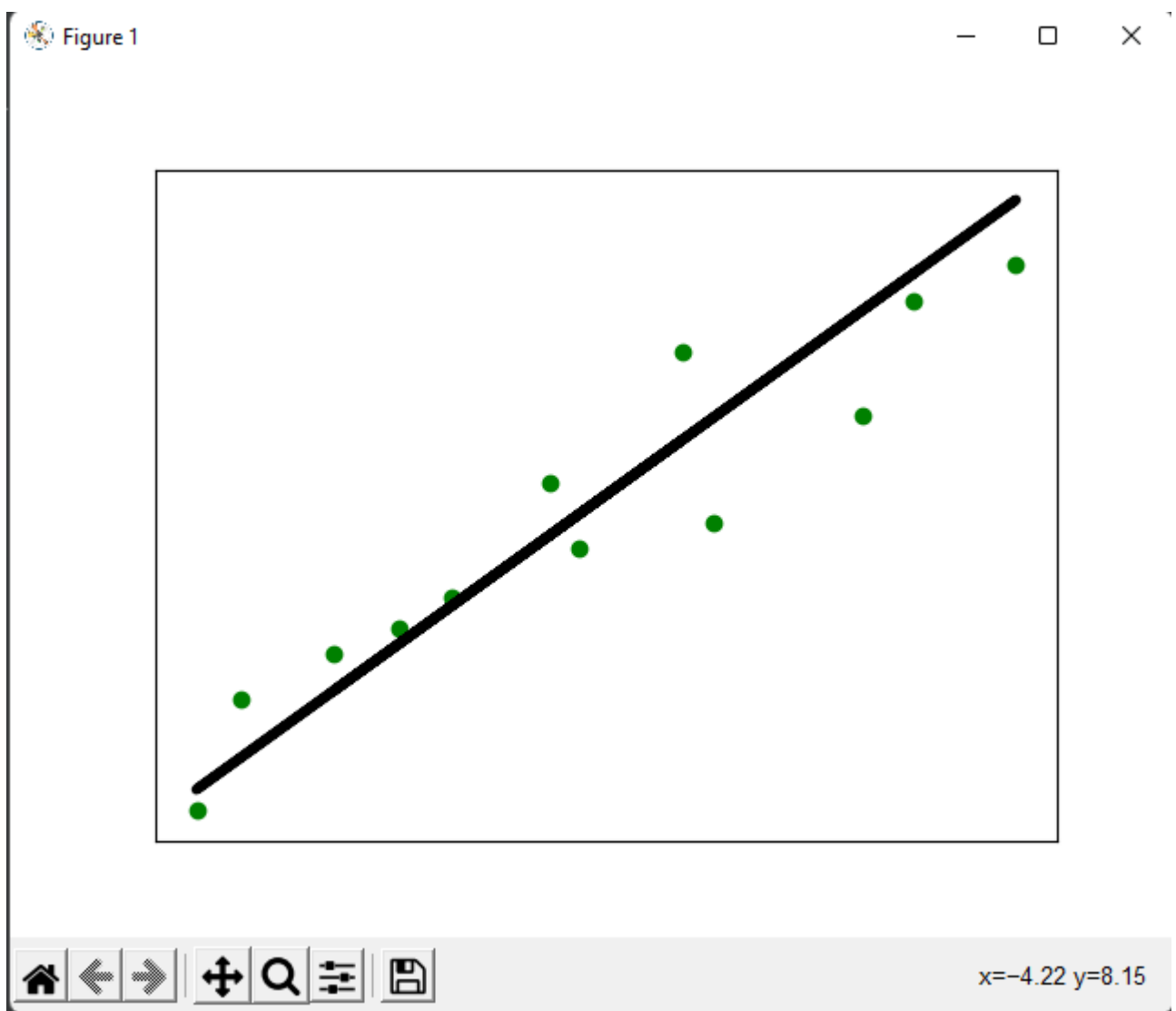


Рисунок 1. Графік функції

					ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3						
Змн.	Арк.	№ докум.	Підпис	Дата	Звіт з лабораторної роботи				Лім.	Арк.	Аркушів
Розроб.		Горелко О. В.									
Перевір.		Пулеко І. В.								1	3
Керівник									ФІКТ Гр. ПІ-59		
Н. контр.											
Зав. каф.											

```

7 input_file = 'data_singlevar_pregn.txt'
8 data = np.loadtxt(input_file, delimiter=',')
9 X, y = data[:, :-1], data[:, -1]
10 num_training = int(0.8 * len(X))
11 num_test = len(X) - num_training
12 X_train, y_train = X[:num_training], y[:num_training]
13 # Тестові дані
14 X_test, y_test = X[num_training:], y[num_training:]
15 regressor = linear_model.LinearRegression()
16 regressor.fit(X_train, y_train)
17 y_test_pred = regressor.predict(X_test)
18 plt.scatter(X_test, y_test, color='green')
19 plt.plot(X_test, y_test_pred, color='black', linewidth=4)
20 plt.xticks(())
21 plt.yticks(())
22 plt.show()
23 print("Linear regressor performance:")
24 print("Mean absolute error =",
25       round(sm.mean_absolute_error(y_test, y_test_pred), 2))
26 print("Mean squared error =",
27       round(sm.mean_squared_error(y_test, y_test_pred), 2))
28 print("Median absolute error =",
29       round(sm.median_absolute_error(y_test, y_test_pred), 2))
30 print("Explain variance score =",
31       round(sm.explained_variance_score(y_test, y_test_pred), 2))
32 print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))
33 output_model_file = 'model.pkl'
34 with open(output_model_file, 'wb') as f:
35     pickle.dump(regressor, f)
36 y_test_pred_new = regressor.predict(X_test)
37 print("\nNew mean absolute error =",
38       round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))
39

```

```

Run: task1 x
D:\University\4(2)\AI\lab3\Scripts\python.exe D:/University/4(2)/AI/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

```

Рисунок 2. Результати оцінки якості

Було обчислено метричні параметри регресора, порівнюючи справжні значення з передбаченими.

Завдання 2.2. Передбачення за допомогою регресії однієї змінної

№ за списком	1	2	3	4	5	6	7	8	9	10
№ варіанту	1	2	3	4	5	1	2	3	4	5

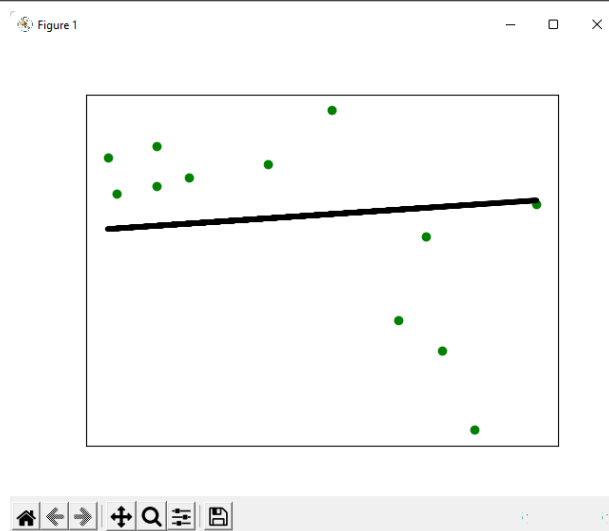


Рисунок 3. Графік функції

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пудеко І. В.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

```

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_5.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]
# Тестові дані
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 = 'model.pkl'
with open(output_model_file, 'wb') as f:
    pickle.dump(regressor, f)

```

```

task1 x
D:\University\4(2)\AI\lab3\Scripts\python.exe D:/University/4(2)/AI/lab3/task1.py
Linear regressor performance:
Mean absolute error = 3.31
Mean squared error = 16.98
Median absolute error = 2.66
Explain variance score = -0.14
R2 score = -0.15

New mean absolute error = 3.31

```

Рисунок 4. Результати оцінки якості

Було обчислено метричні параметри регресора, порівнюючи справжні значення з передбаченими за заданим варіантом.

Завдання 2.3. Створення багатовимірного регресора

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пулеко І. В.				3
Змн.	Арк.	№ докум.	Підпис	Дата		

```

7
8 input_file = 'data_multivar_regg.txt'
9 data = np.loadtxt(input_file, delimiter=',')
10 X, y = data[:, :-1], data[:, -1]
11 num_training = int(0.8 * len(X))
12 num_test = len(X) - num_training
13 X_train, y_train = X[:num_training], y[:num_training]
14 # Тестові дані
15 X_test, y_test = X[num_training:], y[num_training:]
16 regressor = linear_model.LinearRegression()
17 regressor.fit(X_train, y_train)
18 y_test_pred = regressor.predict(X_test)
19
20
21 print("Linear regressor performance:")
22 print("Mean absolute error =",
23       round(sm.mean_absolute_error(y_test, y_test_pred), 2))
24 print("Mean squared error =",
25       round(sm.mean_squared_error(y_test, y_test_pred), 2))
26 print("Median absolute error =",
27       round(sm.median_absolute_error(y_test, y_test_pred), 2))
28 print("Explain variance score =",
29       round(sm.explained_variance_score(y_test, y_test_pred), 2))
30 print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))
31 output_model_file = 'model.pkl'
32 with open(output_model_file, 'wb') as f:
33     pickle.dump(regressor, f)
34 y_test_pred_new = regressor.predict(X_test)
35 print("\nNew mean absolute error =",
36       round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))
37
38 polynomial = PolynomialFeatures(degree=10)
39 X_train_transformed = polynomial.fit_transform(X_train)
40
41 datapoint = [[7.75, 0.35, 5.50]]

```

Run: task3

```

D:\University\4(2)\AI\lab3\Scripts\python.exe D:/University/4(2)/AI/lab3/task3.py
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

New mean absolute error = 3.58

Linear regression:
[36.05286276]

Polynomial regression:
[41.46248888]

```

Рисунок 5. Результати оцінки якості

Можемо побачити, що поліноміальний регресор забезпечує отримання результату, ближчого до значення 41.46, тобто дає кращі результати.

Завдання 2.4. Регресія багатьох змінних

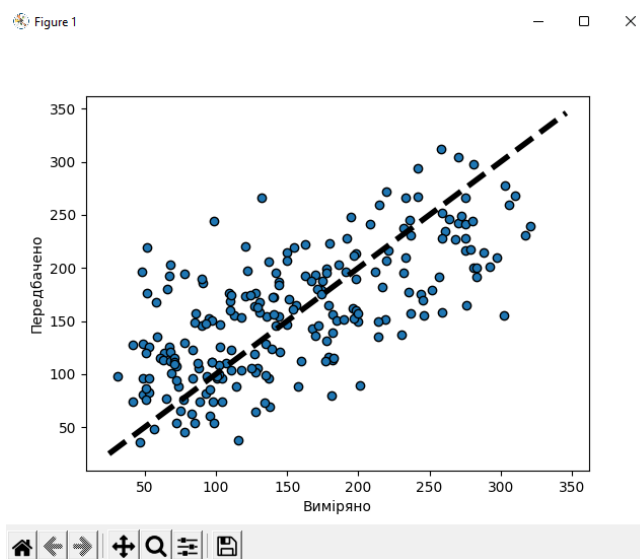


Рисунок 6. Графік функції

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пудеко І. В.				4
Змн.	Арк.	№ докум.	Підпис	Дата		

```

task3.py task4.py task1.py
1 import matplotlib.pyplot as plt
2 import numpy as np
3 from sklearn import datasets, linear_model
4 from sklearn.metrics import mean_squared_error, r2_score
5 from sklearn.metrics import mean_absolute_error
6 from sklearn.model_selection import train_test_split
7 import sklearn.metrics as sm
8
9 diabetes = datasets.load_diabetes()
10 X = diabetes.data
11 y = diabetes.target
12 Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=0.5, random_state=0)
13 regr = linear_model.LinearRegression()
14 regr.fit(Xtrain, ytrain)
15 ypred = regr.predict(Xtest)
16
17 print("Mean absolute error =",
18       round(sm.mean_absolute_error(ytest, ypred), 2))
19 print("Mean squared error =",
20       round(sm.mean_squared_error(ytest, ypred), 2))
21 print("Regression coef =",
22       regr.coef_, 2)
23 print("Regression intercept =",
24       round(regr.intercept_, 2))
25 print("R2 score =", round(sm.r2_score(ytest, ypred), 2))
26
27 fig, ax = plt.subplots()
28 ax.scatter(ytest, ypred, edgecolors=(0, 0, 0))
29 ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
30 ax.set_xlabel('Виміряно')
31 ax.set_ylabel('Передбачено')
32 plt.show()

```

Run: task4

D:\University\4(2)\AI\lab3\Scripts\python.exe D:/University/4(2)/AI/lab3/task4.py

Mean absolute error = 44.8

Mean squared error = 3075.33

Regression coef = [-20.4047621 -265.88518066 564.65086437 325.56226865 -692.16120333 395.55720874 23.49659361 116.36402337 843.94613929 12.71856131] 2

Regression intercept = 154.36

R2 score = 0.44

Рисунок 7. Результати оцінки якості

Створено модель регресії та отримано дані про неї

Висновок: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідив методи регресії даних у машинному навчанні.

Завдання 2.5. Самостійна побудова регресії

Варіант 5

```

m = 100
X = 6 * np.random.rand(m, 1) - 3
y = 0.4 * X ** 2 + X + 4 + np.random.randn(m, 1)

```

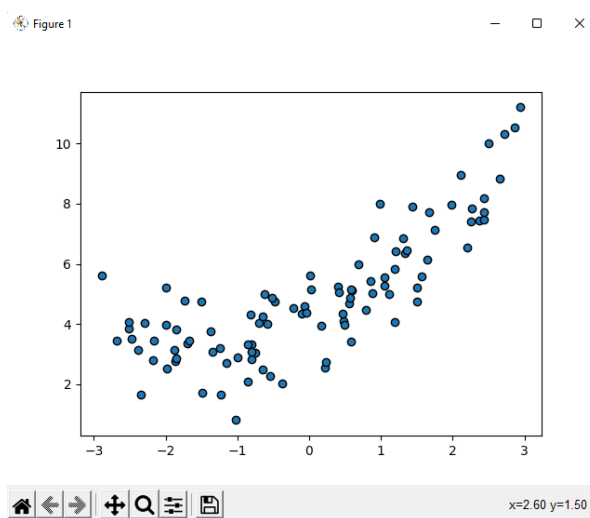


Рисунок 8. Графік функції

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пудеко І. В.				5
Змн.	Арк.	№ докум.	Підпис	Дата		

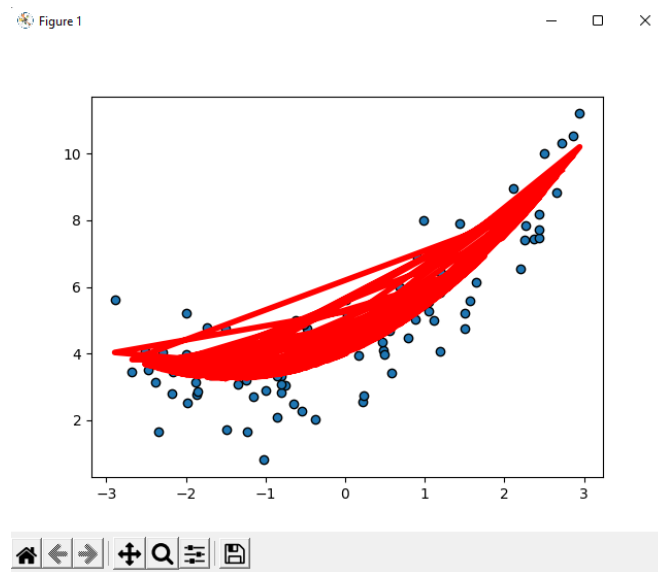


Рисунок 9. Графік функції і регресія

```
task5 x
D:\University\4(2)\AI\lab3\Scripts\python.exe D:/University/4(2)/AI/lab3/task5.py
[3.99400982] [[0.93924116 0.41704405]]
```

Рисунок 10. Результати регресії

Дана модель математичного рівняння: $y = 0.5x^2 + 1x + 2$ + гаусовий шум
 Отримана модель регресії: $y = 0.41x^2 + 0.93x + 0.41$.
 Коефіцієнт близький до модельних, отже модель навчена правильно.

Завдання 2.6. Побудова кривих навчання

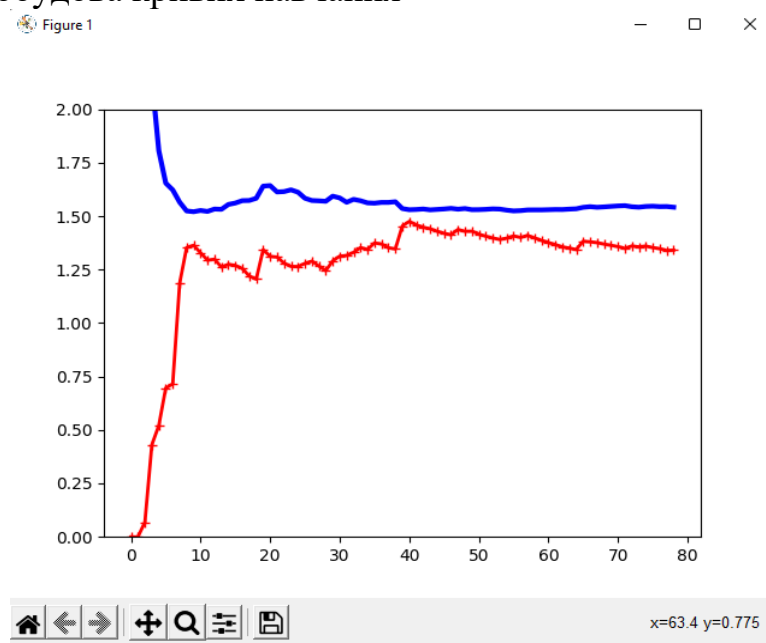


Рисунок 11. Криві навчання для лінійної моделі

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пудеко І. В.				6
Змн.	Арк.	№ докум.	Підпис	Дата		

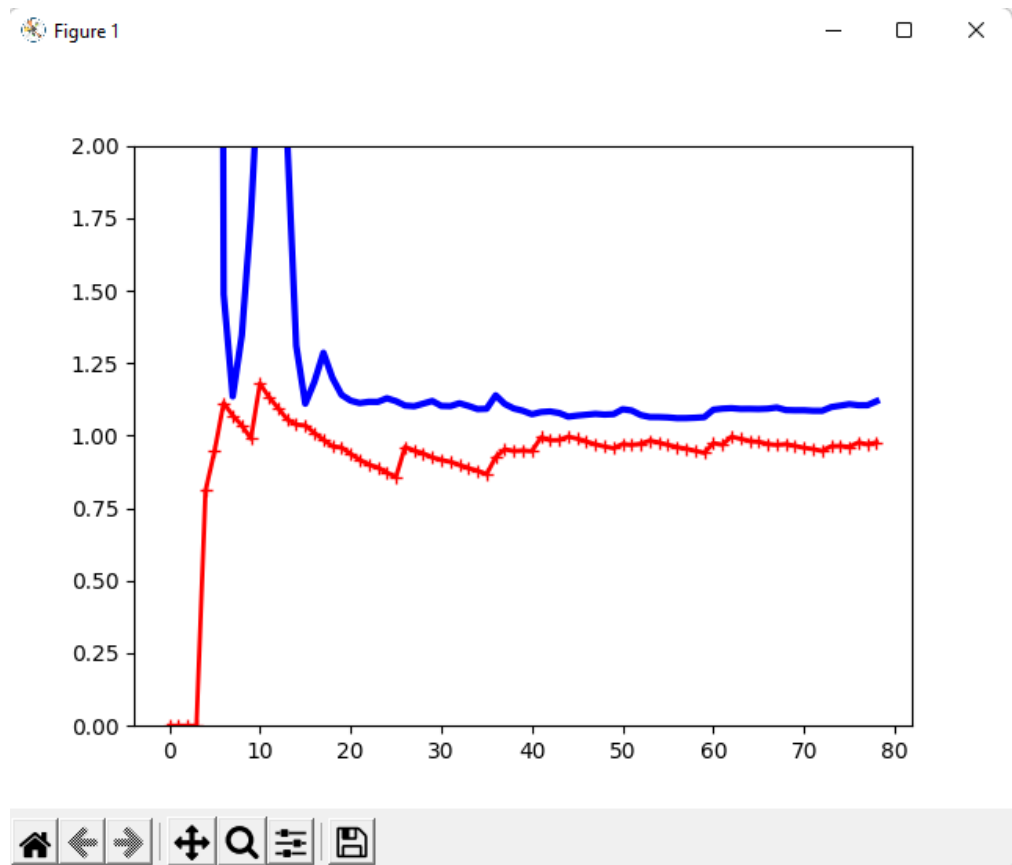


Рисунок 11. Криві навчання для поліноміальної моделі

Висновок: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідив методи регресії даних у машинному навчанні.

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр3	Арк.
		Пудеко І. В.				7
Змн.	Арк.	№ докум.	Підпис	Дата		