

ЛАБОРАТОРНА РОБОТА № 3

ДОСЛІДЖЕННЯ МЕТОДІВ РЕГРЕСІЇ

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

Хід роботи:

Завдання 1: Створення регресора однієї змінної

Лістинг програми:

```
# Diachenko Viktor, ZPI-18, Lab3, Task 1

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]

# Розбивка даних на навчальний та тестовий набори
num_training = int(0.8 * len(X))
num_test = len(X) - num_training

# Тренувальні дані
X_train, y_train = X[:num_training], y[:num_training]
# Тестові дані
X_test, y_test = X[num_training:], y[num_training:]

# Створення об'єкта лінійного регресора
regressor = linear_model.LinearRegression()

regressor.fit(X_train, y_train)
```

					ДУ «Житомирська політехніка».20.121.03.000 – Лр3				
Змн.	Арк.	№ докум.	Підпис	Дата	Звіт з лабораторної роботи	Літ.	Арк.	Аркушів	
Розроб.		Дяченко В. В.							
Перевір.		Пулеко І. В.					1	08	
Керівник						ФІКТ Гр. ЗПІ-18			
Н. контр.									
Зав. каф.									

```

# Прогнозування результату
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()

3
regressor.fit(X_train, y_train)

# Прогнозування результату
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))
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'

# Збереження моделі
with open(output_model_file, 'wb') as f:
    pickle.dump(regressor, f)

# Завантаження моделі
y_test_pred_new = regressor.predict(X_test)
print("\nNew mean absolute error =",
round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))

```

		Дяченко В. В.			ДУ «Житомирська політехніка».20.121.03.000 – Лр3	Арк.
		Пулюко І. В.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

Результат виконання програми:

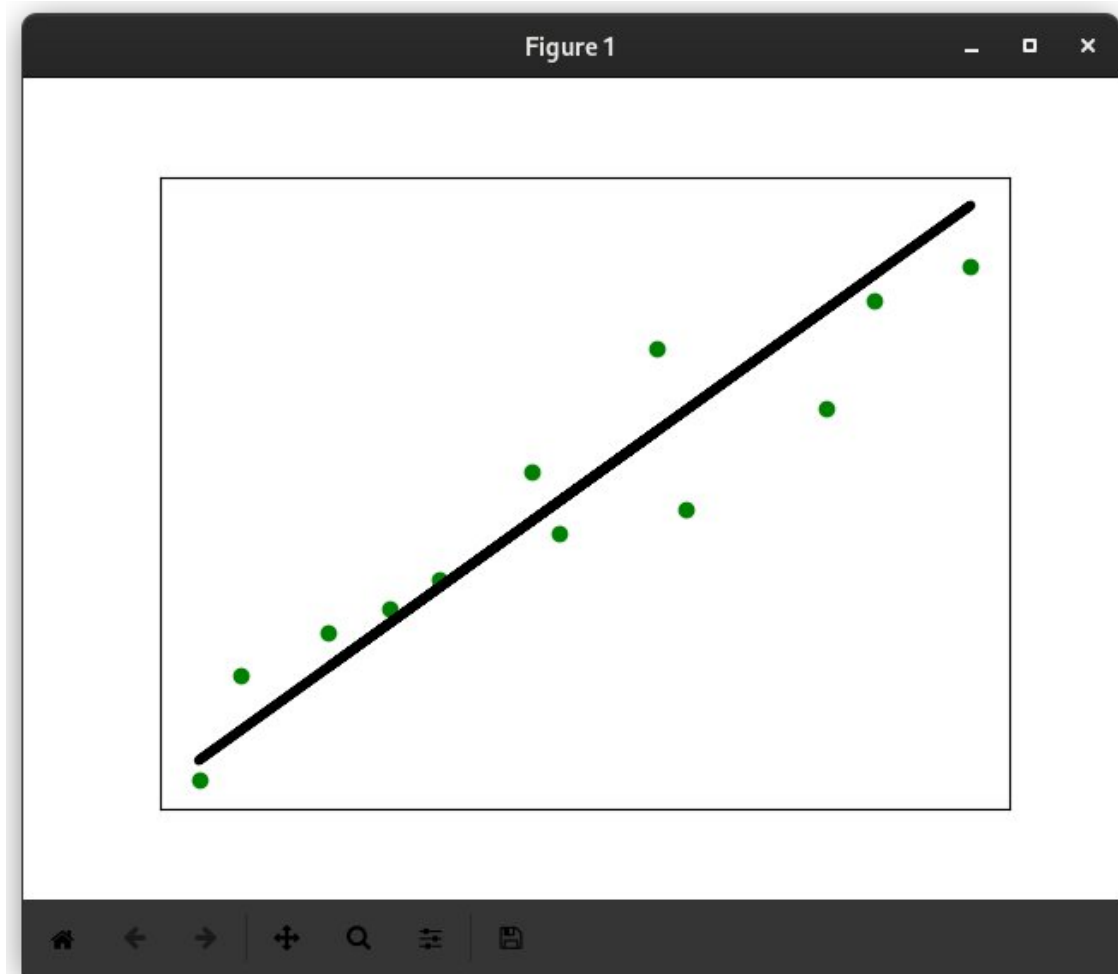


Рис. 1. Регресивна модель на основі однієї змінної

```

albedych@MiWiFi-R4A-srv:~/Документи/AI/13
~/Д/АІ/13 python LR_3_task_1.py 12:08:41
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
~/Д/АІ/13 1m 43s 12:13:28
  
```

Рис. 2. Результати оцінки якості

Завдання 2. Передбачення за допомогою регресії однієї змінної

Лістинг програми:

```
# Diachenko Viktor, ZPI-18, Lab3, Task 2

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_regr_1.txt'

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

# Розбивка даних на навчальний та тестовий набори
num_training = int(0.8 * len(X))
num_test = len(X) - num_training

# Тренувальні дані
X_train, y_train = X[:num_training], y[:num_training]
# Тестові дані
X_test, y_test = X[num_training:], y[num_training:]

# Створення об'єкта лінійного регресора
regressor = linear_model.LinearRegression()

regressor.fit(X_train, y_train)

# Прогнозування результату
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()

3
regressor.fit(X_train, y_train)

# Прогнозування результату
y_test_pred = regressor.predict(X_test)

# Побудова графіка
plt.scatter(X_test, y_test, color='green')
```

		Дяченко В. В.			ДУ «Житомирська політехніка».20.121.03.000 – Лр3	Арк.
		Пулеко І. В.				4
Змн.	Арк.	№ докум.	Підпис	Дата		

```

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))

# Файл для збереження моделі
output_model_file = 'model.pkl'

# Збереження моделі
with open(output_model_file, 'wb') as f:
    pickle.dump(regressor, f)

# Завантаження моделі
y_test_pred_new = regressor.predict(X_test)
print("\nNew mean absolute error =",
round(sm.mean_absolute_error(y_test, y_test_pred_new), 2))

```

Результат виконання програми:

		Дяченко В. В.			ДУ «Житомирська політехніка».20.121.03.000 – Лр3	Арк.
		Пулеко І. В.				
Змн.	Арк.	№ докум.	Підпис	Дата		5

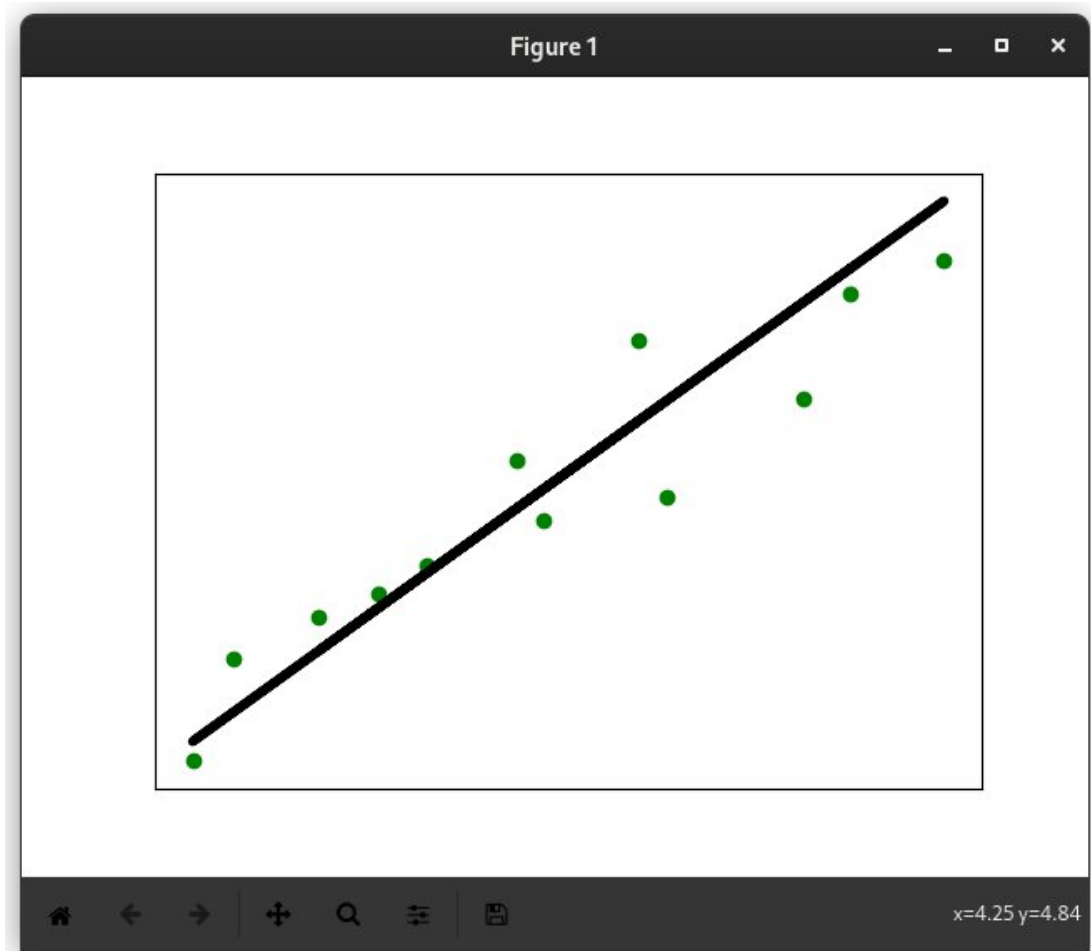


Рис. 3. Регресивна модель на основі однієї змінної

```

albedych@MiWiFi-R4A-srv:~/Документи/AI/L3
~/Д/АІ/Л3 python LR_3_task_2.py 12:24:32
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
~/Д/АІ/Л3 3s 12:24:37

```

Рис. 4. Оцінка результату якості

Завдання 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
from sklearn.metrics import mean_absolute_error
from sklearn.model_selection import train_test_split

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

Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size = 0.5, random_state =
0)

regr = linear_model.LinearRegression()

regr.fit(Xtrain, ytrain)

ypred = regr.predict(Xtest)

fig, ax = plt.subplots()
ax.scatter(ytest, ypred, 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 програми:

		Дяченко В. В.			ДУ «Житомирська політехніка».20.121.03.000 – Лр3	Арк.
		Пулеко І. В.				7
Змн.	Арк.	№ докум.	Підпис	Дата		

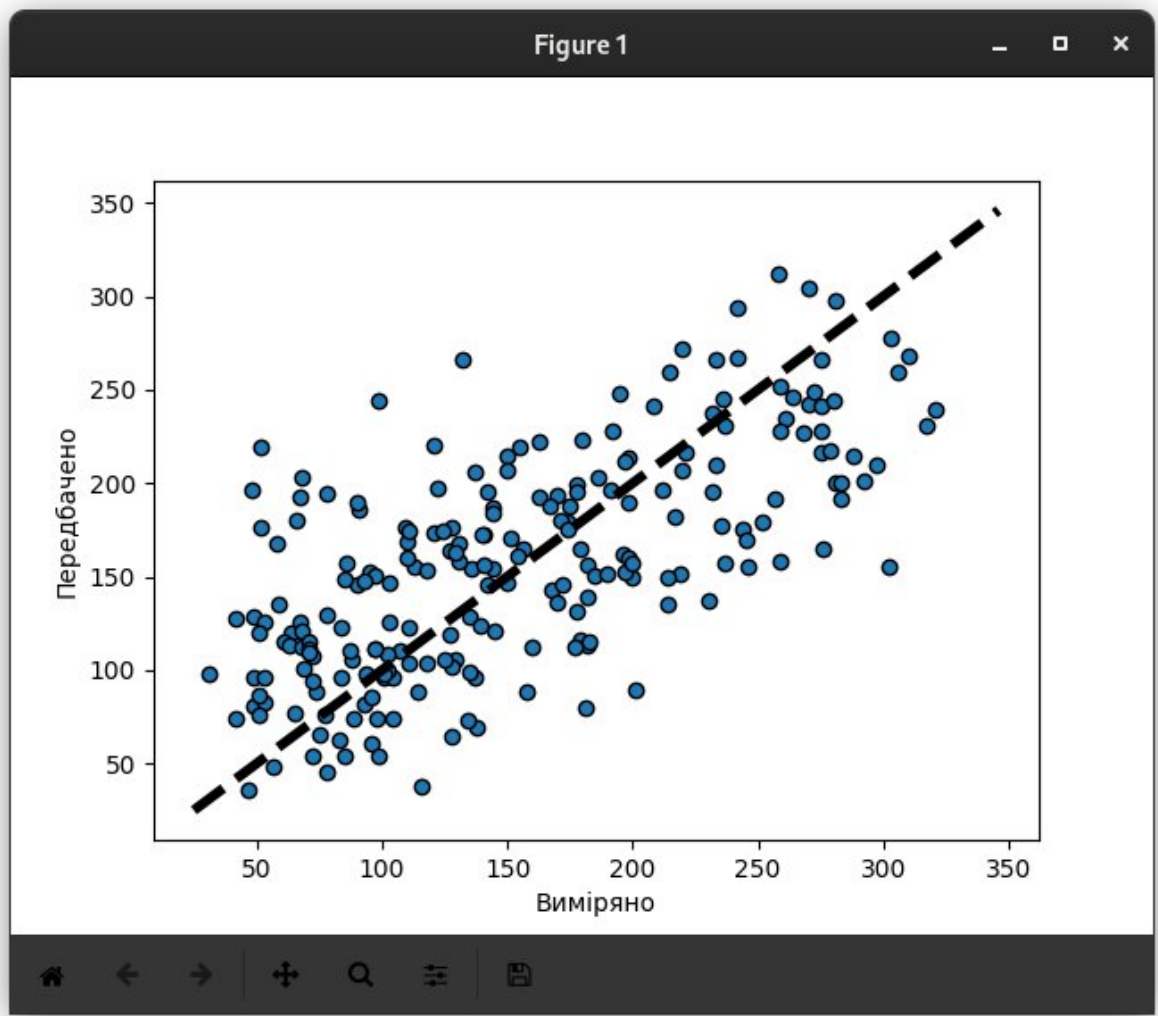


Рис. 5. Регресивна модель на основі багатьох змінних

Посилання на GitHub: https://github.com/diachenkovv/AI_python

Висновки: в ході виконання лабораторної роботи, використовуючи спеціалізовані бібліотеки та мову програмування Python, було досліджено методи регресії даних у машинному навчанні.

		Дяченко В. В.			ДУ «Житомирська політехніка». 20.121.03.000 – Лр3	Арк.
		Пулеко І. В.				8
Змн.	Арк.	№ докум.	Підпис	Дата		