ЛАБОРАТОРНА РОБОТА №3

ДОСЛІДЖЕННЯ МЕТОДІВ РЕГРЕСІЇ ТА НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

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

Завдання 2.1. Створення регресора однієї змінної.

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

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 perpecopa
regressor = linear_model.LinearRegression()

# Тренування моделі
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.tyticks(())
plt.tyticks(())
```

					ДУ «Житомирська політехі	нік <mark>а</mark> ».22	2.121.04	.000 — ЛрЗ	
Змн.	Арк.	№ докум.	Підпис	Дата	•			<u>, </u>	
Розр	00 δ.	Войтко П.О.	·			Літ.	Арк.	Аркушів	
Пере	Перевір.	. Філіпов В.О.		n_ :		1	12		
Кері	зник				Звіт з				
Н. кс	нтр.				лабораторної роботи	ΦΙΚΤ	Гр. ІПЗ	Bκ-20-1[1]	
Зав.	каф.						•		

```
print("Explain variance score =", round(sm.explained variance score(y test,
print("R2 score =", round(sm.r2_score(y_test, y_test_pred), 2))
output model file = 'model.pkl'
y_test_pred_new = regressor_model.predict(X_test)
               K Figure 1
```

Рис. 1. - Результат виконання

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

☆ 🎤 🗣 Q 至 🖺

x=-0.27 y=6.17

```
Run:

LR_3_task_1 ×

C:\Users\Admin\PycharmProjects\labka3\venv\Scripts\python.exe C:\Users\Admin\PycharmProjects\labka3\LR_3_task_1.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

Process finished with exit code 0
```

Рис. 2. - Результат виконання

Висновок: модель для вихідних даних побудована валідно. МАЕ, MSE – середня якість. Показник R2 – добре.

Завдання 2.2. Передбачення за допомогою регресії однієї змінної.

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

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

# Вхідний файд, який містить дані
input_file = 'data_regr_4.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'скта лінійного perpecopa
regressor = linear_model.LinearRegression()
# Тренування моделі
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.plotty test, y test pred, color='black', linewidth=4)
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

```
plt.xticks(())
plt.yticks(())
plt.yticks(())
plt.show()

# Ображування метрик
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:
    plckle.dump(regressor, f)

# Завантаження моделі
with open(output_model_file, 'rb') as f:
    regressor_model = pickle.load(f)

# Регform prediction on test data
y_test_pred_new = regressor_model.predict(X_test)
print("New mean absolute error =", round(sm.mean_absolute_error(y_test,
y_test_pred_new), 2))
```



Рис. 3. - Результат виконання

Арк.

			Войтко П.О.			
			Філіпов В.О.			ДУ «Житомирська політехніка».22.121.04.000 — Лр3
ı	Змн.	Арк.	№ докум.	Підпис	Дата	

```
LR_3_task_2
   C:\Users\Admin\PycharmProjects\labka3\venv\Scripts\python.exe C:\Users\Admin\PycharmProjects\labka3\LR_3_task_2.py
   Linear regressor performance:
   Mean absolute error = 2.72
mean squared error = 13.16
≛ Median absolute error = 1.9
₹ Explain variance score = -0.07
New mean absolute error = 2.72
   Process finished with exit code 0
```

Рис. 4. - Результат виконання

Завдання 2.3. Створення багатовимірного регресора.

```
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)
print("Explained variance score =", round(sm.explained_variance_score(y_test,
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

```
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))
 C:\Users\Admin\PycharmProjects\labka3\venv\Scripts\python.exe C:\Users\Admin\PycharmProjects\labka3\LR_3_task_3.py
 Linear Regressor performance:
 Mean squared error = 20.31
 Explained variance score = 0.86
 R2 score = 0.86
 Linear regression:
  [36.05286276]
 Polynomial regression:
  [41.45562492]
 Process finished with exit code 0
```

Рис. 5. - Результат виконання

Висновок: Якщо порівнювати з з лінійним регресором, поліноміальний регресор демонструє кращі результати. На це вказує значення 41.45

Завдання 2.4. Регресія багатьох змінних.

```
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, ran-dom_state=0)
regr = linear_model.LinearRegression()
regr.fit(Xtrain, ytrain)
ypred = regr.predict(Xtest)
# OбpaxyBaHHH Metpuk
print("regr.coef =", np.round(regr.coef_, 2))
print("R2 score =", round(r2_score(ytest, ypred), 2))
print("Mean absolute error =", round(mean_absolute_error(ytest, ypred), 2))
print("Mean squared error =", round(mean_squared_error(ytest, ypred), 2))
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('BumipsHo')
ax.set_ylabel('ПередбачеНо')
plt.show()
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

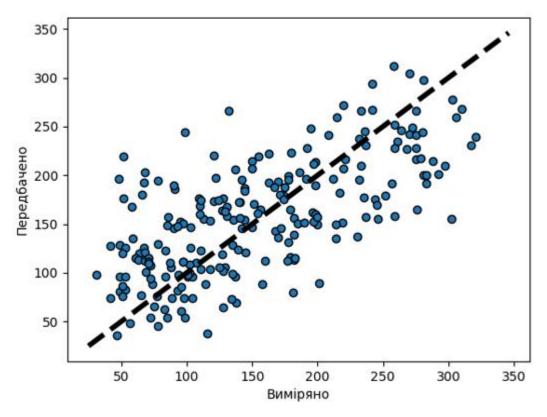


Рис. 6. - Результат виконання

```
regr.coef = [ -20.4 -265.89 564.65 325.56 -692.16 395.56 23.5 116.36 843.95 12.72]
regr.intercept = 154.36
R2 score = 0.44
Mean absolute error = 44.8
Mean squared error = 3075.33

Process finished with exit code 0
```

Рис. 7. - Результат виконання

Завдання 2.5. Самостійна побудова регресії.

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

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

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

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

```
# Лінійна perpeciя
linear_regressor = linear_model.LinearRegression()
linear_regressor.fit(X, y)

# Поліноміальна perpeciя
polynomial = PolynomialFeatures(degree=2, include_bias=False)
X poly = polynomial.fit_transform(X)
polynomial.fit(X_poly, y)

poly_linear_model = linear_model.LinearRegression()
poly_linear_model.fit(X_poly, y)
y_pred = poly_linear_model.predict(X_poly)

print("\nr2: ", sm.r2_score(y, y_pred))

# Лінійна perpeciя
plt.scatter(X, y, color='red')
plt.plot(X, linear_regressor.predict(X), color='blue', linewidth=1)
plt.show()

# Поліноміальна perpeciя
plt.scatter(X, y, color='red')
plt.plot(X, y_pred, "+", color='blue', linewidth=2)
plt.title("Поліноміальна perpeciя")
plt.show()
```

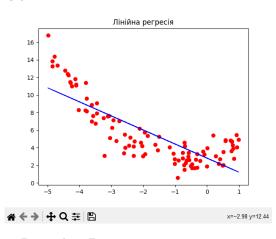


Рис. 8. - Результат виконання

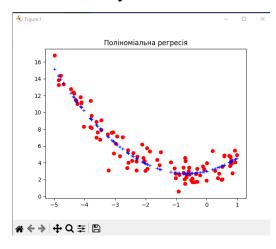


Рис. 9. - Результат виконання

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

Завдання 2.6. Побудова кривих навчання.

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

```
import matplotlib.pyplot as plt
m = 100
X = 6 * np.random.rand(m, 1) - 5
lin reg = linear model.LinearRegression()
from sklearn.pipeline import Pipeline
polynomial_regression = Pipeline([
plot learning curves(polynomial regression, X, y)
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

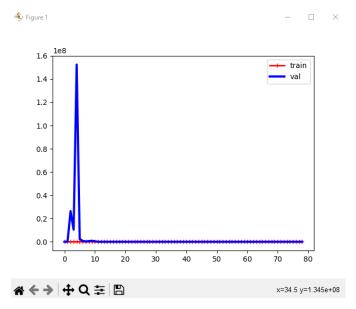


Рис. 10. Криві навчання для поліноміальної моделі 10 ступеня.

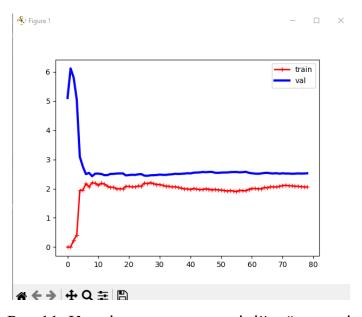


Рис.11. Криві навчання для лінійної моделі.

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

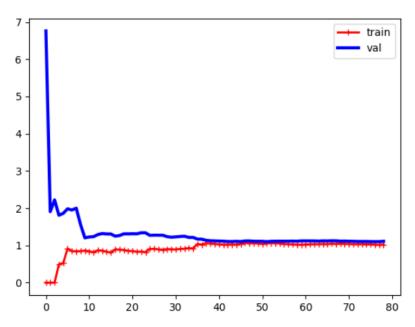


Рис.12. Криві навчання для поліноміальної моделі 2 ступеня.

Висновок: для з'ясування ступеня складності необхідної моделі використовуються криві навчання. Для досягнення успіху необхідно досягти компромісу між зміщенням та дисперсією. В нашому випадку найкращий результат показала модель 2 ступеня.

Завдання 2.7. Кластеризація даних за допомогою методу к-середніх.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn import metrics

# Завантаження вхідних даних
X = np.loadtxt('data_clustering.txt', delimiter=',')
num_clusters = 5

# Включення вхідних даних до графіка
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.ylim(y_min, y_max)
plt.ylim(y_min, y_max)
plt.yticks(())

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

# Навчання моделі кластеризації KMeans
kmeans.fit(X)

# Визначення кроку сітки
step_size = 0.01
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

```
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
# Передбачення вихідних міток для всіх точок сітки
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',
            cmap=plt.cm.Paired,
plt.scatter(X[:, 0], X[:, 1], marker='o', facecolors='none',
cluster_centers = kmeans.cluster_centers_
plt.scatter(cluster_centers[:, 0], cluster_centers[:, 1],
x \min, x \max = X[:, 0].\min() - 1, X[:, 0].\max() + 1
plt.title('Межі кластерів')
plt.xlim(x min, x max)
plt.ylim(y min, y max)
plt.xticks(())
plt.yticks(())
plt.show()
                                         □ × ﴿ Figure 2
                          Input data
                                                            Межі кластерів
          ☆ ♦ ♦ ♦ ♀ ₽
                                             ~ ~ → | + Q = | B
                                  plt.xlim(x_min, x_max)
plt.ylim(y_min, y_max)
```

Рис. 13. Вхідні дані + кластери.

Rum: | R.J. Lask.5 × | R.J. Lask.7 × | R.J. Lask.7 × | R.J. Lask.5 × | R.J. La

		Войтко П.О.		
		Філіпов В.О.	·	
Змн.	Арк.	№ докум.	Підпис	Дата

Висновок: метод k-середніх валідно працює, але за умови, відомої кількісті кластерів.

Завдання 2.8. Кластеризація K-середніх для набору даних Iris.

```
import matplotlib.pyplot as plt
import numpy as np
iris = datasets.load iris()
X = iris.data[:, :2]
# Визначаємо початкові кластери
kmeans.fit(X)
print("n_clusters: 3, n_init: 10, max_iter: 300, tol: 0.0001, verbose: 0, ran-
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()
   i = rng.permutation(X.shape[0])[:n clusters]
print("using find clusters():")
centers, labels = 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 = 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()
```

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

Рис. 14. Кластеризація для набору даних Iris.

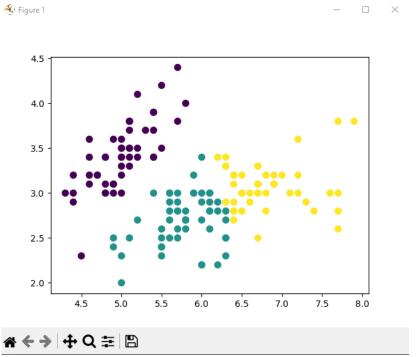
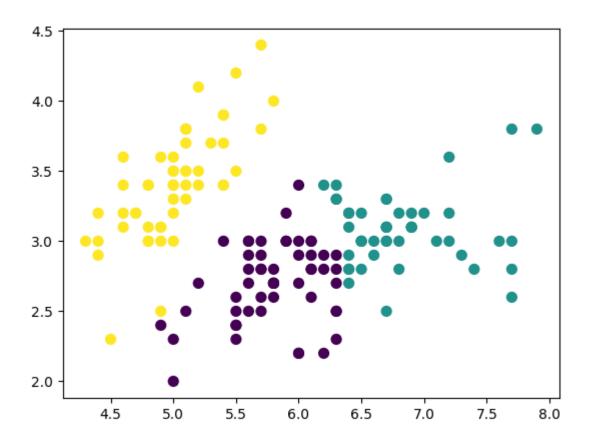


Рис. 15. Кластеризація для набору даних Iris.

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата





```
C:\Users\Admin\PycharmProjects\labka3\venv\Scripts\python.exe C:\Users\Admin\PycharmProjects\labka3\LR_3_task_8.py
n_clusters: 3, n_init: 10, max_iter: 300, tol: 0.0001, verbose: 0, ran-dom_state: None, copy_x: True
110011110101010111111100111011101
using find_clusters():
```

Рис. 16. Кластеризація для набору даних Iris.

Завдання 2.9. Оцінка кількості кластерів з використанням методу зсуву середнього.

```
bandwidth_X = estimate_bandwidth(X, quantile=0.1,
```

		Войтко П.О.		
·	·	Філіпов В.О.	·	·
Змн.	Арк.	№ докум.	Підпис	Дата

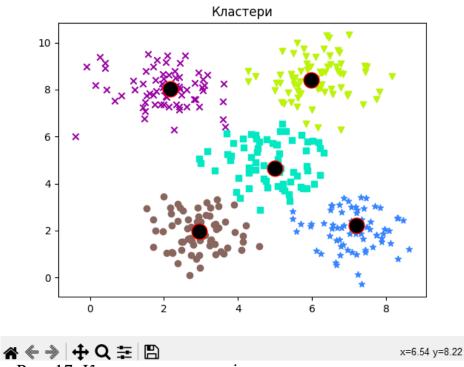


Рис. 17. Кластери, отримані методом зсуву середнього.

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата

```
C:\Users\Admin\PycharmProjects\labka3\venv\Scripts\python.exe C:\Users\Admin\PycharmProjects\labka3\LR_3_task_9.py

Centers of clusters:
   [[2.95568966 1.95775862]
   [7.20690909 2.20836364]
   [2.17603774 8.03283019]
   [5.97960784 8.39078431]
   [4.99466667 4.658444444]]

Number of clusters in input data = 5
```

Рис. 18. Центри кластерів.

Метод зсуву середнього — доволі валідний алгоритм, головною перевагою якого ϵ непотрібність жодних припущень щодо базового розподілу даних, має змогу обробляти довільні простори функцій, проте важливу роль відіграє обрана ширина вікна (bandwidth).

Висновок по лабораторній роботі:

Під час виконання завдань лабораторної роботи я навчився працювати з використанням спеціалізованих бібліотек та мови програмування Python було досліджено методи регресії та неконтрольованої класифікації даних у машинному навчанні.

Github: https://github.com/pvoitko/II

		Войтко П.О.		
		Філіпов В.О.		
Змн.	Арк.	№ докум.	Підпис	Дата