Жарқын Ақтілек

7 Есеп Үлкен деректер жиынтығының ішінде шағын ауытқулардың жалпы модель нәтижелеріне қалай әсер ететінін зеттеңіз. мұндай әсерлердің тұрақтылығы мен өзгеруі

import numpy as np

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

# Деректер жиынтығын жасау

np.random.seed(42)

X = 10 \* np.random.rand(1000, 1) # 1000 мән

y = 5 \* X + 3 + np.random.randn(1000, 1) \* 2 # Нақты мәндер

# Модельді үйрету

model = LinearRegression()

model.fit(X, y)

y\_pred = model.predict(X)

original\_mse = mean\_squared\_error(y, y\_pred)

# Шағын ауытқулар енгізу

X\_noisy = X.copy()

X\_noisy[:50] += np.random.randn(50, 1) \* 5 # 50 мәнге үлкенірек ауытқу енгізу

y\_noisy = y.copy()

y\_noisy[:50] += np.random.randn(50, 1) \* 5

# Жаңа модельді үйрету

model\_noisy = LinearRegression()

model\_noisy.fit(X\_noisy, y\_noisy)

y\_noisy\_pred = model\_noisy.predict(X\_noisy)

noisy\_mse = mean\_squared\_error(y\_noisy, y\_noisy\_pred)

# Нәтижелерді салыстыру

print(f"Түпнұсқа деректер бойынша MSE: {original\_mse:.4f}")

print(f"Ауытқыған деректер бойынша MSE: {noisy\_mse:.4f}")

# График арқылы көрсету

plt.figure(figsize=(10, 5))

plt.scatter(X, y, color='blue', alpha=0.3, label='Нақты деректер')

plt.scatter(X\_noisy[:50], y\_noisy[:50], color='red', alpha=0.6, label='Ауытқыған деректер')

plt.plot(X, y\_pred, color='green', label='Түпнұсқа модель')

plt.plot(X, model\_noisy.predict(X), color='purple', linestyle='dashed', label='Ауытқыған модель')

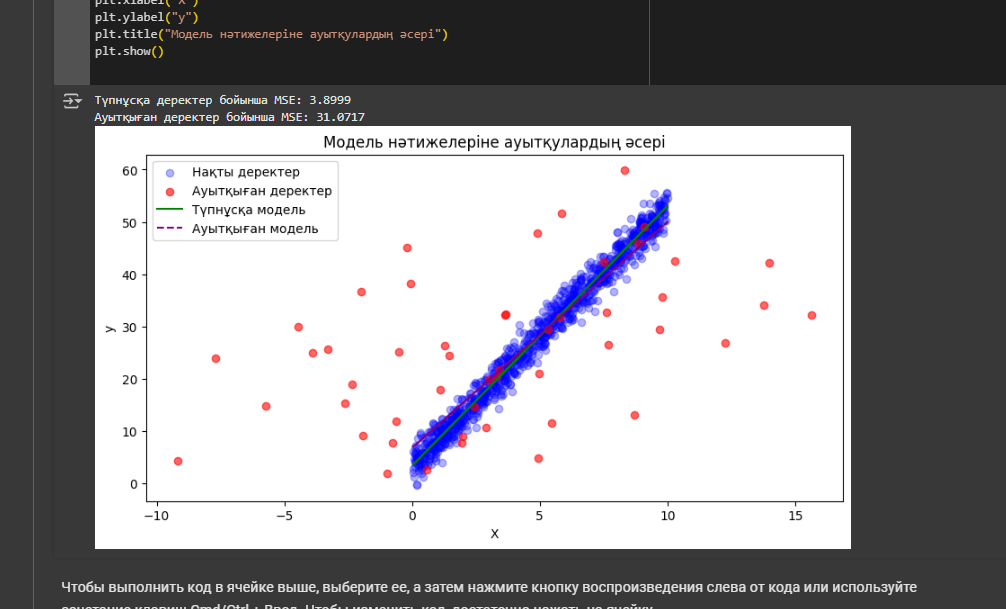
plt.legend()

plt.xlabel("X")

plt.ylabel("y")

plt.title("Модель нәтижелеріне ауытқулардың әсері")

plt.show()



8 Есеп Гаусс үлестірімінен ауытқулары бар деретер жиынтығының модельдің дәлдігін тексер

import numpy as np

import matplotlib.pyplot as plt

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

# Деректерді генерациялау

np.random.seed(42)

X = np.random.normal(loc=50, scale=10, size=(100, 1)) # Гаусс үлестірімі

Y = 3 \* X + 7 + np.random.normal(loc=0, scale=5, size=(100, 1)) # Негізгі заңдылық

# Ауытқулар қосу (outliers)

outliers\_X = np.random.uniform(20, 80, size=(10, 1))

outliers\_Y = np.random.uniform(100, 200, size=(10, 1))

X\_with\_outliers = np.vstack((X, outliers\_X))

Y\_with\_outliers = np.vstack((Y, outliers\_Y))

# Модельді құру және үйрету

model = LinearRegression()

model.fit(X\_with\_outliers, Y\_with\_outliers)

# Болжам жасау

Y\_pred = model.predict(X\_with\_outliers)

# Модельді бағалау

mse = mean\_squared\_error(Y\_with\_outliers, Y\_pred)

r2 = r2\_score(Y\_with\_outliers, Y\_pred)

print(f"MSE: {mse:.2f}")

print(f"R² Score: {r2:.2f}")

# Нәтижені визуализациялау

plt.scatter(X\_with\_outliers, Y\_with\_outliers, label='Деректер (ауытқулармен)', alpha=0.7)

plt.plot(X\_with\_outliers, Y\_pred, color='red', label='Регрессия сызығы')

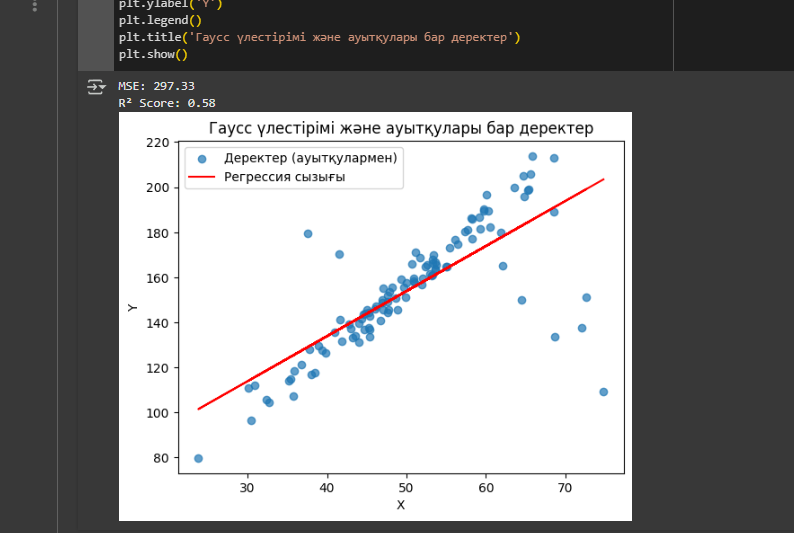
plt.xlabel('X')

plt.ylabel('Y')

plt.legend()

plt.title('Гаусс үлестірімі және ауытқулары бар деректер')

plt.show()



9 Есеп Әртүрлі қолданушылардан алынған деректердің біркелкі болмауы модельдің нәтижелеріне қалай әсер ететінін талда

import numpy as np

import pandas as pd

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_absolute\_error

# Біркелкі емес деректер жасау

data = {

"user": ["A", "B", "C", "D", "E"], # Бұл баған сандық емес

"feature1": [10, 200, np.nan, 40, 50], # NaN мәні бар

"feature2": [1, 2, 3, 4, 100], # Аномалды мән бар

"target": [15, 210, 30, 45, 55]

}

df = pd.DataFrame(data)

# Тек сандық бағандарды алу және NaN мәндерді толтыру

df\_numeric = df.select\_dtypes(include=[np.number])

df[df\_numeric.columns] = df\_numeric.fillna(df\_numeric.mean())

# Фичаларды (ерекшеліктерді) және мақсатты мәнді бөлу

X = df[["feature1", "feature2"]]

y = df["target"]

# Модельді үйрету

model = LinearRegression()

model.fit(X, y)

# Болжам жасау

y\_pred = model.predict(X)

# Қателікті өлшеу

mae = mean\_absolute\_error(y, y\_pred)

print(f"Болжам нәтижелері: {y\_pred}")

print(f"Орташа абсолютті қате (MAE): {mae}")

