

Võ Đình Khôi

24c15046

✓ Đầu bài:  Dự đoán giá cổ phiếu bằng Tối ưu lồi dựa trên dữ liệu lịch sử

1. Mô tả bài toán

Chúng ta sẽ xây dựng mô hình hồi quy tuyến tính để dự đoán giá đóng cửa của một cổ phiếu dựa trên dữ liệu lịch sử. Hai phương pháp chính sẽ được sử dụng:

- Hồi quy đơn biến:** Dự đoán giá cổ phiếu dựa trên giá trị của 1 ngày trước đó.
- Hồi quy đa biến:** Dự đoán giá cổ phiếu dựa trên k ngày trước đó.

Mô hình này sử dụng giả định rằng giá cổ phiếu có xu hướng phụ thuộc vào giá trị trong quá khứ và có thể tìm ra mối quan hệ tuyến tính giữa chúng.

2. Dữ liệu đầu vào

Dữ liệu lịch sử giá cổ phiếu bao gồm:

- Ngày giao dịch**
- Giá đóng cửa (Close Price)**

Ví dụ về dữ liệu:

Ngày	Giá đóng cửa (Close)
2024-02-01	100
2024-02-02	102
2024-02-03	101
2024-02-04	105

3. Yêu cầu

3.1 Hồi quy đơn biến (1 ngày trước đó)

- Xây dựng mô hình hồi quy tuyến tính với đầu vào $X_t = \text{Close}_{t-1}$ và đầu ra $Y_t = \text{Close}_t$.
- Mô hình có dạng:

$$Y_t = aX_t + b$$

- Huấn luyện mô hình bằng phương pháp tối ưu lồi (Gradient Descent hoặc Least Squares).
- Dự đoán giá cho ngày tiếp theo.

3.2 Hồi quy đa biến (k ngày trước đó)

- Sử dụng hồi quy tuyến tính đa biến với đầu vào $X_t = (\text{Close}_{t-1}, \text{Close}_{t-2}, \dots, \text{Close}_{t-k})$.
- Mô hình có dạng:

$$Y_t = w_1X_{t-1} + w_2X_{t-2} + \dots + w_kX_{t-k} + b$$

- Huấn luyện mô hình và dự đoán giá cho ngày tiếp theo.

4. Tìm nghiệm cho w bằng đạo hàm

Hàm mất mát của hồi quy tuyến tính là:

$$J(w) = \frac{1}{2m} \sum_{i=1}^m (Y_i - (w^T X_i + b))^2$$

Đạo hàm theo w :

$$\frac{\partial J}{\partial w} = -\frac{1}{m} \sum_{i=1}^m X_i (Y_i - (w^T X_i + b))$$

Giải phương trình đạo hàm bằng cách đặt $\nabla J(w) = 0$:

$$w = (X^T X)^{-1} X^T Y$$

Đây là nghiệm tối ưu của w trong hồi quy tuyến tính sử dụng Least Squares.

5. Cách đánh giá mô hình

- Sử dụng **Mean Squared Error (MSE)** để đo độ chính xác:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_{\text{true},i} - Y_{\text{pred},i})^2$$

- So sánh kết quả giữa mô hình đơn biến và đa biến.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split

# Tạo dữ liệu giả lập
np.random.seed(42)
days = np.arange(1, 101)
prices = 100 + np.cumsum(np.random.randn(100))

df = pd.DataFrame({'Day': days, 'Close': prices})

# Thêm cột giá của ngày trước đó
df['Close_prev1'] = df['Close'].shift(1)
df = df.dropna()

X_single = df[['Close_prev1']]
y_single = df['Close']

X_train, X_test, y_train, y_test = train_test_split(X_single, y_single, test_size=0.2, random_state=42)

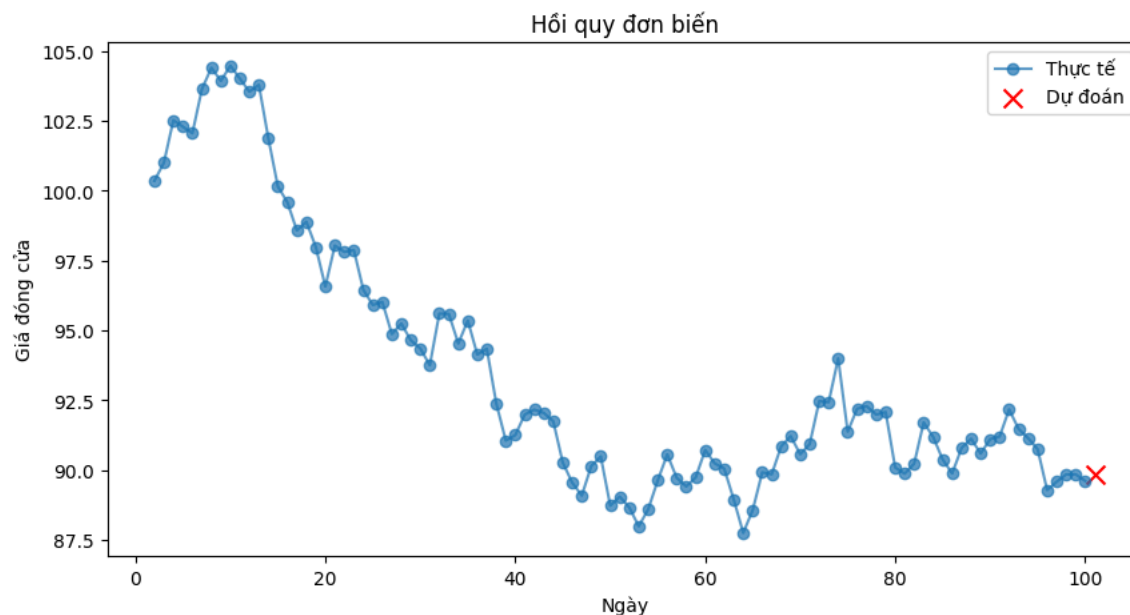
model_single = LinearRegression()
model_single.fit(X_train, y_train)
y_pred_single = model_single.predict(X_test)

mse_single = mean_squared_error(y_test, y_pred_single)
print(f"MSE (1 ngày trước): {mse_single:.4f}")

# Dự đoán giá ngày tiếp theo
latest_price = df.iloc[-1]['Close_prev1']
next_price_single = model_single.predict([[latest_price]])[0]
print(f"Dự đoán giá cổ phiếu ngày tiếp theo (1 ngày trước): {next_price_single:.2f}")

# Vẽ biểu đồ
plt.figure(figsize=(10, 5))
plt.plot(df['Day'], df['Close'], label='Thực tế', marker='o', linestyle='-', alpha=0.7)
plt.scatter(df['Day'].iloc[-1] + 1, next_price_single, color='red', label='Dự đoán', marker='x', s=100)
plt.xlabel("Ngày")
plt.ylabel("Giá đóng cửa")
plt.legend()
plt.title("Hồi quy đơn biến")
plt.show()
```

↩ MSE (1 ngày trước): 0.9184
Dự đoán giá cổ phiếu ngày tiếp theo (1 ngày trước): 89.85
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRe
warnings.warn()



```
k = 5 # Số ngày trước đó cần dùng
for i in range(1, k + 1):
    df[f'Close_prev{i}'] = df['Close'].shift(i)
df = df.dropna()

X_multi = df[[f'Close_prev{i}' for i in range(1, k + 1)]]
y_multi = df['Close']

X_train, X_test, y_train, y_test = train_test_split(X_multi, y_multi, test_size=0.2, random_state=42)

model_multi = LinearRegression()
model_multi.fit(X_train, y_train)
y_pred_multi = model_multi.predict(X_test)

mse_multi = mean_squared_error(y_test, y_pred_multi)
print(f"MSE (k ngày trước, k={k}): {mse_multi:.4f}")

# Dự đoán giá cổ phiếu ngày tiếp theo
latest_data = df.iloc[-1][[f'Close_prev{i}' for i in range(1, k + 1)]].values.reshape(1, -1)
next_price = model_multi.predict(latest_data)[0]
print(f"Dự đoán giá cổ phiếu ngày tiếp theo: {next_price:.2f}")

# Vẽ biểu đồ
plt.figure(figsize=(10, 5))
plt.plot(df['Day'], df['Close'], label='Thực tế', marker='o', linestyle='-', alpha=0.7)
plt.scatter(df['Day'].iloc[-1] + 1, next_price, color='red', label='Dự đoán', marker='x', s=100)
plt.xlabel("Ngày")
plt.ylabel("Giá đóng cửa")
plt.legend()
plt.title("Hồi quy đa biến")
plt.show()
```

↩ MSE (k ngày trước, k=5): 1.0251
Dự đoán giá cổ phiếu ngày tiếp theo: 89.71
/usr/local/lib/python3.11/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not have valid feature names, but LinearRe
warnings.warn()

