

TÌM HIỂU TRÍ TUỆ NHÂN TẠO AI

Lời Mở Đầu

Trong kỷ nguyên số, Trí tuệ Nhân tạo (Artificial Intelligence - AI) không chỉ đơn thuần là một công nghệ, mà đã trở thành biểu tượng cho khả năng mô phỏng và mở rộng tư duy của con người thông qua máy móc.

Mặc dù khái niệm về AI đã manh nha từ những năm 50 của thế kỷ trước, nhưng phải đến thời gian gần đây, nó mới thực sự có những bước nhảy vọt ngoạn mục. Những tiến bộ này không chỉ phản ánh sự trưởng thành của công nghệ, mà còn cho thấy một bước tiến vĩ đại trong nhận thức của con người. Chúng ta đang dần khai phá các mô hình của tự nhiên thông qua việc thấu hiểu chính bản thân mình, trong một quá trình có thể được xem như sự "bắt chước có ý thức" một cách tinh vi và sáng tạo.

Ngày nay, AI đang len lỏi vào hầu hết mọi lĩnh vực - từ y tế, tài chính, giáo dục cho đến nghệ thuật - và khẳng định vị trí là trái tim của cuộc Cách mạng Công nghiệp 4.0.

Mục tiêu của tài liệu này là trình bày những kiến thức toán học nền tảng của AI một cách đơn giản và dễ hiểu nhất, giúp bất kỳ ai cũng có thể tiếp cận và nhận biết được sức mạnh của nó.

Tài liệu được biên soạn với các mục đích cụ thể:

- Tính dễ tiếp cận: Giúp những người chưa có nền tảng về toán học hay AI có thể hiểu được những khái niệm cốt lõi một cách cơ bản nhất.
- Tính thực tiễn: Kết hợp hài hòa giữa lý thuyết, minh họa thực nghiệm và các ứng dụng trong thực tế.

Biên soạn: Trần Hoàng Linh petertranhoanglinh@gmail.com

Chương trình: AI ứng dụng từ zero

Chương 1 Toán học cơ bản

Bài 1 Hồi quy tuyến tính

1.1 Hồi quy tuyến tính cơ bản Linner Regression Basic

a. công thức

$$y = b_0 + b_1x$$

a. công thức tìm b_1

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

b. công thức tìm b_0

$$b_0 = \bar{y} - b_1\bar{x}$$

bài tập ví dụ.

- Diện tích: 30, 45, 60, 80, 100, 120 đơn vị m^2
- Giá nhà: 480, 630, 780, 1020, 1200, 1440 đơn vị triệu đồng
- Dùng hồi quy tuyến tính để dự đoán giá nhà khi diện tích là 80 m^2

Giải bài toán:

- Bước 1: Tìm giá trị X và Y trung bình:
 - $\bar{x} = \frac{30+45+60+80+100+120}{6} = 72.5$
 - $\bar{y} = \frac{480+630+780+1020+1200+1440}{6} = 925$
- Bước 2: Tìm b_1
 - lập bảng so sách phương sai của từng giá trị với giá trị trung bình

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$((x_i - \bar{x})^2)$	$((x_i - \bar{x})(y_i - \bar{y}))$
30	480	-42.5	-445	1806.25	18962.5
45	630	-27.5	-295	756.25	8112.5
60	780	-12.5	-145	156.25	1812.5
80	1020	7.5	95	56.25	712.5
100	1200	27.5	275	756.25	7562.5

x	y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$((x_i - \bar{x})^2)$	$((x_i - \bar{x})(y_i - \bar{y}))$
120	1440	47.5	515	2256.25	24462.5

◦ tổng các cột

- $\sum (x_i - \bar{x})^2 = 5787.5$

- $\sum (x_i - \bar{x})(y_i - \bar{y}) = 61625$

◦ tính hệ số $b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{5787}{61625} \approx 10.65$

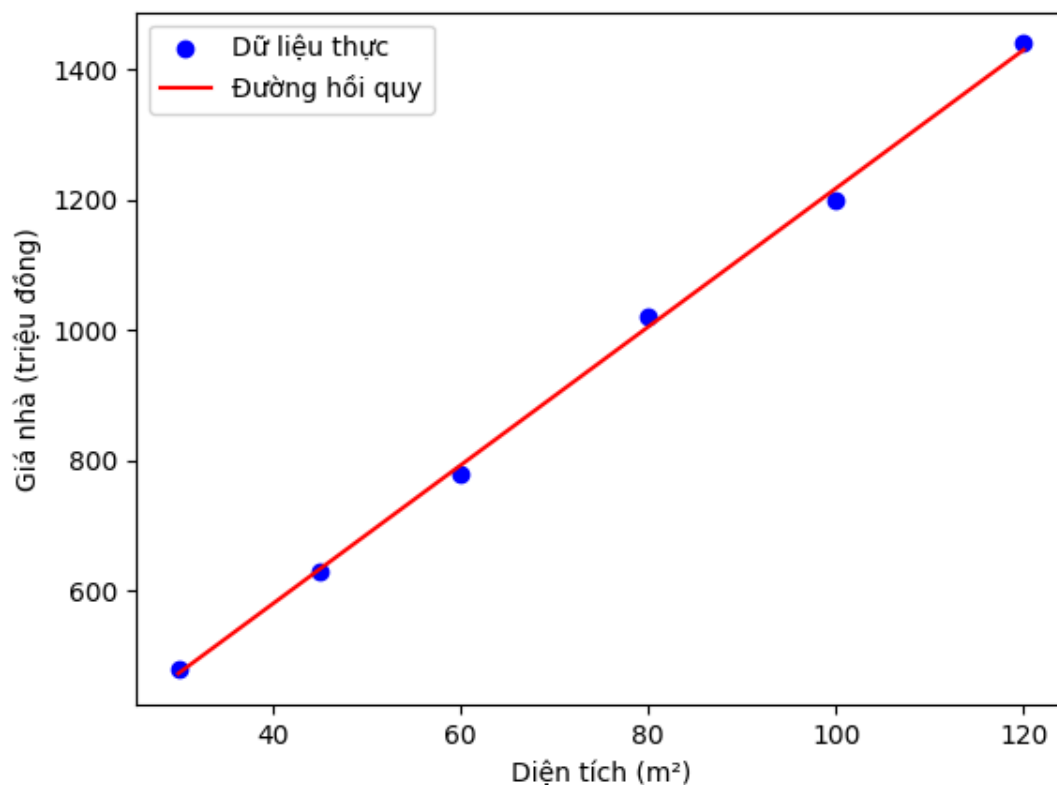
◦ tính hệ số $b_0 = \bar{y} - b_1 \bar{x} = 925 - 10.65(72.5) = 153.875$

Phương trình hồi quy dự đoán giá nhà theo m² có dạng:

$$y = 153.875 + 10.65x$$

Dựa vào phương trình hồi quy ta có thể dự đoán giá nhà khi diện tích là 80 m² là :

$$PriceHouse = 153.875 + 10.65 \times 80 = 1005.875(triệu)$$



Sử dụng trong python cơ bản

```
algorithms > linearRegressionBasic.py > ...
1  import numpy as np
2  from sklearn.linear_model import LinearRegression
3  import matplotlib.pyplot as plt
4
5  # Dữ liệu
6  x = np.array([30, 45, 60, 80, 100, 120]).reshape(-1, 1) # phải là dữ liệu 2d
7  y = np.array([480, 630, 780, 1020, 1200, 1440])
8
9  # Mô hình hồi quy
10 model = LinearRegression()
11 model.fit(x, y)
12
13 # Hệ số
14 b0 = model.intercept_
15 b1 = model.coef_[0]
16 r2 = model.score(x, y)
17
18 # Dự đoán giá nhà 90 m²
19 predict_90 = model.predict([[90]])
20
21 print(f"Phương trình hồi quy: y = {b0:.2f} + {b1:.2f}x")
22 print(f"Hệ số xác định R² = {r2:.3f}")
23 print(f"Giá dự đoán cho 90 m²: {predict_90[0]:.2f} triệu đồng")
24
25 # Vẽ biểu đồ
26 plt.scatter(x, y, color='blue', label='Dữ liệu thực')
27 plt.plot(x, model.predict(x), color='red', label='Đường hồi quy')
28 plt.xlabel('Diện tích (m²)')
29 plt.ylabel('Giá nhà (triệu đồng)')
30 plt.legend()
31 plt.show()
32
```

Ưu nhược điểm của thuật toán

Ưu điểm:

- Dễ triển khai
- Tính toán nhanh và đơn giản
- Ít tham số phức tạp

Nhược điểm:

- Chỉ thật sự hiệu quả và có độ chính xác cao với mô hình dữ liệu có tính tuyến tính
- Dễ bị nhiễu nếu dữ liệu chỉ cần 1 vài dữ liệu ko chính xác sẽ dẫn đến hệ thống bị sụp đổ
- Không phù hợp với bài toán đánh giá cần nhiều tiêu chí đầu vào.

1.2 Hồi quy tuyến tính bội Multiple Linear Regression

a. Khái niệm

Nếu ở mô hình trước bạn dự đoán giá nhà theo diện tích, thì trong hồi quy tuyến tính bội, bạn có thể dự đoán giá nhà theo nhiều yếu tố, ví dụ:

- Diện tích nhà (m²)

- Số phòng ngủ
- Khoảng cách đến trung tâm
- Năm xây dựng, v.v.

--> Một mô hình dự đoán có nhiều yếu tố quyết định chúng ta cần sử dụng tuyến tính bội để tạo mô hình thành vì tuyến tính cơ bản.

b. Công thức

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

c. Bàn luận logic

Nhìn vào công thức của mô hình hồi quy tuyến tính bội, ta thấy rằng để tìm được mối quan hệ giữa các biến độc lập x_n và biến phụ thuộc y , mô hình cần dựa trên một số giả định thống kê nhất định. Hồi quy tuyến tính không đơn thuần là vẽ một đường thẳng “cho đẹp”, mà là ước lượng chính xác nhất mối quan hệ thật giữa các biến, dựa trên những giả định giúp mô hình phản ánh đúng bản chất dữ liệu.

Ta cần dùng giả thuyết thống kê để tìm sự liên hệ giữa chúng

- Giả thuyết vô hiệu: $H_0 : b_1 = b_2 \dots = b_k = 0$

Nếu bạn chứng minh giả thuyết này đúng thì phương trình của bạn là vô nghĩa vì y không hề phụ thuộc x .