Machine learning cơ bản cho người lười

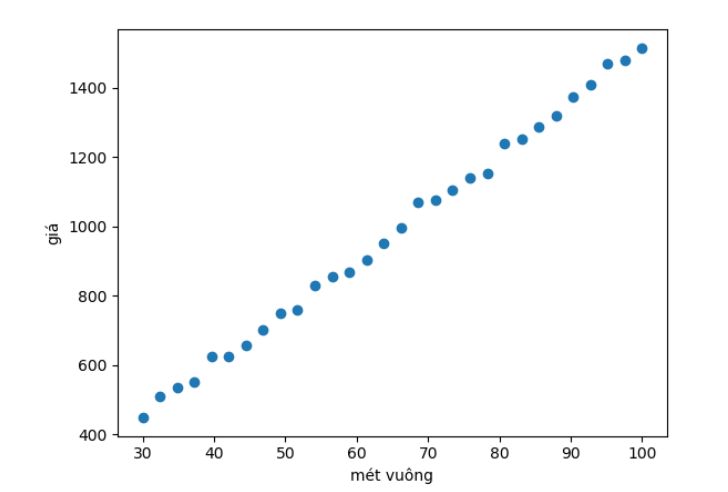
I, Mở đầu, yêu cầu khi học Machine learing(máy học)

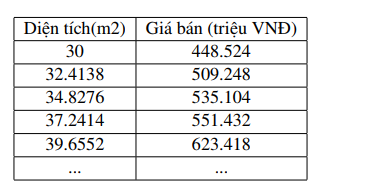
* ở đây sẽ ví dụ các thuật toán và công thức để tính toán sao cho đơn giản nhất
* với yêu cầu:
* biết lập trình python căn bản
* biết các *install* và *import* các thư viện trong máy

code: <https://colab.research.google.com/drive/1xsgvE0UTIU7esPQB-3lDQ51Y4a-Dr4hE?usp=sharing>

II, các thuật toán Machine Learning cơ bản(hiện tại 6)

1. linear Regression(hồi quy tuyến tính)
2. ví dụ

* Thuật toán trên giúp ta sử lý những bài toán với các kiểu dữ liệu phân bố theo tuyến tính:
* ví dụ : một miếng đất A có diện tích là X sẽ có giá thành là Y. khi X tăng thì Y tăng và khi X giảm thì Y giảm
* Nhược điểm: độ chính xác không cao

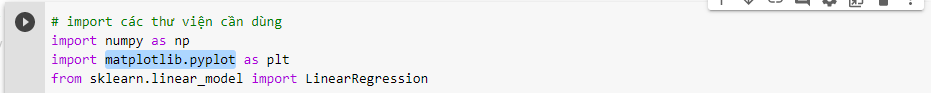


1. cách tính toán

B1: import các thư viện cần dùng:

ta cần thư viện *numpy* để xử lý ma trận. *matplotlib.pyplot*  để vẽ model.

và thư viện ***sklearn.linear\_model*** sau đó import *LinearRegression*



B2: nhập dữ liệu vào 2 mảng x, y. chú ý ta nên cho data về dạng matrận x[size \* len(x)]

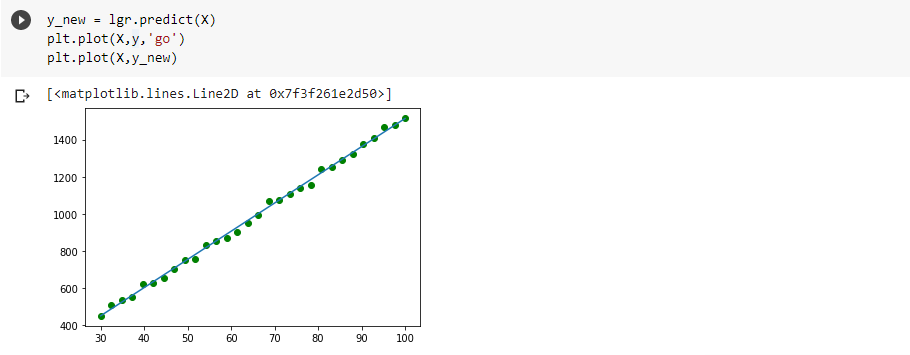
B3: dùng 1 biến để nhận thư viện *LinearRegression* VD: lgr = linearRegression()

* sau đó ta tiến hành fit data theo biến X\_ và Y\_ cho trước
* VD : lgr.fit(x\_, y\_)

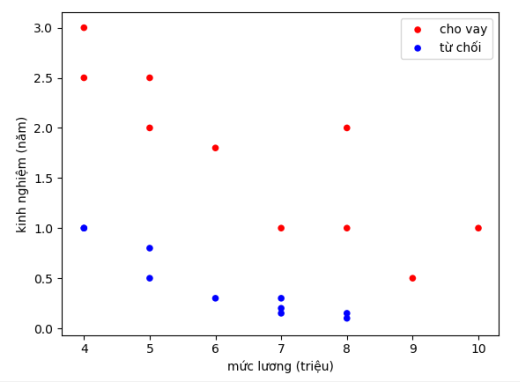
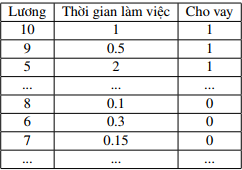


B4: dự đoán dữ liệu:

* ta dự đoán x theo y bằng cách: y\_new = lgr.predict(x) #ta có thể dự đoán bất cứ dữ liệu nào miễn sao nó hợp lệ
* sau đó dùng *matplotlib.pyplot*  để in ra biểu đồ



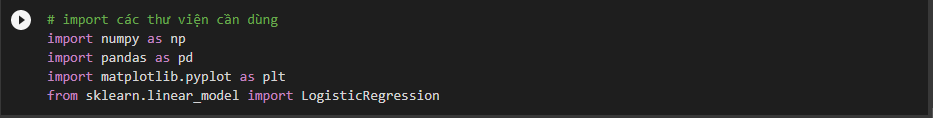
1. Logistic Regression
2. ví dụ

* Ngân hàng bạn đang làm có chương trình cho vay ưu đãi cho các đối tượng mua chung cư. Tuy nhiên gần đây có một vài chung cư rất hấp dẫn (giá tốt, vị trí đẹp,...) nên lượng hồ sơ người nộp cho chương trình ưu đãi tăng đáng kể. Bình thường bạn có thể duyệt 10-20 hồ sơ một ngày để quyết định hồ sơ có được cho vay hay không, tuy nhiên gần đây bạn nhận được 1000-2000 hồ sơ mỗi ngày. Bạn không thể xử lý hết hồ sơ và bạn cần có một giải pháp để có thể dự đoán hồ sơ mới là có nên cho vay hay không
* Sau khi phân tích thì bạn nhận thấy là hai yếu tố chính quyết định đến việc được vay tiền đó  
  là mức lương và thời gian công tác. Đây là dữ liệu bạn có từ trước đến nay:
* công việc của Logistic Regression: phân tích xem với những giá trị nào là 1 hay 0.
* đối với những dữ liệu lớn mình cũng sử phân tích theo ví dụ trên.

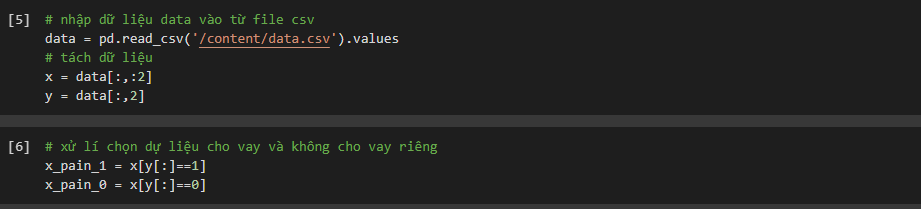
1. cách tính toán

B1: import các thư viện cần dùng:

quan trọng nhất thư viện ***sklearn.linear\_model*** sau đó import *LogisticRegression*



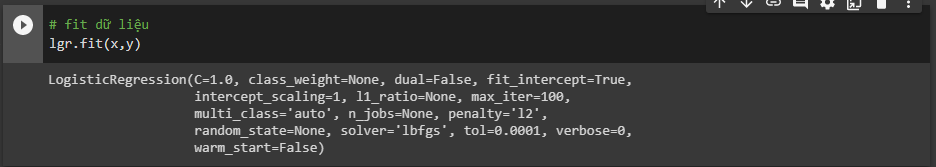
B2: nhập và xử lý các dữ liệu train:



B3: khởi tạo thư viện logisticRegression



B4: fit data



B5: dự đoán dữ liệu, với tập dữ liệu test

