**TRƯỜNG ĐẠI HỌC ĐIỆN LỰC**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO HỌC PHẦN HỌC MÁY NÂNG CAO**

**ĐỀ TÀI:Kỹ Thuật Xây Dựng Hàm Loss Với Tensorflow**

|  |  |
| --- | --- |
|  |  |
| |  |  | | --- | --- | | **Sinh viên thực hiện** | **: Nguyễn Đình Dương** | | **Giảng viên hướng dẫn** | **: Phạm Thị Kim Dung** | | | **Ngành** | **: CÔNG NGHỆ THÔNG TIN** | | | **Chuyên ngành** | **: CÔNG NGHỆ PHẦN MỀM** | | | **Lớp** | **: D12 CNPM2** | | | **Khóa** | **: 2017-2021** | | |  |
| **Hà Nội,tháng 1 năm 2021** |  |

**PHIẾU CHẤM ĐIỂM**

Sinh viên thực hiện:

|  |  |  |  |
| --- | --- | --- | --- |
| **Họ và tên** | **Chữ ký** | **Nội dung thực hiện** | **Điểm** |
| Nguyễn Đình Dương |  |  |  |

Giảng viên chấm:

|  |  |  |
| --- | --- | --- |
| **Họ và tên** | **Chữ ký** | **Ghi chú** |
| Giảng viên chấm 1 : |  |  |
| Giảng viên chấm 2 : |  |  |

# 

# **Lời cảm ơn**

Đầu tiên em xin gửi lời cảm ơn chân thành xâu sắc đến các thầy cô trường Đại Học Điện Lực khia Công Nghệ Thông Tin đã tạo cho em điều kiện cho chúng em học tập và làm việc.Và em cũng xin chân thành cảm ơn cô Phạm Thị Kim Dung đã nhiệt tình hướng dẫn chúng em hoàn thành báo cáo kết thúc môn học này.

Có được báo cáo này, chúng em xin bày tỏ lòng biết ơn chân thành và sâu sắc, cũng như trong quá trình làm báo cáo, khó tránh khỏi sai sót, rất mong các thầy, cô giáo bỏ qua. Đồng thời do trình độ lý luận cũng như kinh nghiệm thực tiễn còn hạn chế nên báo cáo không thể tránh khỏi những thiếu sót, em rất mong được ý kiến đóng góp của thầy, cô để em học thêm được nhiều kinh nghiệm và sẽ hoàn thành tốt hơn những bản báo cáo sắp tới.

**CHƯƠNG II.BÀI TOÁN HỌC MÁY**

**I.Mô tả dữ liệu**

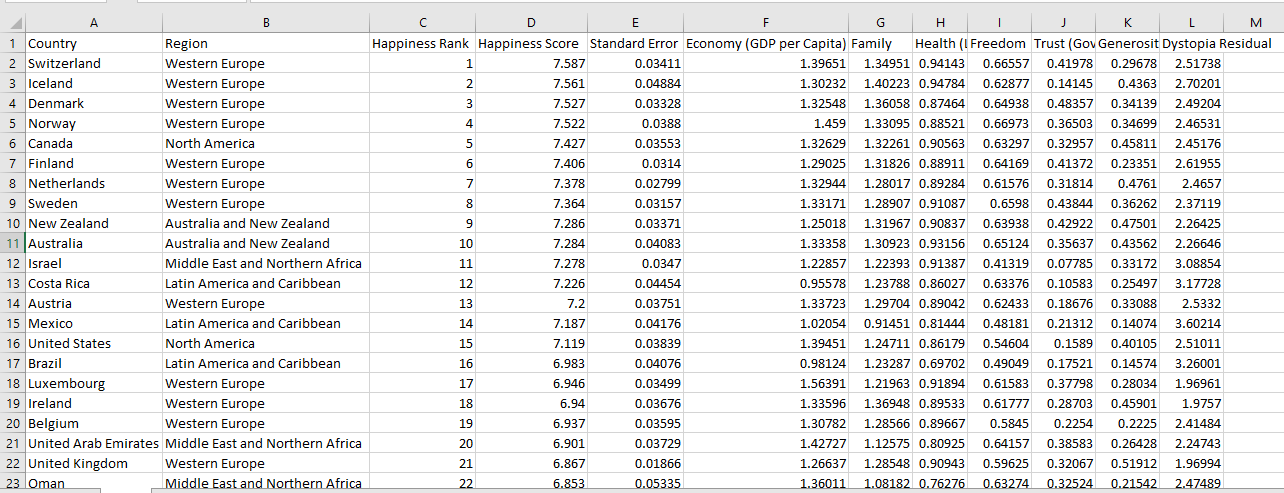
Ta có bộ dữ liệu Word Happiness Report.csv (báo cáo hạnh phúc thế giới 2015)

Báo cáo Hạnh phúc Thế giới là một cuộc khảo sát mang tính bước ngoặt về tình trạng hạnh phúc toàn cầu. Báo cáo đầu tiên được xuất bản vào năm 2012, lần thứ hai vào năm 2013, lần thứ ba vào năm 2015 và lần thứ tư trong Bản cập nhật năm 2016. Hạnh phúc Thế giới 2017, xếp hạng 155 quốc gia theo mức độ hạnh phúc, đã được Liên hợp quốc công bố tại một sự kiện kỷ niệm Ngày Quốc tế Hạnh phúc vào ngày 20 tháng 3. Báo cáo tiếp tục nhận được sự công nhận trên toàn cầu khi các chính phủ, tổ chức và xã hội dân sự ngày càng sử dụng các chỉ số hạnh phúc để thông báo cho các quyết định hoạch định chính sách của họ. Các chuyên gia hàng đầu trong các lĩnh vực - kinh tế, tâm lý học, phân tích khảo sát, thống kê quốc gia, y tế, chính sách công và hơn thế nữa - mô tả cách các phép đo mức độ hạnh phúc có thể được sử dụng hiệu quả để đánh giá sự tiến bộ của các quốc gia.

Các phần còn lại, hoặc các thành phần không giải thích được, khác nhau ở mỗi quốc gia, phản ánh mức độ mà sáu biến đánh giá trung bình quá mức hoặc không giải thích được về tuổi thọ 2014-2016. Các phần dư này có giá trị trung bình xấp xỉ 0 trên toàn bộ các quốc gia. Hình 2.2 cho thấy phần dư trung bình của mỗi quốc gia khi áp dụng phương trình trong Bảng 2.1 cho dữ liệu trung bình 2014-2016 cho sáu biến ở quốc gia đó. Chúng tôi kết hợp các phần dư này với ước tính cho các đánh giá tuổi thọ trong Dystopia để thanh kết hợp sẽ luôn có giá trị dương. Như có thể thấy trong Hình 2.2, mặc dù một số phần dư đánh giá tuổi thọ khá lớn, đôi khi vượt quá một điểm trên thang điểm từ 0 đến 10, chúng luôn nhỏ hơn nhiều so với giá trị tính toán trong Dystopia, nơi tuổi thọ trung bình được đánh giá là 1,85 trên thang điểm từ 0 đến 10.

**Các cột kế tiếp Điểm hạnh phúc (như Gia đình, Sự hào phóng, v.v.) mô tả điều gì?**

Các cột sau: GDP trên mỗi đầu người, Gia đình, Kỳ vọng cuộc sống, Tự do, Sự hào phóng, Chính phủ tin cậy Tham nhũng mô tả mức độ mà những yếu tố này đóng góp vào việc đánh giá mức độ hạnh phúc ở mỗi quốc gia.



**1.Bảng thông số thuộc tính của dữ liệu**

**-**Dữ liệu gồm có 11 cột và 157 dòng

|  |  |  |
| --- | --- | --- |
| STT | Tên thuộc tính | Mô tả |
| 1 | Country | Tên đất nước |
| 2 | Region | Khu vực quốc gia thuộc về Châu Phi cận Sahara,Trung và Đông Âu |
| 3 | Happiness Rank | Xếp hạng các quốc gia dựa trên Điểm hạnh phúc |
| 4 | Happiness Score | Xếp hạng điểm hạnh phúc |
| 5 | Standard Error | Sai số tiêu chuẩn của điểm dạnh phúc |
| 6 | Economy(GDP per Capita) | Mức độ đóng góp GDP vào việc tính điểm hạnh phúc |
| 7 | Family | Mức độ mà gia đình đóng góp vào việc tính điểm hạnh phúc |
| 8 | Health(Life Expectancy) | Mức độ mà tuổi thọ đã đóng góp vào việc tính điểm hạnh phúc |
| 9 | Freedom | Mức độ mà tự do đã đóng góp vào tính điểm hạnh phúc |
| 10 | Trust(Government Corruption) | Mức độ mà nhận thức về tham những đóng góp vào điểm hạnh phúc |
| 11 | Generosity | Mức độ về sự hào phóng đóng góp vào điểm hạnh phúc |

**2.Mô tả bài toán học máy:**

+ Bài toán dự đoán về mức độ chỉ số hạnh phúc, của các quốc gia thuộc về Châu Phi cận Sahara, Trung và Đông Âu dựa trên các tiêu chuẩn như (mức độ đóng góp GDP, độ tuổi, chỉ số tự do,nhận thức, và sự hào phóng). Để từ đó đánh giá được mức độ hạnh phúc

+ Do dữ liệu bài toán có sự liên tục nên chúng ta chọn bài toán Hồi Quy và sử dụng kỹ thuật học máy là Linear Regression

+ Phương pháp đánh giá chọn sai số bình phương trung bình (MSE) của công cụ ước tính (của thủ tục ước tính số lượng không quan sát được) đo trung bình bình phương của các lỗi – nghĩa là chênh lệch bình phương trung bình giữa các giá trị ước tính và giá trị ước tính. MSE là một hàm rủi ro, tương ứng với giá trị dự kiến ​​của mất lỗi bình phương. Việc MSE hầu như luôn luôn tích cực (chứ không phải bằng không) là do tính ngẫu nhiên hoặc do công cụ ước tính không tính đến thông tin có thể tạo ra ước tính chính xác hơn, và độ lệch chuẩn của phần dư (RMSE lỗi dự đoán ) . Phần dư là thước đo khoảng cách từ các điểm dữ liệu đường hồi quy; RMSE là thước đo mức độ lan truyền của những phần dư này. Nói cách khác, nó cho bạn biết mức độ tập trung của dữ liệu xung quanh dòng phù hợp nhất . Lỗi bình phương trung bình thường được sử dụng trong khí hậu học, dự báo và phân tích hồi quy để xác minh kết quả thí nghiệm.

**3.Mô tả chi tiết cách sử lý dữ liệu:**

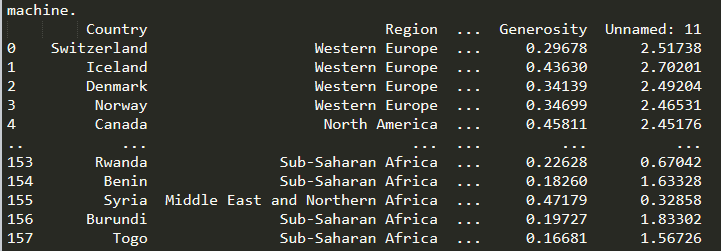
* Load dữ liệu đã được lấy từ Kaggle

+Code

df = pd.read\_csv('C:/hoc\_may/2015.csv')

print(df)

KQ:



* Đưa những dữ liệu dạng chữ về dạng số

+Code

from sklearn.preprocessing import LabelEncoder

lb\_make = LabelEncoder()

#X["outlook"] = lb\_make.fit\_transform(X["outlook"])

# X\_LabelEncoder = X.apply(lb\_make.fit\_transform)

from sklearn import preprocessing

df\_LabelEncoder = df.apply(lambda col: preprocessing.LabelEncoder().fit\_transform(col))

X\_LabelEncoder = df\_LabelEncoder.drop('Economy (GDP per Capita)',axis=1)

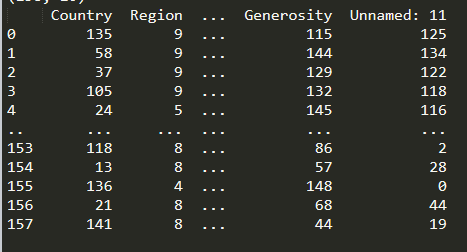
y\_LabelEncoder = df\_LabelEncoder['Economy (GDP per Capita)']

print(X\_LabelEncoder)

print(y\_LabelEncoder)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X\_LabelEncoder, y\_LabelEncoder, test\_size=0.2)

KQ:



-Huấn luyện và kỹ thuật học máy

+Code

mode = DecisionTreeRegressor()

# Huấn luyện mô hình với X\_train, y\_bina\_train

mode = mode.fit(X\_train, y\_train)

# Dự đoán mô hình với X\_test, y\_bina\_test

y\_predicted = mode.predict(X\_test)

# Hiện thị KQ

# The mean squared error

print('Mean squared error DecisionTreeRegressor: %.2f'

% mean\_squared\_error(y\_test, y\_predicted))

# from sklearn.metrics import confusion\_matrix,classification\_report

# print(classification\_report(y\_test, y\_predicted))

# print(confusion\_matrix(y\_test, y\_predicted))

######################################################

regr = linear\_model.LinearRegression()

# Train the model using the training sets

regr.fit(X\_train, y\_train)

# Make predictions using the testing set

y\_pred = regr.predict(X\_test)

# The coefficients

print('Coefficients: \n', regr.coef\_)

print('Bias: \n', regr.intercept\_)

regr = linear\_model.LinearRegression()

# Train the model using the training sets

regr.fit(X\_train, y\_train)

# Make predictions using the testing set

y\_pred = regr.predict(X\_test)

# The coefficients

print('Coefficients: \n', regr.coef\_)

print('Bias: \n', regr.intercept\_)

KQ :



+Sai số trung bình bình phương tuyến tính

+Code

# The mean squared error(sai số bình phương trung bình)

print('Mean squared error LinearRegression: %.2f'

% mean\_squared\_error(y\_test, y\_pred))

KQ:



-Hình vẽ

+

+Code

from sklearn.linear\_model import LinearRegression

model = LinearRegression()

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

y\_predict = model.predict(X\_test)

##y\_train = PCA(1).fit\_transform(y\_train)

#X\_test = PCA(1).fit\_transform(X\_test)

#plt.scatter(X\_train, y\_train, color='green')

plt.scatter(y\_test, y\_predict, color='black')

plt.plot(y\_predict, y\_predict, color='blue', linewidth=3)

plt.xticks(())

plt.yticks(())

plt.title('Linear regression for 2015')

plt.xlabel('X')

plt.ylabel('y')

plt.show()

n\_features = X\_train.shape[1]

# define model

model = Sequential()

model.add(Dense(10, activation='relu', kernel\_initializer='he\_normal', input\_shape=(n\_features,)))

model.add(Dense(8, activation='relu', kernel\_initializer='he\_normal'))

model.add(Dense(1))

# compile the model

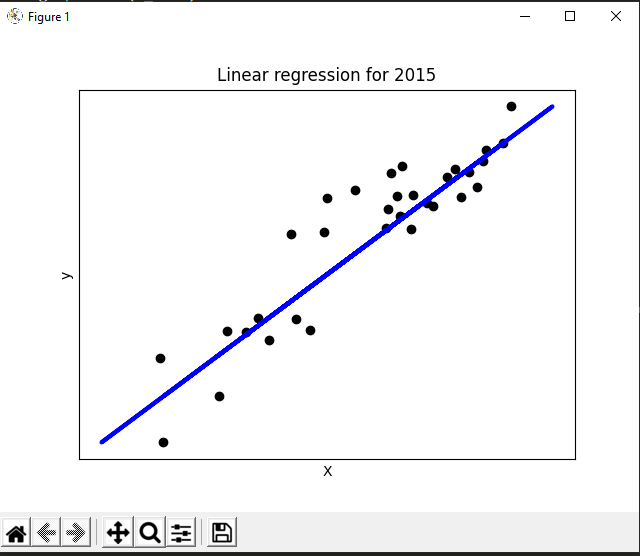
model.compile(optimizer='adam', loss='mse')

#‘mse‘ (mean squared error) for regression.

# fit the model

model.fit(X\_train, y\_train, epochs=250, batch\_size=32, verbose=0)

KQ:



-Phương pháp đánh giá:

-Dùng phương pháp MSE và RMSE

+Code:

error = model.evaluate(X\_test, y\_test, verbose=0)

print('MSE: %.3f, RMSE: %.3f' % (error, sqrt(error)))

KQ:



-Hiển thị ra y\_test và y\_predict

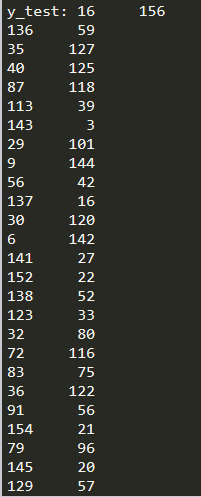
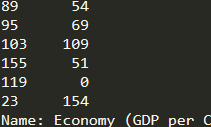
+Code:

print("y\_test:", y\_test)

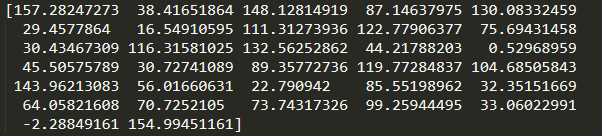
print(y\_predict)

KQ:

Y test

Y predict



CHƯƠNG III KẾT LUẬN