import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import random

from sklearn.model\_selection import train\_test\_split #para dividir los datos

from sklearn.linear\_model import LinearRegression

from mpl\_toolkits.mplot3d import Axes3D

dataset = pd.read\_csv('Salarios.csv')

#Agregar columna países

paises = ['CO', 'MX', 'PE', 'BR', 'US']

dataset['Pais'] = [paises[np.random.randint(0,len(paises))] for i in range(len(dataset)) ]

dataset['IDPais'] = pd.factorize(dataset['Pais'])[0]

print('Dataset inicial\n\n', dataset.head())

#Asigno valores para x,y

x = dataset[['Aexperiencia','IDPais']]

y = dataset['Salario']

#x  = dataset.iloc[:, :-1].values  # Primer campo (de dos)

#y = dataset.iloc[:, 1].values     # Ultimo campo

#30% prueba 70% a confirmar

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(x,y, test\_size = 0.3, random\_state=0)

# ENTRENA CON fit

regressor.fit(X\_train, Y\_train)

ajuste = regressor.score(X\_test,Y\_test)

print('\nEl porcentaje de ajuste es: ', ajuste\*100, '%\n')

# ploteamos datos de entrenamiento para ver si realmente el

# modelo lo hizo bien, se entreno como se debe

#Grafico de los resultados de entrenamiento

fig = plt.figure(dpi=150)

viz\_train = fig.add\_subplot(111, projection='3d')

viz\_train.scatter(X\_train['Aexperiencia'], X\_train['IDPais'], Y\_train, color = 'blue', label = 'Train')

viz\_train.scatter(X\_test['Aexperiencia'], X\_test['IDPais'], regressor.predict(X\_test), color = 'red', label = 'Test')

viz\_train.plot\_trisurf(X\_train['Aexperiencia'],X\_train['IDPais'], regressor.predict(X\_train),color = 'black', alpha = 0.5)

viz\_train.set\_title('Salario, Experiencia y País')

viz\_train.set\_xlabel('Experiencia (años)')

viz\_train.set\_ylabel('País')

viz\_train.set\_zlabel('Salario (USD)')

viz\_train.set\_yticks(range(len(paises)))

viz\_train.set\_yticklabels(paises)

plt.legend(loc="upper left")

fig.show

Dataset inicial

Aexperiencia Salario Pais IDPais

0 10.0 42500 MX 0

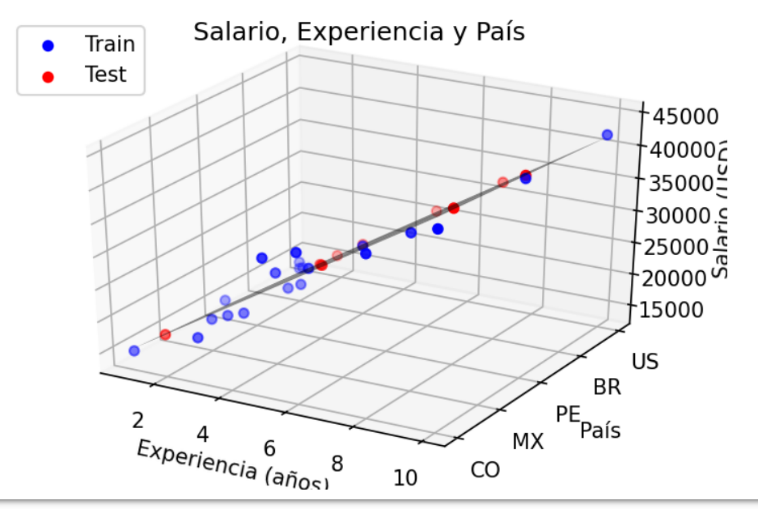
1 8.0 36302 US 1

2 5.4 35000 US 1

3 2.5 18103 US 1

4 1.0 15000 BR 2

El porcentaje de ajuste es: 85.49850635695397 %



#Grafico de los resultados de entrenamiento

fig = plt.figure()

viz\_train = fig.add\_subplot(111, projection='3d')

viz\_train.scatter(X\_train['Aexperiencia'],X\_train['IDPais'], Y\_train, color = 'blue')

viz\_train.plot\_trisurf(X\_train['Aexperiencia'],X\_train['IDPais'], regressor.predict(X\_train),color = 'black', alpha = 0.5)

viz\_train.set\_title('Salario Experiencia y pais')

viz\_train.set\_xlabel('Experiencia')

viz\_train.set\_ylabel('Pais')

viz\_train.set\_zlabel('Salario')

from sklearn import preprocessing

le1 = preprocessing.LabelEncoder()

paises\_encoded = le1.fit\_transform(paises)

viz\_train.set\_yticks(range(len(paises\_encoded)))

viz\_train.set\_yticklabels(le1.inverse\_transform(paises\_encoded))

viz\_train.azim=-10

fig.show



viz\_train.azim=80

