

## Tópicos Especiais em Ciência da Computação II

Maria Elisa Lima Pedro

### Atividade - Regressão múltipla

```
from sklearn.linear_model import LinearRegression
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from scipy import stats

dados = pd.read_csv('precos_planos_saude.csv')

dadosOriginais = dados.copy()

dicionario = {'Sedentário': 0, 'Leve': 1, 'Moderado': 2, 'Intenso': 3}
dados['Fumante'] = dados['Fumante'].map(lambda x: 1 if x == 'Sim' else 0)
dados['Histórico_Familiar'] = dados['Histórico_Familiar'].map(lambda x: 1 if x == 'Sim' else 0)
dados['Atividade_Física'] = dados['Atividade_Física'].map(lambda x: dicionario[x])
dadosTratados = dados.copy()

precoPlano = dados["Preço_Plano_R$"]
dados.drop('Região', axis='columns', inplace=True)
dados.drop("Preço_Plano_R$", axis='columns', inplace=True)
dados.drop("ID_Cliente", axis='columns', inplace=True)
correlacao = dados.corr()

for i in range(len(correlacao)):
    for j in range(i + 1, len(correlacao.iloc[i])):
        if abs(correlacao.iloc[i,j]) > 0.5:
            print(correlacao.iloc[i,j])
            print(i,j)
            print(correlacao.index[i])
            print(correlacao.columns[j])
            print()
def reta(a, b, x):
    return ((a*x) + b)

plt.title("Peso x IMC")
plt.xlabel("Peso (Kg) ")
plt.ylabel("IMC")
```

```

plt.scatter(dados["Peso_kg"], dados["IMC"], label = "IMC em funcao do
Peso")

a,b,r,p,std_err=stats.linregress(dados["Peso_kg"], dados["IMC"])
print(a,b)
valoresY = []
valoresX = []
for x in range(np.min(dados["Peso_kg"]), np.max(dados["Peso_kg"])):
    valoresX.append(x)
    valoresY.append(reta(a,b,x))
plt.plot(np.array(valoresX), np.array(valoresY), c="r", label =
"Regressao Linear do IMC em funcao do Peso")
plt.legend()
plt.show()
regressaoMultipla = LinearRegression()

x = dadosTratados[["Idade", "Peso_kg", "Altura_m", "IMC", "Fumante",
"Histórico_Familiar", "Atividade_Física"]]

regressaoMultipla.fit(x, dadosTratados["Preço_Plano_R$"])

yPred = regressaoMultipla.predict(x)

plt.scatter(dadosTratados["Idade"], dadosTratados["Preço_Plano_R$"],
c="blue", label="Preço real")
plt.scatter(dadosTratados["Idade"], yPred, c="orange",
label="Previsão")
plt.xlabel("Idade (anos)")
plt.ylabel("Preço do Plano de Saúde (R$)")
plt.legend()
plt.show()

```

