

Aula 3 - aprendizado de máquina com scikit-learn

Jayme Anchante

24 de fevereiro de 2021

Motivação

carregando os dados

```
1 from sklearn.datasets import load_iris
2 iris = load_iris()
```

descrição dos dados

```
1 print(iris.DESCR)
```

montando base única

Crie um objeto chamado `df` que seja um `pd.DataFrame` com os atributos e as espécies de iris.

Dica: qual o tipo do objeto iris? veja as chaves.

montando base

```
1 df = pd.DataFrame(iris.data, columns=iris.feature_names)
2 df.loc[:, "species"] = iris.target
3 def f(x):
4     return iris.target_names[x]
5 df.loc[:, "species"] = df["species"].apply(f)
```

Desafio

Dada a iris

```
1  vals = [[2, 3, 2.5, 0.7]]  
2  uma_iris = pd.DataFrame(vals, columns=iris.feature_names)
```

A qual espécie de íris esta flor pertence?

Algoritmo

Crie um algoritmo/função que receba como argumento `uma_iris` e retorne um texto com a espécie de iris que ela seria?

Dica: Comece escrevendo em linguagem natural (português) as etapas que você gostaria de seguir, depois tente implementar as etapas.

Dica: explore os dados, medidas de centralidade, dispersão, agrupamento etc.

Regras de negócios

```
1 col = "petal length (cm)"
2 print(df.groupby("species")[col].mean())
3 if uma_iris.loc[0, col] < 3:
4     print("setosa")
5 elif uma_iris.loc[0, col] >= 3 and uma_iris.loc[0, col] < 4.5:
6     print("versicolor")
7 elif uma_iris.loc[0, col] >= 4.5:
8     print("virginica")
```

Algoritmo das médias

Regressão linear

```
1 medias = df.groupby("species").mean()
2 diff = medias - uma_iris.loc[0, ]
3 diff_abs = diff.abs()
4 diff_abs.idxmin()
5 vc = s.value_counts()
6 vc.index[0]
```

Vizinhos

```
1 df = df.set_index("species")
2 diff = df - uma_iris.loc[0]
3 diff_abs = diff.abs()
4 soma = diff_abs.sum(axis=1)
5 soma.idxmin()
6 # n é a quantidade de vizinhos
7 n = 3
8 soma.sort_values().head(n)
```


