



LISBON  
SCHOOL OF  
ECONOMICS &  
MANAGEMENT  
UNIVERSIDADE DE LISBOA

Carlos J. Costa

**SCIKIT-LEARN**



# scikit-learn

- A scikit-learn (anteriormente scikits.learn)
- É uma biblioteca de machine learning em Open Source
- Inclui vários algoritmos de classificação, regressão, clustering, redução dimensional, selecção de modelos, pré-processamento
- <https://scikit-learn.org/>
- <https://scikit-learn.org/stable/index.html>

# scikit-learn

- Dividir a amostra em treino e teste

```
from sklearn.model_selection import train_test_split
```

```
x_train, x_test, y_train, y_test = train_test_split(X, Y,  
test_size=0.4, random_state=50)
```

# scikit-learn

- Vamos então fazer regressão...

- Importar modulo relevante:

```
from sklearn.linear_model import LinearRegression
```

- Criar modelo

```
lm = LinearRegression()
```

- Ajustar modelo

```
lm.fit(x_train,y_train)
```

# scikit-learn

- coeficiente de determinação  $R^2$
- `Im.score(X, Y)`
- Coeficientes
- `Im.coef_`

# scikit-learn

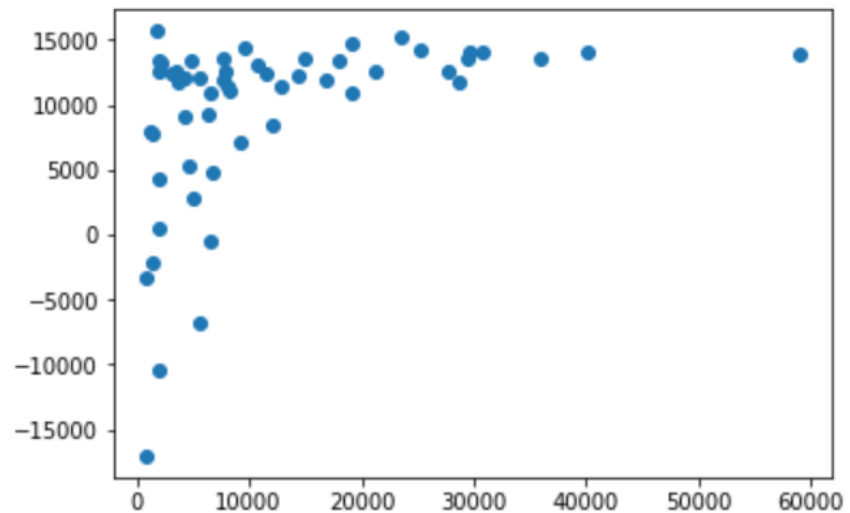
Não confundir  
com perdição!

- Vamos agora fazer predição  
predicao = lm.predict(x\_test)

# scikit-learn

- A utilização de uma nova biblioteca agora pode ser relevante a matplotlib (serve para fazer gráficos)

```
import matplotlib.pyplot as plt  
plt.scatter(y_test, predicao)
```



# Bibliografia

- <https://pandas.pydata.org/>
- [https://pandas.pydata.org/pandas-docs/stable/getting\\_started/10min.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html)
- <https://scikit-learn.org/>
- <https://scikit-learn.org/stable/index.html>
- <https://www.statsmodels.org/stable/index.html>