

# seance6\_graphes\_ml\_correction

November 28, 2021

## 1 Graphes en machine learning - correction

Correction (en cours de rédaction) des exercices autour des graphes courants en machine learning.

```
[1]: %matplotlib inline
     %load_ext pyensae
```

```
[2]: import matplotlib.pyplot as plt
     plt.style.use('ggplot')
```

```
[3]: from jyquickhelper import add_notebook_menu
     add_notebook_menu()
```

```
[3]: <IPython.core.display.HTML object>
```

Le module utilise des données issue de [Wine Quality Data Set](https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/) pour lequel on essaye de prédire la qualité du vin en fonction de ses caractéristiques chimiques.

```
[4]: from pyensae.datasources import download_data, DownloadDataException
     uci = "https://archive.ics.uci.edu/ml/machine-learning-databases/wine-quality/"
     try:
         download_data("winequality-red.csv", url=uci)
         download_data("winequality-white.csv", url=uci)
     except DownloadDataException:
         print("backup")
         download_data("winequality-red.csv", website="xd")
         download_data("winequality-white.csv", website="xd")
```

```
[5]: %head winequality-red.csv
```

```
[5]: <IPython.core.display.HTML object>
```

```
[6]: import pandas
     red_wine = pandas.read_csv("winequality-red.csv", sep=";")
     red_wine["red"] = 1
     white_wine = pandas.read_csv("winequality-white.csv", sep=";")
     white_wine["red"] = 0
     wines = pandas.concat([red_wine, white_wine])
     wines.head()
```

```
[6]:   fixed acidity  volatile acidity  citric acid  residual sugar  chlorides \
0             7.4                0.70         0.00             1.9         0.076
```

1	7.8	0.88	0.00	2.6	0.098
2	7.8	0.76	0.04	2.3	0.092
3	11.2	0.28	0.56	1.9	0.075
4	7.4	0.70	0.00	1.9	0.076

	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	\
0	11.0	34.0	0.9978	3.51	0.56	
1	25.0	67.0	0.9968	3.20	0.68	
2	15.0	54.0	0.9970	3.26	0.65	
3	17.0	60.0	0.9980	3.16	0.58	
4	11.0	34.0	0.9978	3.51	0.56	

	alcohol	quality	red
0	9.4	5	1
1	9.8	5	1
2	9.8	5	1
3	9.8	6	1
4	9.4	5	1

On découpe en base d'apprentissage, base de test :

```
[7]: from sklearn.model_selection import train_test_split
X = wines[[c for c in wines.columns if c != "quality"]]
Y = wines["quality"]
x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size=0.33,
↳random_state=42)
type(x_train), type(y_train)
```

```
[7]: (pandas.core.frame.DataFrame, pandas.core.series.Series)
```

```
[8]: wines.shape, x_train.shape, y_train.shape
```

```
[8]: ((6497, 13), (4352, 12), (4352,))
```

### 1.1 Exercice 1 : créer une fonction pour automatiser la création de ce graphe

```
[9]:
```

### 1.2 Exercice 2 : simplifier l'apprentissage de chaque modèle

```
[10]:
```

### 1.3 Exercice 3 : grid\_search

Considérer un modèle et estimer au mieux ses paramètres.

```
[11]:
```