

wines_knn

August 6, 2022

1 Plus proches voisins

On cherche à prédire la note d'un vin avec un modèle des plus proches voisins.

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

```
[2]: from papierstat.datasets import load_wines_dataset
df = load_wines_dataset()
df.head()
```

```
[2]:
```

	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	color
0	9.4	5	red
1	9.8	5	red
2	9.8	5	red
3	9.8	6	red
4	9.4	5	red

Les premières colonnes contiennent les composants du vins ou plutôt les résultats de mesures chimiques. Cela exclut la qualité et la couleur qui n'est pas une information numérique. On les appelle les **variables** ou **features**. C'est ce qui est connu au moment de faire une prédiction.

```
[3]: X = df.drop(['quality', 'color'], axis=1)
```

La colonne *qualité* est ce qu'on cherche à prédire.

```
[4]: y = df['quality']
```

On cale un modèle des plus proches voisins. Celui mémorise simplement tous les exemples. Le paramètre *n_neighbors* précise le nombre de voisins qui prennent pas à la prédiction. Dans notre cas, c'est juste 1.

```
[5]: from sklearn.neighbors import KNeighborsRegressor
knn = KNeighborsRegressor(n_neighbors=1)
knn.fit(X, y)
```

```
[5]: KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=1, n_neighbors=1, p=2,
weights='uniform')
```

On regarde la prédiction pour un vin quelconque tiré de la base.

```
[6]: import random
i = random.randint(0, df.shape[0]-1)
vin = X.iloc[i:i+1,:]
knn.predict(vin)
```

```
[6]: array([5.])
```

On la compare à la note enregistrée dans la base.

```
[7]: y[i]
```

```
[7]: 5
```

Tout se passe bien.

```
[8]:
```

```
[9]:
```