

ml_scikit_learn_simple

December 23, 2020

1 Rappels sur scikit-learn et le machine learning

Quelques exercices simples sur *scikit-learn*. Le notebook est long pour ceux qui débutent en machine learning et sans doute sans suspens pour ceux qui en ont déjà fait.

```
[1]: from jupyterhelper import add_notebook_menu
      add_notebook_menu()
```

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

```
[2]: %matplotlib inline
```

1.1 Des données synthétiques

On simule un jeu de données aléatoires.

```
[3]: from numpy import random
      n = 1000
      X = random.rand(n, 2)
      X[:5]
```

```
[3]: array([[0.25324685, 0.97811479],
            [0.32928095, 0.40816327],
            [0.44178633, 0.51600754],
            [0.76893618, 0.34170807],
            [0.00282938, 0.49371721]])
```

```
[4]: y = X[:, 0] * 3 - 2 * X[:, 1] ** 2 + random.rand(n)
      y[:5]
```

```
[4]: array([-0.18666718,  1.29326419,  1.64748543,  2.39341326,  0.06048883])
```

1.2 Exercice 1 : diviser en base d'apprentissage et de test

```
[5]:
```

1.3 Exercice 2 : caler une régression linéaire

Et calculer le coefficient R^2 .

```
[6]:
```

1.4 Exercice 3 : améliorer le modèle en appliquant une transformation bien choisie

[7] :

1.5 Exercice 4 : caler une forêt aléatoire

[8] :

1.6 Exercice 5 : un peu de math

Comparer les deux modèles sur les données suivantes ? Que remarquez-vous ? Expliquez pourquoi ?

[9] :

```
X_test2 = random.rand(n, 2) + 0.5
y_test2 = X_test2[:, 0] * 3 - 2 * X_test2[:, 1] ** 2 + random.rand(n)
```

1.7 Exercice 6 : faire un graphe avec...

Le nuage de points du premier et second jeu, les prédictions des deux modèles, une légende, un titre... avec [pandas](#) ou directement avec [matplotlib](#) au choix.

[10] :

1.8 Exercice 7 : illustrer l'overfitting avec un arbre de décision

Sur le premier jeu de données.

[11] :

1.9 Exercice 8 : augmenter le nombre de features et régulariser une régression logistique

L'objectif est de regarder l'impact de la régularisation des coefficients d'une régression logistique lorsque le nombre de features augmentent.

[12] :