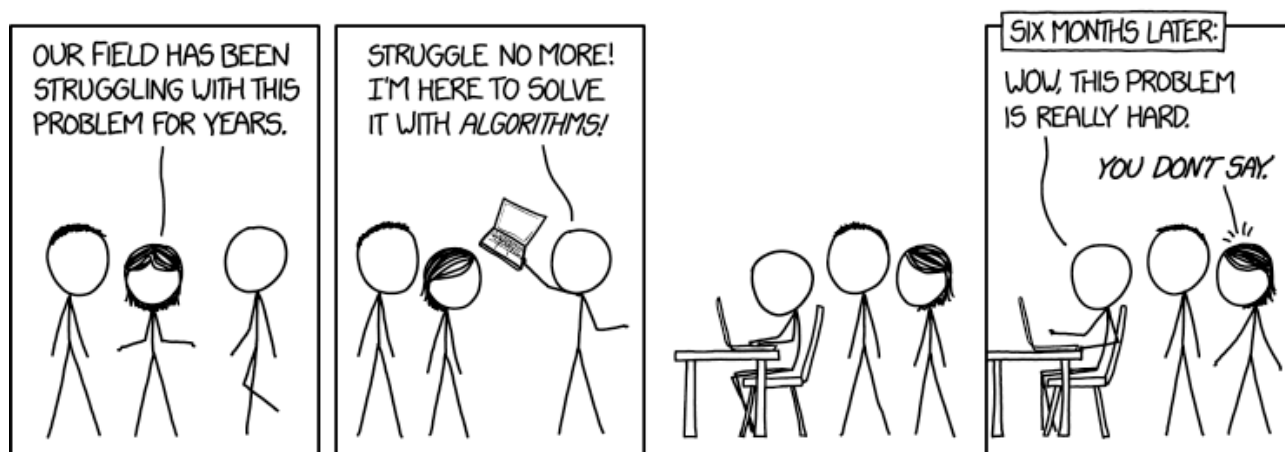


Algorithmes de Machine Learning

Partie I : Courbes d'apprentissage

Data Scientist - 11/07/2025



Algorithms are hard

Source: [xkcd](https://xkcd.com/1595/)

Etape 1

Algorithmes ML et courbes d'apprentissage

Objectifs de l'activité

- Être capable d'expliquer le fonctionnement des algorithmes suivants : l'Arbre de Décision, la Forêt Aléatoire, les Machines à Vecteurs de Support (SVM)
- Connaître et comprendre les hyperparamètres qui existent pour "régler" ces modèles
- Tracer des courbes d'apprentissage pour ces différents algorithmes de ML et les interpréter pour expliquer comment le choix du type de modèle et des hyperparamètres affecte le **sur-ajustement** et le **sous-ajustement**

Compétences

- Expliquer ce qu'est un arbre de décision, une forêt aléatoire, un SVM
- Tracer et interpréter les courbes d'apprentissage utilisant ces différents modèles

Consignes

- Prenez le temps de relire les sections du livre pour l'arbre de décision (pages 175-180), la forêt aléatoire (ch. Random Forests) et les algorithmes SVM (pages 153-156).
- Lisez les Ressources 1 et 2 sur les courbes d'apprentissage.
- Implémentez le code donné dans le livre p. 132 sur le dataset iris.
- Importez ces algorithmes à partir de scikit-learn et mettez en œuvre le code du livre sur les courbes d'apprentissage (en utilisant les mêmes données). Assurez-vous de bien comprendre ce que fait le code. Parlez-en à vos voisins ! Essayez d'adapter le code dans la Ressource 1 pour l'appliquer au même dataset que dans le livre.
- Choisissez un algorithme pour lequel vous allez ensuite produire les courbes d'apprentissage. Vous pouvez essayer le jeu de données des arbres de Grenoble depuis ML1 ou le dataset sur les maladies cardiaques que vous avez vues en ML2.1.
- Créez un modèle ML qui montre des signes de sous-ajustement et un autre qui montre des signes de sur-ajustement. Essayez de modifier les hyperparamètres du modèle et voyez comment cela affecte la courbe d'apprentissage. Expliquez comment le choix du type de modèle et des hyperparamètres affecte le sur-ajustement et le sous-ajustement. Parlez-en à vos voisins !

Livrables

- Script / notebook python qui contient :
 - Un exemple de courbe d'apprentissage d'un modèle qui est « underfitted »
 - Un exemple de courbe d'apprentissage d'un modèle qui est « overfitted »

Pour aller plus loin

- Essayez d'implémenter des courbes d'apprentissage sur un problème de classification que vous avez déjà rencontré dans les modules précédents.
- Essayez de produire les courbes d'apprentissage pour les autres algorithmes de ML.
- Essayez de produire des courbes de validation comme montré dans la ressource Courbes de validation 1.

Ressources

General :

Hands on Machine Learning with Scikit-Learn and Tensorflow, Chapitres 4, 5, 6, 7.

Courbes d'apprentissage :

1. https://scikit-learn.org/stable/auto_examples/model_selection/plot_learning_curve.html
2. <https://towardsdatascience.com/why-you-should-be-plotting-learning-curves-in-your-next-machine-learning-project-221bae60c53>

Courbes de validation :

3. https://scikit-learn.org/stable/modules/learning_curve.html

Ressources supplémentaires :

Arbres de décision:

<https://scikit-learn.org/stable/modules/tree.html>

<https://www.lovelyanalytics.com/2016/08/16/decision-tree-comment-ca-marche/>
<https://www.youtube.com/watch?v=ZVR2Way4nwQ>
<https://www.youtube.com/watch?v=sgQAhG5Q7iY>

Random forest:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=random%20forest#sklearn.ensemble.RandomForestClassifier>
<https://towardsdatascience.com/random-forest-in-python-24d0893d51c0>
<https://www.youtube.com/watch?v=v6VJ2RO66Ag>

SVM:

<https://scikit-learn.org/stable/modules/svm.html>
<https://towardsdatascience.com/a-brief-introduction-to-support-vector-machine-adf0f103a80f>
<https://www.youtube.com/watch?v=YPSrcrkx28>
<https://www.youtube.com/watch?v=Q7vT0--5VII>