

Vérification ponctuelle des algorithmes de CLASSIFICATION + sélection de modèles

Atelier #10





Qu'est-ce que la vérification ponctuelle (*"spot-check"*)?

Définition de vérification ponctuelle (*"spot-check"*)

C'est une méthode qui compare plusieurs algorithmes entre eux dans le but d'identifier celui ou ceux qui performant le mieux pour la problématique d'apprentissage automatique à solutionner et les données disponibles.



Pourquoi faire une vérification ponctuelle?

- Très difficile de savoir à l'avance quel algorithme est le meilleur pour un problème d'apprentissage automatique.
- Pour se concentrer dès le début sur un ou quelques algorithmes qui sont les plus prometteurs.

Algorithmes de classification

Terminologie		Scikit-learn		
Français	Anglais	Module	Documentation	Algorithme
Régression Logistique	<i>Logistic Regression</i>	linear_model	https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression	<i>LogisticRegression</i>
Analyse Discriminante Linéaire (ADL)	<i>Linear Discriminant Analysis (LDA)</i>	discriminant_analysis	https://scikit-learn.org/stable/modules/lda_qda.html#linear-and-quadratic-discriminant-analysis	<i>LinearDiscriminantAnalysis</i>
K plus proches voisins	<i>k-Nearest Neighbors (KNN)</i>	neighbors	https://scikit-learn.org/stable/modules/neighbors.html#nearest-neighbors-classification	<i>KNeighborsClassifier</i>
Classification Naïve Bayesienne	<i>Naive Bayes Classifier</i>	naive_bayes	https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes	<i>GaussianNB</i>
Arbres de décision	<i>Decision Trees</i>	tree	https://scikit-learn.org/stable/modules/tree.html#classification	<i>DecisionTreeClassifier</i>
Machines à vecteurs de support (MVS)	<i>Support Vector Machine (SVM)</i>	svm	https://scikit-learn.org/stable/modules/svm.html#classification	<i>SVC</i>



Sélection de modèles

Définition : Comparer plusieurs modèles entres eux en utilisant un métrique commun dans le but de sélectionner le ou les meilleur(s) modèle(s).

Étapes :

1. Définir les algorithmes à tester.
2. Sélectionner le métrique approprié à la problématique.
3. Faire rouler les modèles de base, sans hyperparamétrisation, tous ensemble.
4. Classer les modèles du meilleur au pire.
5. Sélectionner le modèle le plus performant.
 - Si plusieurs modèles ont une performance équivalente, voir la leçon 13.

**NE VOUS EN FAITES PAS SI
VOUS NE COMPRENEZ PAS.
- ANDREW NG**



Note

Certains algorithmes seront plus approfondis, mais on ne va pas approfondir le fonctionnement de tous les algorithmes qui seront vu dans le cadre théorique de ce cours.

Données au format LARGE vs LONG

Large

	A	B	C
0	1	2	3
1	4	5	6
2	7	8	9

Long

	variable	value
0	A	1
1	A	4
2	A	7
3	B	2
4	B	5
5	B	8
6	C	3
7	C	6
8	C	9

Préférable