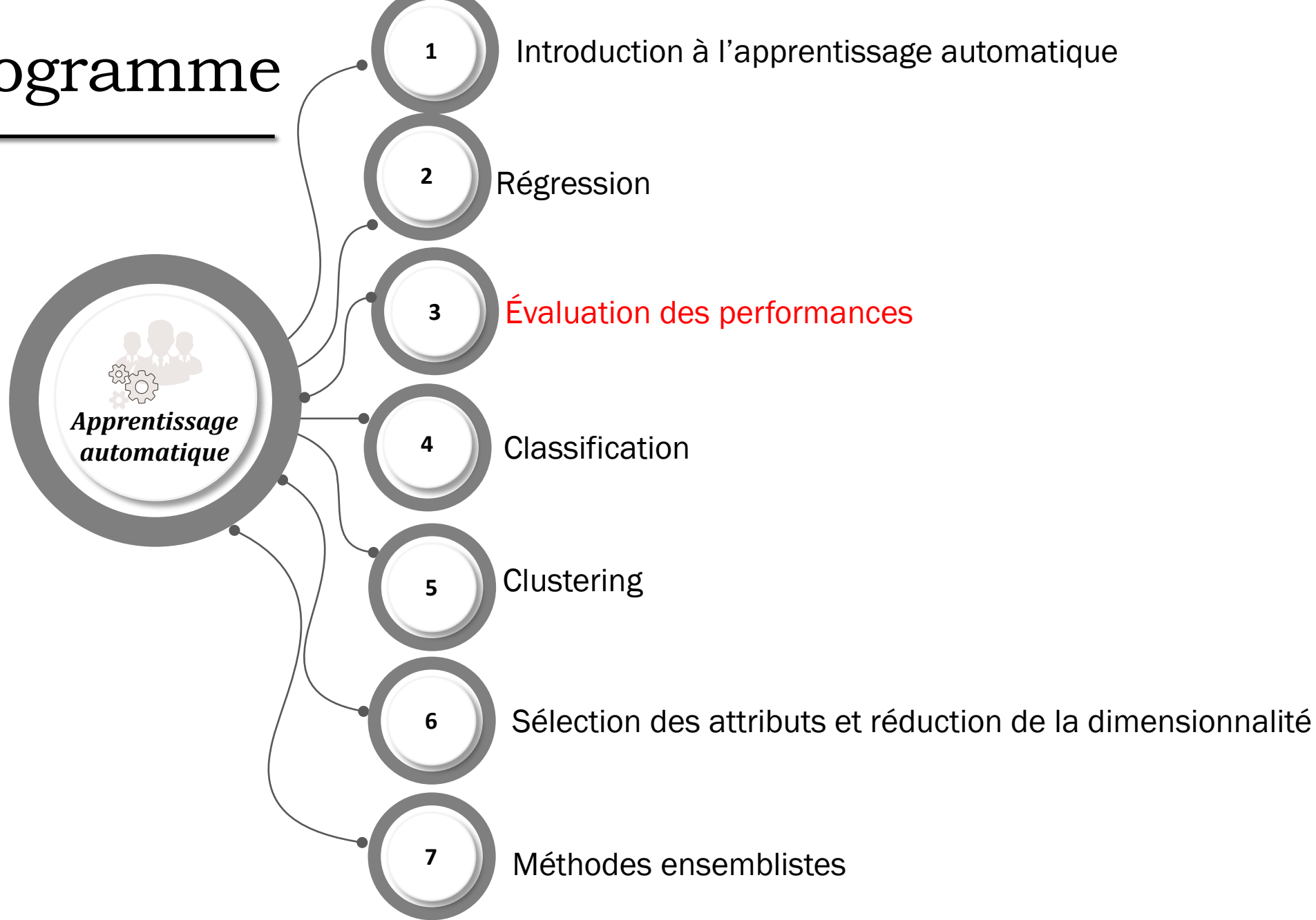


Apprentissage Automatique

*Intelligence Artificielle et
Sciences de Données
(IASD)*

DR N. DIF

Programme



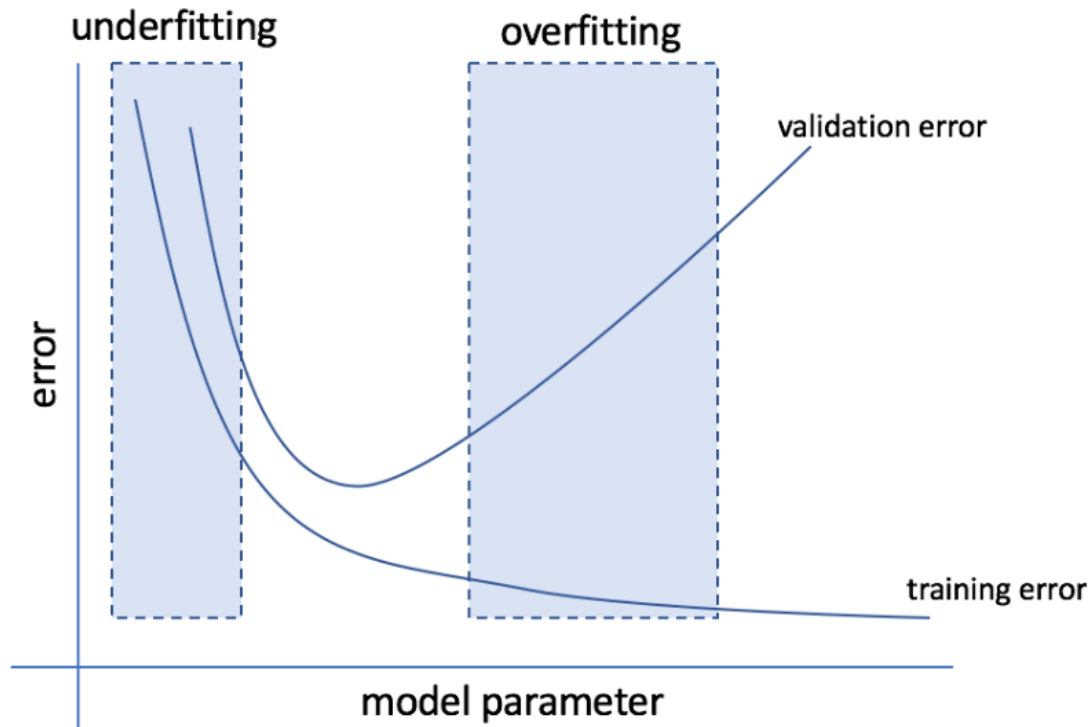
LES MESURES D'ÉVALUATION

- Les mesures d'évaluation sont utilisées pour mesurer la performance du modèle sur la base de teste.
- Il existe deux types de mesures d'évaluation : des mesures à maximiser comme l'accuracy et des mesures à minimiser comme l'erreur. Le choix de la mesure approprié dépend du domaine d'application et de la nature des données.
- $\text{Accuracy} = \frac{\text{Nombre d'instance bien classées}}{\text{Nombre total d'instances}}$, $\text{Error} = \frac{\text{Nombre d'instance mal classées}}{\text{Nombre total d'instances}}$

LES MÉTHODES D'ÉVALUATION

- Le processus de data mining est divisé en trois étapes principales : prétraitement des données, apprentissage, et **évaluation des performances**.
- La seule méthode exploitée en data mining **pour comparer les modèles** générés de l'apprentissage c'est de **mesurer leurs performances**.
- L'étape d'évaluation est cruciale dans le développement des modèles prédictifs en apprentissage automatique, car, **le but principal de l'apprentissage ce n'est pas de créer seulement des modèles, mais plutôt des modèles performants** caractérisés par des résultats prometteurs en prédiction.

2. APPRENTISSAGE SUPERVISÉ



Surajustement : Le modèle est performant sur la base d'apprentissage et très mauvais en généralisation sur la base de test. Cela engendre une grande marge entre les performances sur la base d'apprentissage et de test.

Sous-ajustement : La performance du modèle est mauvaise sur les deux bases : d'apprentissage et de test.

From : <https://www.jeremyjordan.me/evaluating-a-machine-learning-model/>

2. APPRENTISSAGE SUPERVISÉ

- Une bonne dataset est caractérisée par un **nombre intéressant d'exemples**, où un maximum de cas est pris en considération en respectant plusieurs critères : **généralisation, équilibrage des données...**
- Dans les problèmes du monde réel, il est impossible **de considérer tous les scénarios possibles**. La solution à ce problème, c'est **d'assurer la généralisation du modèle** créé sur un maximum d'exemple.
- Un modèle performant n'est pas seulement caractérisé par une bonne performance, mais plutôt caractérisé par **une bonne généralisation**, et qui a prouvé son efficacité sur des données non vue durant l'apprentissage.
- Le **surapprentissage** est le contraire de la généralisation, il est défini par l'ensemble des modèles caractérisés par une **grande variance** entre les performances sur la base d'apprentissage et les instances non vu durant l'apprentissage.

2. APPRENTISSAGE SUPERVISÉ

- Il est important de **mesurer la performance du modèle sur des données non vues durant l'apprentissage** pour éviter le fameux problème de **surapprentissage** et d'assurer la généralisation.
- Les **méthodes d'évaluation** sont parmi les techniques utilisées pour éviter ce problème. En général, la stratégie consiste à **diviser la dataset en base d'apprentissage et de test**.
- La base d'apprentissage est utilisée pour évaluer le modèle et la base de test pour tester sa performance.
- Dans certaines situations, la dataset est divisée en 3 sets : **base d'apprentissage, base de validation, et base de test**. La base d'évaluation est utilisée durant l'apprentissage pour fixer la valeur de quelques hyperparamètres.

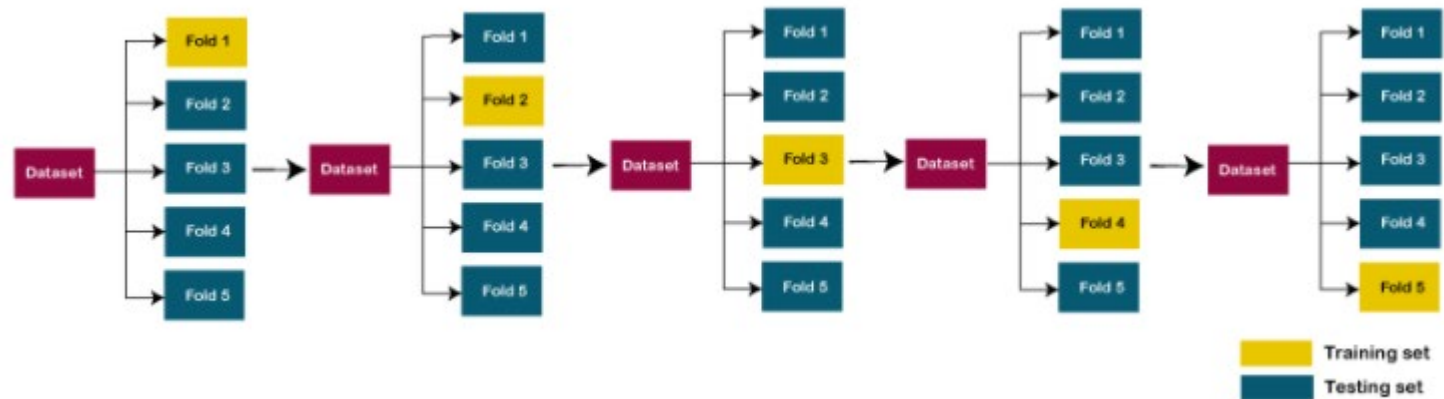
2. APPRENTISSAGE SUPERVISÉ

-
- **Hold out:** dans cette approche, la dataset est divisée aléatoirement en base d'apprentissage et de test (ou de validation), généralement $2/3$ de la dataset est utilisée pour l'apprentissage, et le $1/3$ restant pour le test. Cette approche peut poser un **problème pour les datasets déséquilibrés** (classe non homogène) ou les datasets caractérisés par des attributs associés à des valeurs prédominantes.
 - **Stratified Hold out:** le même principe de hold out, mais, elle **résout le problème des bases déséquilibrées**, où, elle propose de diviser équitablement les classes ou les valeurs d'attributs sur les bases d'apprentissage et de teste. L'inconvénient de cette approche est le nombre d'instances considérées lors de l'apprentissage et de test, car, un algorithme qui prend en considération un maximum d'exemples (évite la perte d'information) lors de l'apprentissage génère un modèle caractérisé par une bonne généralisation, et pour tester la fiabilité du modèle généré, il est conseiller de le tester sur un maximum d'exemples.

2. APPRENTISSAGE SUPERVISÉ

2.1. Les méthodes d'évaluation

- **K-Cross-Validation** : cette méthode d'évaluation résout le problème de la stratified hold out cité auparavant. Elle propose de diviser la dataset en K folds, ensuite k-1 fold sont utilisés pour l'apprentissage et 1 fold pour le test (ou l'évaluation). Le processus est répété K fois, ensuite, la performance est calculée en fonction de la moyenne des performances de test sur chaque fold . Généralement, $k = 10$.



2. APPRENTISSAGE SUPERVISÉ

2.1. Les méthodes d'évaluation

- **Stratified K-cross-validation:** cette technique est similaire à la K-cross-validation, sauf qu'elle exploite le principe de **réorganisation équitable**. Il s'agit d'un processus de réorganisation des données pour s'assurer que chaque groupe est un bon représentant de l'ensemble de données complet. C'est l'une des meilleures approches pour traiter le biais et la variance,
- **Leave-one-out :** utilise le même principe de la K-cross-validation avec $K = N - 1$, où, N est le nombre d'instances dans la dataset. Cette technique utilise seulement une seule instance pour le test dans chaque itération, et le nombre itération est équivalent au nombre d'instances N dans la dataset.