

Intelligence Artificielle 3

Cours 4 - Modèles ensemblistes

Steve Lévesque, Tous droits réservés © où applicables

- 1 Qu'est-ce que le modèle Ensembliste (Ensemble)
- 2 Types de modèles
 - Méthodes Ensemblistes (avec d'autres modèles traditionnels)
 - Voting
 - Stacking
 - Bagging
 - Boosting
 - Algorithmes basées sur Bagging ou Boosting (tout-en-un)

Qu'est-ce que le modèle Ensembliste (Ensemble)

Une méthode d'ensemble est une technique puissante qui permet de combiner plusieurs modèles d'apprentissage automatique en un seul modèle, fonctionnant souvent mieux que n'importe lequel des modèles individuels laissés à eux-mêmes.

NB : Les modèle Ensemblistes ne garantissent pas une hausse de performance.

Qu'est-ce que le modèle Ensembliste (Ensemble)

Exemple avec la méthode par vote majoritaire.

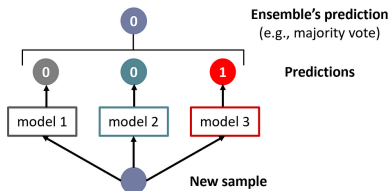


Figure: <https://pub.towardsai.net/introduction-to-ensemble-methods-226a5a421687>

Méthodes Ensemblistes (avec d'autres modèles traditionnels)

- Voting : Avoir plusieurs modèles et passer la décision au vote, comme un conseil d'administration ou un comité
- Stacking : Construction d'un nouveau modèle à partir des prédictions des autres modèles, maintenant considérées comme les nouveaux "features" du nouveau modèle empilé
- Bagging : Même principe que "Voting", mais créer des sous-ensembles de données distincts pour chaque modèle, en plus de garder la grandeur (nombre d'entrées) originale du jeu de donnée original
- Boosting : Comme "Bagging", mais en plus corrige les erreurs (mauvaises prédictions) de manière séquentielle pour que chaque modèle améliore réellement la performance de l'ensemble

Voting

Le but conceptuel du “Voting” est de combiner plusieurs modèles différents, mais d'attribuer une caractéristique de vote pour décider de la décision finale.

Il y a 3 types de votes principaux :

- Max (Majority/Hard Voting)
- Moyenne (Soft Voting)
- Weighted (poids)

Voting - Max (Majority/Hard Voting)

Prends la valeur la plus votée ($50\% + 1$ normalement) comme valeur de vérité.

Par exemple : 3 modèles sur 5 vote pour la valeur 0 et les 2 autres modèles pour 1. La valeur votée est 0.

Voting - Moyenne (Soft Voting)

Une moyenne d'un résultat est faite pour généraliser la réponse.

Attention : ne peut pas être catégorique (classification).

Voting - Weighted (poids)

Un vote peut valoir plus qu'un vote normal.

Par exemple : un modèle très performant pourrait valoir 2 votes, puis les autres modèles 1 seul.

Stacking

Contrairement au Voting, Stacking utilise plusieurs modèles et passe leurs résultats à un modèle final qui permet de faire la prédiction finale à partir des prédictions des modèles précédents.

Stacking

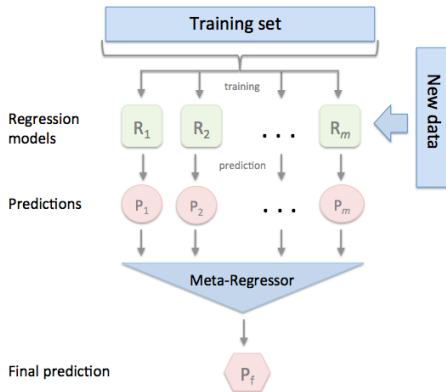


Figure: https://rasbt.github.io/mlxtend/user_guide/regressor/StackingRegressor/

Bagging

Bagging permet d'augmenter la confiance du modèle en offrant une variance et un biais plus basse.

Pour y arriver, plusieurs instances (modèles) sont faites d'un modèle de choix, les données sont séparées en sous-ensembles différents (bootstrapping) et le résultat est généralisé.

Il ne garanti pas une meilleure performance sur nos données locales, mais une variance et un biais bas est un atout important pour lorsque le modèle sera en production.

Bagging

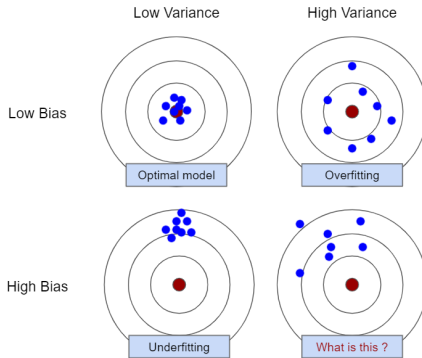


Figure: <https://towardsdatascience.com/what-bias-variance-bulls-eye-diagram-really-represent-ff6fb9670993>

Bagging



Bagging Classifier Process Flow

Figure: <https://medium.com/ml-research-lab/bagging-ensemble-meta-algorithm-for-reducing-variance-c98ffa5489f>

Boosting

Boosting permet de prendre en considération un défaut de Bagging, le fait que si une prédiction est toujours fausse, il n'y avait pas d'ajustements faits pour essayer de bien prédire celle-ci.

Maintenant, le modèle actuel est dépendant du modèle précédent, et ainsi de suite.

Un poids est attribué à chaque entrée de donnée. En fonction des erreurs, les poids sont ajustés pour permettre aux nouvelles itérations d'essayer de corriger les erreurs.

Boosting

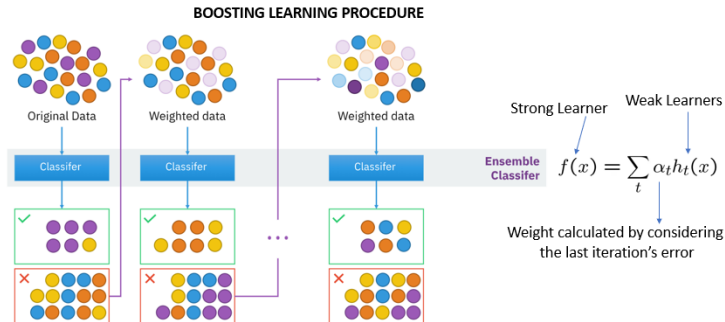


Figure: <https://medium.com/analytics-vidhya/ensemble-methods-in-machine-learning-31084c3740be>

Boosting

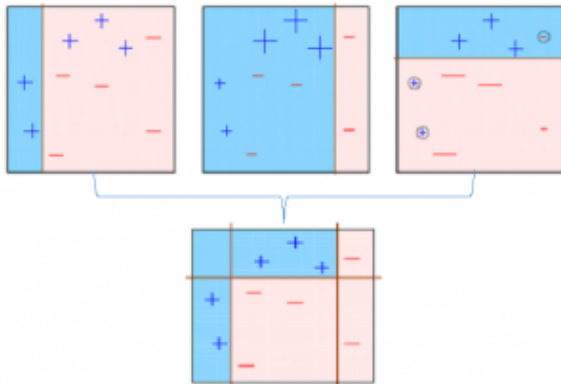


Figure: <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>

Algorithmes basées sur Bagging ou Boosting (tout-en-un)

- Bagging : Bagging meta-estimator, Random forest
- Boosting : AdaBoost, GBM, XGBM, Light GBM, CatBoost

Bibliographie

- <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>
- <https://pub.towardsai.net/introduction-to-ensemble-methods-226a5a421687>
- https://rasbt.github.io/mlxtend/user_guide/regressor/StackingRegressor/
- <https://towardsdatascience.com/what-bias-variance-bulls-eye-diagram-really-represent->
- <https://medium.com/ml-research-lab/bagging-ensemble-meta-algorithm-for-reducing-variance->
- <https://medium.com/analytics-vidhya/ensemble-methods-in-machine-learning-31084c3740be>
- <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>