

L'ensemble learning

1





La variabilité inter et intra-opérateur

- ▶ La variabilité intra-opérateur est la variabilité des résultats que l'on obtient en comparant les résultats d'un professionnel sur le même problème à intervalle de temps différent.
- ▶ La variabilité inter-opérateur est la variabilité des résultats que l'on obtient en comparant les résultats de plusieurs professionnels.



Segmentation de tumeur

- ▶ Dans cette article de recherche paru dans Elsevier en 2004, ils montrent que lors de la segmentation manuel de tumeur par des radiologues expert la variabilité intra-opérateur était de 20% +/- 15% et de 28% +/- 12% pour la variabilité inter-opérateur.
- ▶ Les résultats d'un expert n'est donc pas toujours le plus fiable.



La sagesse de la foule

- ▶ C'est une théorie qui présuppose que la perception et la résolution d'un problème sont plus efficaces par une foule que par n'importe quel individu en faisant partie ou non.
- ▶ Selon ce concept, une foule d'amateur peut mieux répondre à un problème d'un seul expert.
- ▶ La foule cependant doit répondre à trois hypothèses pour valider cette théorie :
 - ▶ la diversité : avoir des personnes de divers milieux avec des idées originales ;
 - ▶ l'indépendance : permettre à ces avis divers de s'exprimer sans aucune influence ;
 - ▶ la décentralisation : laisser ces différents jugements s'additionner plutôt que de laisser une autorité supérieure choisir les idées qu'elle préfère.



L'ensemble learning

- ▶ De ce concept philosophique on en a retiré une application mathématique l'ensemble learning.
- ▶ Les méthodes d'ensemble learning utilisent plusieurs algorithmes d'apprentissage et prennent en compte les résultats de ces modèles afin d'obtenir de meilleures performances prédictives que les modèles pris séparément.

Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
1	1	1	0	0	1
2	1	1	0	1	0
3	0	0	1	1	1
4	0	1	1	1	0
5	1	0	1	0	1



Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
1	1	1	0	0	1
2	1	1	0	1	0
3	0	0	1	1	1
4	0	1	1	1	0
5	1	0	1	0	1



Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
1	1	1	0	0	1
2	1	1	0	1	0
3	0	0	1	1	1
4	0	1	1	1	0
5	1	0	1	0	1



Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5
1	1	1	0	0	1
2	1	1	0	1	0
3	0	0	1	1	1
4	0	1	1	1	0
5	1	0	1	0	1
Perf	60%	60%	60%	60%	60%



Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Mélange
1	1	1	0	0	1	1
2	1	1	0	1	0	1
3	0	0	1	1	1	1
4	0	1	1	1	0	1
5	1	0	1	0	1	1
Perf	60%	60%	60%	60%	60%	100%

Exemple

Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Mélange
1	1	1	0	0	1	1
2	1	1	0	1	0	1
3	0	0	1	1	1	1
4	0	1	1	1	0	1
5	1	0	1	0	1	1
Perf	60%	60%	60%	60%	60%	100%

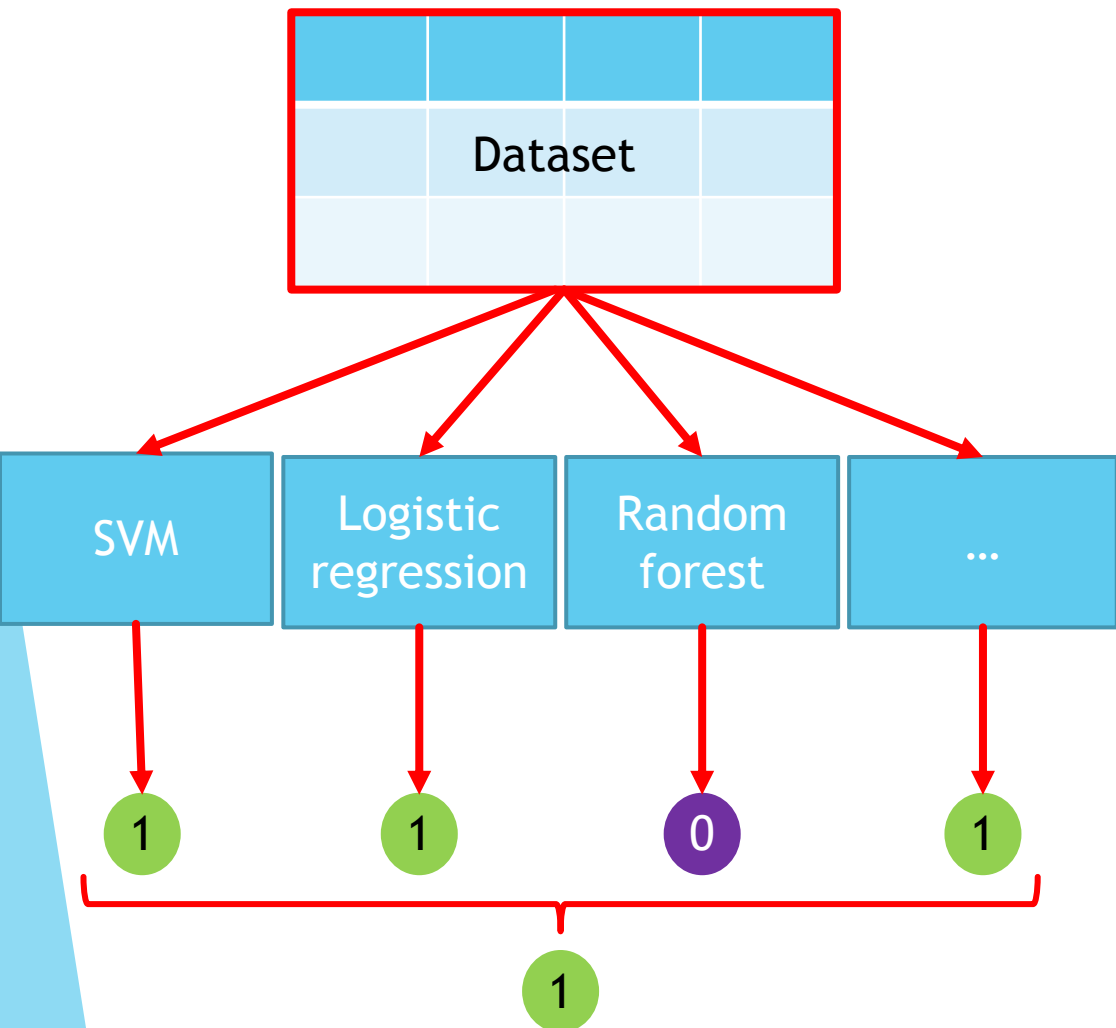
L'importance de la diversité

- ▶ Comme on l'a vu avec la sagesse des foules, la diversité est une notion importante. Les modèles doivent être différents, construits sur des données différents ou avec des algorithmes d'apprentissage différents.

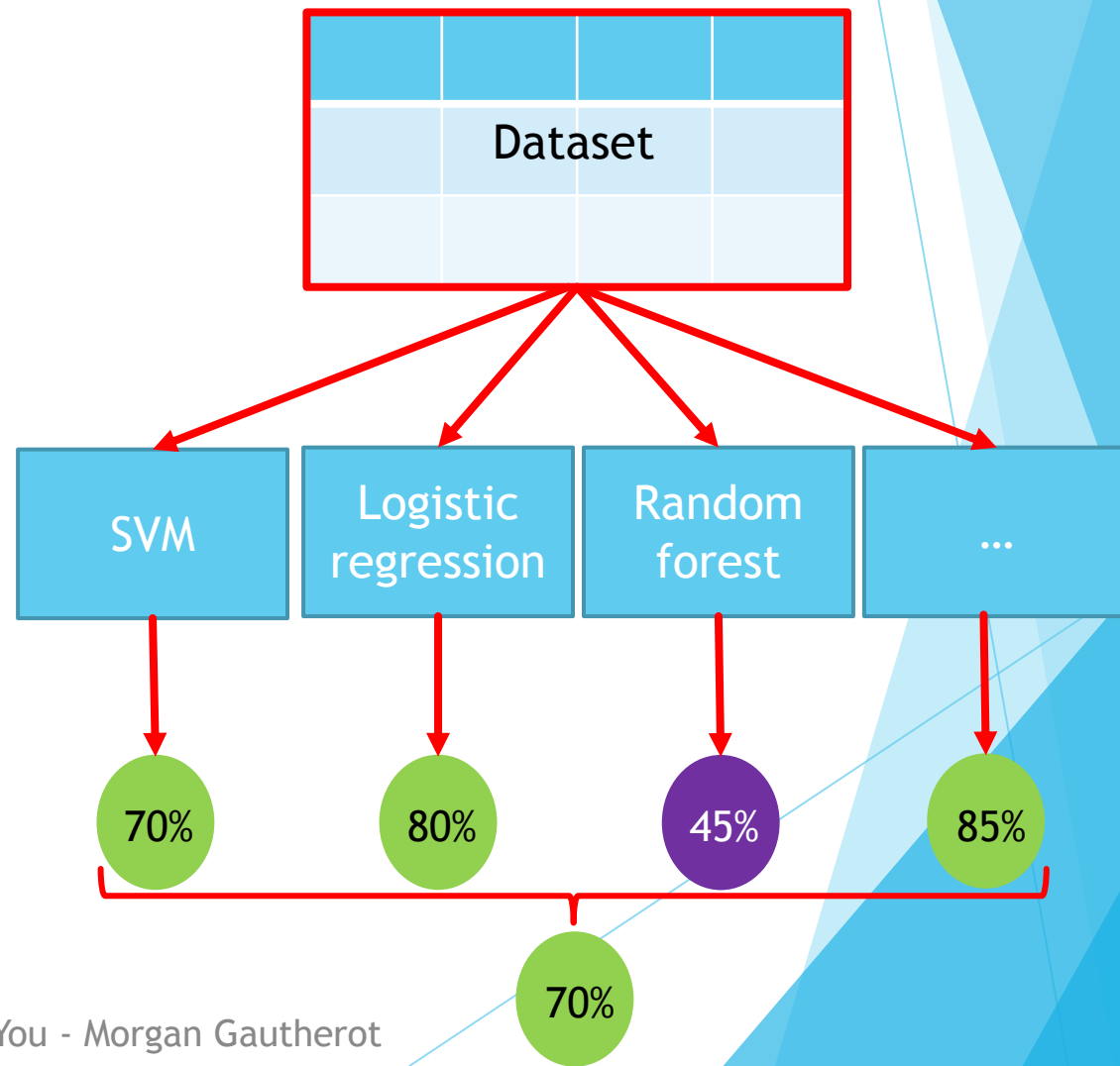
Index	Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Mélange
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	1	1	1	1	1	1
Perf	60%	60%	60%	60%	60%	60%

Voting

Hard Voting



Soft Voting





Utilisation de ce concept

- ▶ Bagging/Pasting (Random forest, bagging de regression)
- ▶ Boosting (AdaBoost, Gradient Boosting)
- ▶ Stacking (Utilisation d'un modèle afin de mélanger les prédictions des modèles)