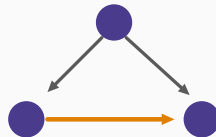


Inférence causale et machine learning

Chapitre 11 : Effets hétérogènes et forêts causales



Pourquoi la moyenne ne suffit pas

La difficulté

La forêt causale

L'usage, et ses garde-fous

Ce qu'il faut retenir

Pourquoi la moyenne ne suffit pas

Un effet moyen peut cacher l'essentiel

L'effet conditionnel

$$\tau(x) = E[Y(1) - Y(0) \mid X = x]$$

C'est l'effet moyen **pour un profil donné**. L'ATE en est la moyenne sur toute la population.

Un effet moyen nul, et deux effets réels

Supposons que le programme de formation :

- améliore l'emploi de **quinze points** chez les jeunes ;
- le dégrade de **quinze points** chez les seniors, en les détournant d'une recherche directe.

L'effet moyen est **nul**. Conclure que le programme ne sert à rien serait une erreur coûteuse : il faut le maintenir pour les uns et l'arrêter pour les autres.

Application en gestion

La question du décideur n'est presque jamais « faut-il ce programme ? » mais « **pour qui ?** ».

Les budgets sont limités; cibler ceux qui en bénéficient le plus multiplie l'effet total à dépense constante. **C'est ici que le machine learning apporte le plus à l'inférence causale** — bien plus qu'en estimant mieux une moyenne.

La difficulté

On n'observe jamais l'effet

Pourquoi ce n'est pas une régression ordinaire

Pour prévoir Y , on dispose de Y observé et l'on minimise une erreur.

Pour estimer $\tau(x)$, on ne dispose de **rien** : l'effet individuel n'est jamais observé — c'est le problème fondamental du chapitre 2.

Il n'existe donc **aucune erreur à minimiser directement**. Les méthodes du cours précédent ne se transposent pas telles quelles.

La solution : construire une cible

On construit une variable dont l'**espérance conditionnelle** est l'effet cherché, puis on la prédit.

Le score doublement robuste joue ce rôle : il combine le modèle de résultat et le score de propension, de sorte qu'il suffit que **l'un des deux** soit correct.

La forêt causale

Un arbre qui cherche l'hétérogénéité

Le changement de critère

Un arbre ordinaire découpe pour rendre les feuilles **homogènes en Y** — c'est le Gini du chapitre 8 d'*Apprentissage statistique*.

Un **arbre causal** découpe pour que les feuilles diffèrent le plus possible **en effet de traitement**. Il cherche là où l'effet change, non là où le résultat change.

L'échantillonnage honnête

Chaque arbre partage ses données en deux :

Moitié	Usage
Première	déterminer la structure des découpages
Seconde	estimer l'effet dans chaque feuille

Intuition

Sans lui, l'arbre découpe là où l'effet **paraît** grand par hasard, puis estime l'effet sur ces mêmes données : il retrouve mécaniquement un effet grand.

C'est le sur-ajustement de la sélection, déjà rencontré au chapitre 5 d'*Apprentissage statistique*. Ici, il produit des **effets hétérogènes fictifs** — des sous-groupes qui n'existent pas.

Le partage honnête est ce qui rend les intervalles de confiance valides. C'est la même discipline que le fractionnement croisé du chapitre 10.

Définition

On agrège de nombreux arbres causaux, comme au chapitre 9 d'*Apprentissage statistique*.

On obtient $\hat{\tau}(x)$ pour chaque profil, **avec** des intervalles de confiance — résultat théorique important, car les forêts ordinaires n'en fournissent pas.

L'usage, et ses garde-fous

Trois usages

1. **Cibler** — concentrer le programme sur les profils à fort effet.
2. **Comprendre** — identifier les caractéristiques associées à l'hétérogénéité.
3. **Tester** — l'effet varie-t-il réellement, ou l'apparente hétérogénéité est-elle du bruit?

Le danger du sous-groupe trouvé après coup

Chercher parmi cent sous-groupes en fera apparaître plusieurs à effet « significatif » par pur hasard.

C'est le problème des comparaisons multiples, et il a produit une abondante littérature d'effets non reproductibles.

Trois protections : **préenregistrer** les sous-groupes d'intérêt, **corriger** pour la multiplicité, et **valider** l'hétérogénéité sur un échantillon réservé.

Hétérogénéité n'est pas modération causale

Constater que l'effet est plus fort chez les jeunes ne signifie pas que **l'âge** cause cette différence.

L'âge peut être corrélé à ce qui compte vraiment — le niveau de formation initiale, l'ancienneté du chômage, le secteur visé.

Ce qu'il faut retenir

Cinq points

1. $\tau(x)$ est l'effet **conditionnel**; un ATE nul peut cacher deux effets opposés.
2. L'effet n'étant jamais observé, on **construit une cible** — le score doublement robuste.
3. L'**arbre causal** découpe selon l'hétérogénéité de l'effet, non du résultat.
4. L'**échantillonnage honnête** — structure et estimation séparées — rend l'inférence valide.
5. On identifie **où** l'effet diffère, pas **pourquoi**; et l'identification reste requise.

La suite

Reste à passer de l'effet estimé à la décision — ce qui suppose de traiter l'incertitude, les coûts, et les limites de la transposition.