

Bagging

Véronique Tremblay

Objectifs

- Comprendre le fonctionnement du bagging
- Connaître et comprendre certaines propriétés théoriques de ces modèles.

Origine

Bootstrap aggregation

(voir Brieman 1996 et ESL section 8.7)

L'algorithme

1 Pour b de 1 à B

a. On sélectionne avec remise un échantillon des données de taille c

$$Z^b$$

b. On construit un modèle prédictif (un arbre) et on calcule une prévision pour chaque observation

$$\hat{f}^b(x) = \sum_{m=1}^M \hat{c}_m \mathbf{I}(x \in R_m)$$

2 On agrège les valeurs des B arbres¹

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x)$$

¹Si Y est nominal, on peut aussi utiliser un vote. On choisit la majorité.

Bagging

└ Pourquoi ça fonctionne?

Pourquoi ça fonctionne?

Rappel: nous voulons réduire l'erreur

$$E[(Y - \hat{f}(x))^2] = \sigma_\epsilon^2 + \text{Var}[\hat{f}(x)] + \text{Biais}[\hat{f}(x)]^2$$

Posons

$$\hat{Y} = \hat{f}^b(x)$$

et

$$\text{Var}[\hat{f}^b(x)] = \text{Var}[\hat{Y}] = \sigma_b^2$$

Quelle est la $\text{Var}[\hat{f}_{bag}(x)]$?

$$\begin{aligned} \text{Var}(\bar{\hat{Y}}_i) &= \text{Var}\left(\frac{\sum_{i=1}^B \hat{Y}_i}{B}\right) \\ &= \frac{1}{B^2} \text{Var}\left(\sum_{i=1}^B \hat{Y}_i\right) \\ &= \frac{1}{B^2} \left(\sum_{i=1}^B \text{Var}(\hat{Y}_i) + \sum_{i \neq j} \text{Cov}(\hat{Y}_i, \hat{Y}_j) \right) \\ &= \frac{1}{B^2} \left(B\sigma_b^2 + \sum_{i \neq j} \text{Cov}(\hat{Y}_i, \hat{Y}_j) \right) \\ &= \frac{1}{B^2} (B\sigma_b^2 + (B^2 - B)\rho\sigma_b^2) \\ &= \rho\sigma_b^2 + \frac{1-\rho}{B}\sigma_b^2 \end{aligned}$$

On obtient

$$\text{Var}[\hat{f}_{bag}(x)] = \rho \sigma_b^2 + \frac{1 - \rho}{B} \sigma_b^2$$

où ρ est la corrélation entre les $\hat{f}^b(x)$.

Caractéristiques

Avantages

- Plus stable

Avantages

- Plus stable
- Meilleures prévisions (surtout pour les modèles très instables comme les arbres et les RN, peut réduire la performance d'estimateurs plus stables comme les KNN)

Avantages

- Plus stable
- Meilleures prévisions (surtout pour les modèles très instables comme les arbres et les RN, peut réduire la performance d'estimateurs plus stables comme les KNN)
- Échantillon OOB (lire la section 15.3.1)

Inconvénients

- Deux hyperparamètres de plus à choisir (B et c)
- Long si B et c sont grands!

En pratique

En R, vous pouvez utiliser la fonction `bagging` de la librairie `adabag`, mais...