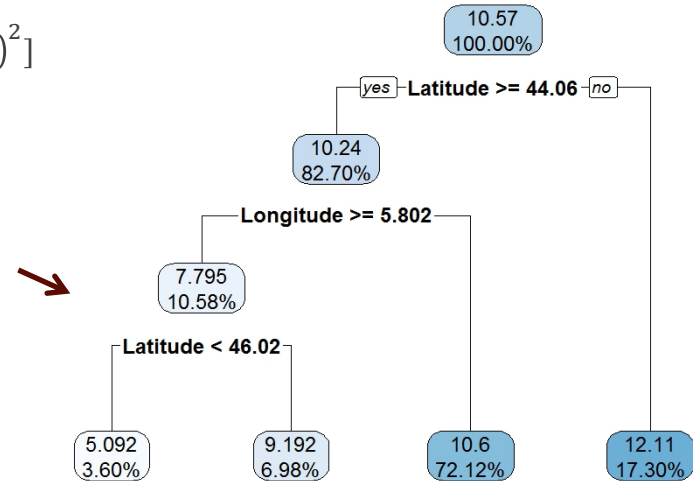


CART (CLASSIFICATION AND REGRESSION TREE) – PRINCIPE DU MODÈLE

- Principe du modèle CART

- On cherche la fonction f^{**} telle que $f^{**} = \underset{f}{\operatorname{argmin}} E[(Y - f(X))^2]$
- Comme $E[(Y - f(X))^2]$ est inconnue, en pratique on cherche $f^* = \underset{f}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N (y_i - f(x_i))^2 \right]$
- On construit un estimateur \tilde{f} de f^* qui prend la forme d'un arbre :



- Comparaison GLM

- Dans le GLM, on suppose $g(E[Y|X]) = {}^t\beta \cdot X$ et le problème est de trouver $\beta^* = \underset{\beta}{\operatorname{argmin}} E[(Y - g^{-1}({}^t\beta \cdot X))^2]$
- Le problème à résoudre dans le GLM est donc un cas particulier du problème à résoudre pour CART : celui où $f(x) = g^{-1}({}^t\beta \cdot x)$
- Dans le GLM on fait une hypothèse plus forte et l'estimation de β est plus simple (β est de dimension finie). Dans CART on ne fait aucune hypothèse et le problème à résoudre est plus complexe (f est de dimension infinie) : on peut néanmoins chercher une solution approchée par la construction d'un arbre.