

## ALGORITHME DE CONSTRUCTION DE L'ARBRE : CRITÈRE D'ARRÊT

- Notations :
  - On note  $F$  un arbre et  $Var(F)$  sa variance, i.e. la somme des variances de ses feuilles pondérée par les effectifs des feuilles
  - Soit  $F_0$  l'arbre initiale (tous les individus sont dans la feuille initiale) et soient  $F_1, F_2, \dots, F_{10}, \dots$  les arbres successifs construits pendant la procédure
  - Remarques :
    - $Var(F_0) = Var(Y)$
    - Le  $R^2$  (train) associé à un arbre  $F$  est :  $R_F^2 = 1 - Var(F)/Var(Y)$
- Critères d'arrêt :
  - L'algorithme s'arrête lorsque  $Var(F_i) - Var(F_{i+1}) < c \cdot Var(Y)$  (soit  $R_{F_{i+1}}^2 - R_{F_i}^2 < c$ ), avec  $c$  le paramètre de complexité de l'arbre (noté « cp » dans  $R$ )
  - C'est-à-dire : lorsque l'augmentation du  $R^2$  dû à l'ajout d'une feuille est trop faible, on s'arrête.
  - Le critère d'arrêt précédent est le critère d'arrêt « principal ». En pratique les critères suivants se combinent avec le critère principal dans le choix de l'arrêt (fonction « rpart.control » dans  $R$ ):
    - Nombre minimal d'individus d'un nœud pour qu'une coupure soit cherchée.
    - Nombre minimal d'individus d'une feuille
    - Profondeur maximale : longueur maximale d'un chemin qui descend depuis le haut de l'arbre vers les feuilles.