

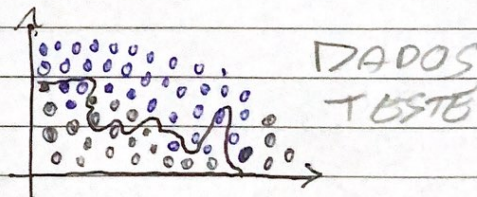
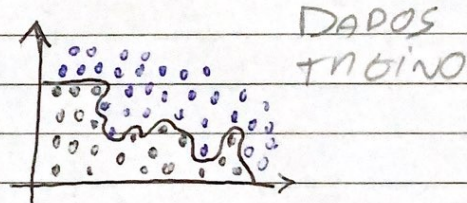
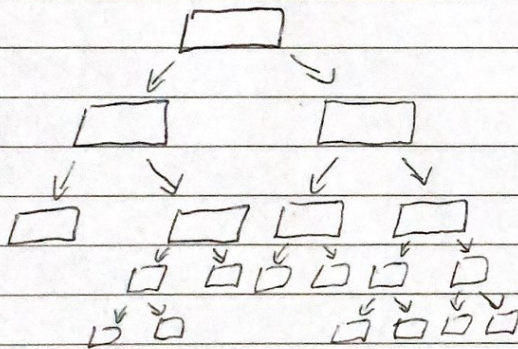
Decision Trees: Evitando Overfitting

$$GINI = 1 - \sum_{i=1}^C p_i^2$$

$$GINI A = 0,46$$

$$GINI B = 0,74$$

OBS. Pode ser que não criamos uma árvore completa (até o fim), pois ao fazermos isso poderíamos ter um problema de overfitting, ou seja, que criamos uma árvore perfeita para os dados de treino, mas que ao ser submetida aos dados de teste ela não vai ter um resultado tão bom.



Para evitar o overfitting é necessário, portanto definir critérios de parada na criação da árvore de decisão para torná-la mais generalizável. Esses critérios são definidos limitando o número máximo de nós (ou folhas) que a árvore vai ter, além da profundidade dessa árvore.