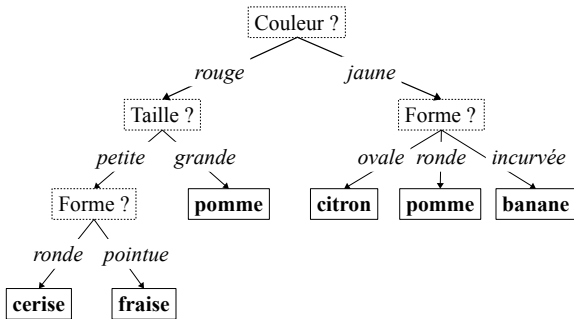


Cours 3 – Arbres et méthodes ensemblistes

1. Arbres de décision

1.1 Exemple



1.2 Construction d'un arbre

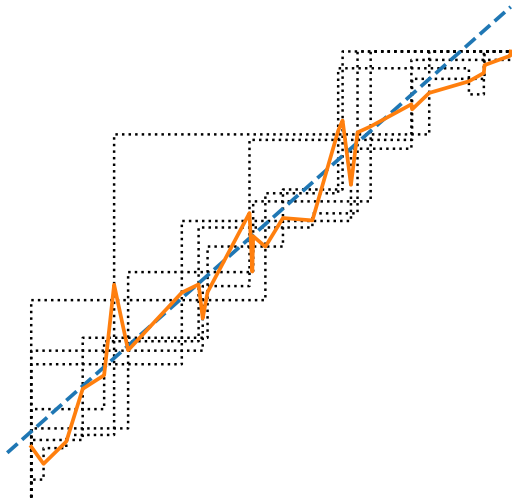
- Exemple :
 - 4 petites pommes jaunes, 5 grosses pommes vertes, 3 bananes, 6 mini-concombres
 - attributs : vert/jaune, rond/pas rond, gros/petit

1.3 Avantages et inconvénients

2. Méthodes ensemblistes

2.1 Sagesse des foules

- Combiner des **apprenants faibles**.



2.2 Bagging

2.3 Forêts aléatoires

2.4 Boosting

- À l'itération m
 - Apprendre le modèle f_m qui minimise l'erreur empirique de

$$F_m = \sum_{l=1}^m \alpha_l f_l = F_{m-1} + \alpha_m f_m$$

- L'erreur pour (\vec{x}_i, y_i) est pondérée de sorte à donner plus d'importances aux exemples pour lesquels F_{m-1} se trompe.
- **AdaBoost** (Schapire & Freund 1997) : erreur exponentielle, f_m est un arbre de décision de profondeur 1 (decision stump)
- **Gradient Boosting** (Friedman 2001) : forme générale