Machine Learning

Session 4 - Decision trees and ensemble methods


hadrien.salem@centralelille.fr


introduction-to-data-science

Introduction

What did we do last time?

Course outline

Machine learning course

Session 1: Regression

Session 2: Supervised classification

Session 3: Clustering

Session 4: Decision trees and ensemble methods

Session 5: Introduction to neural networks

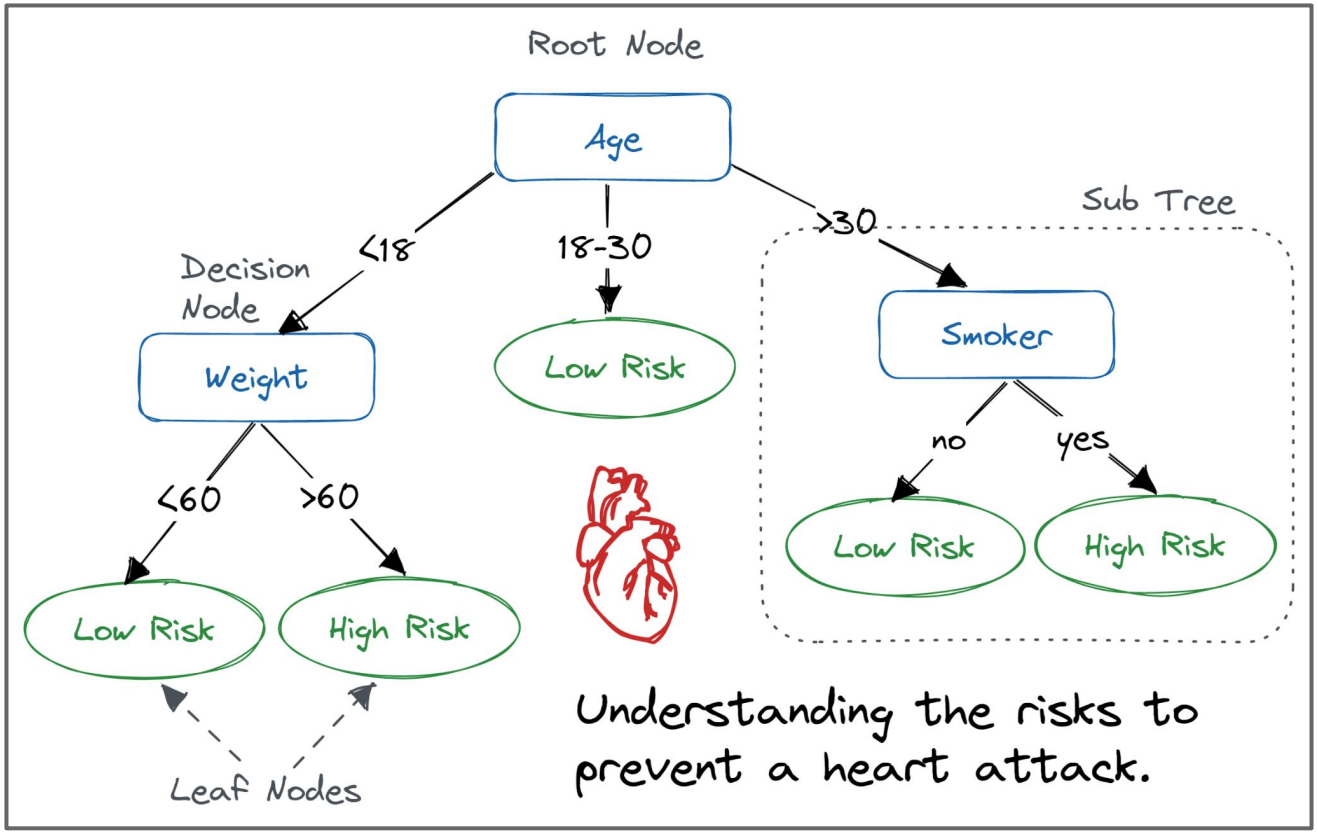
Session 6: Advanced neural networks

Session 7: Introduction to reinforcement learning

Session 8: Reading science papers

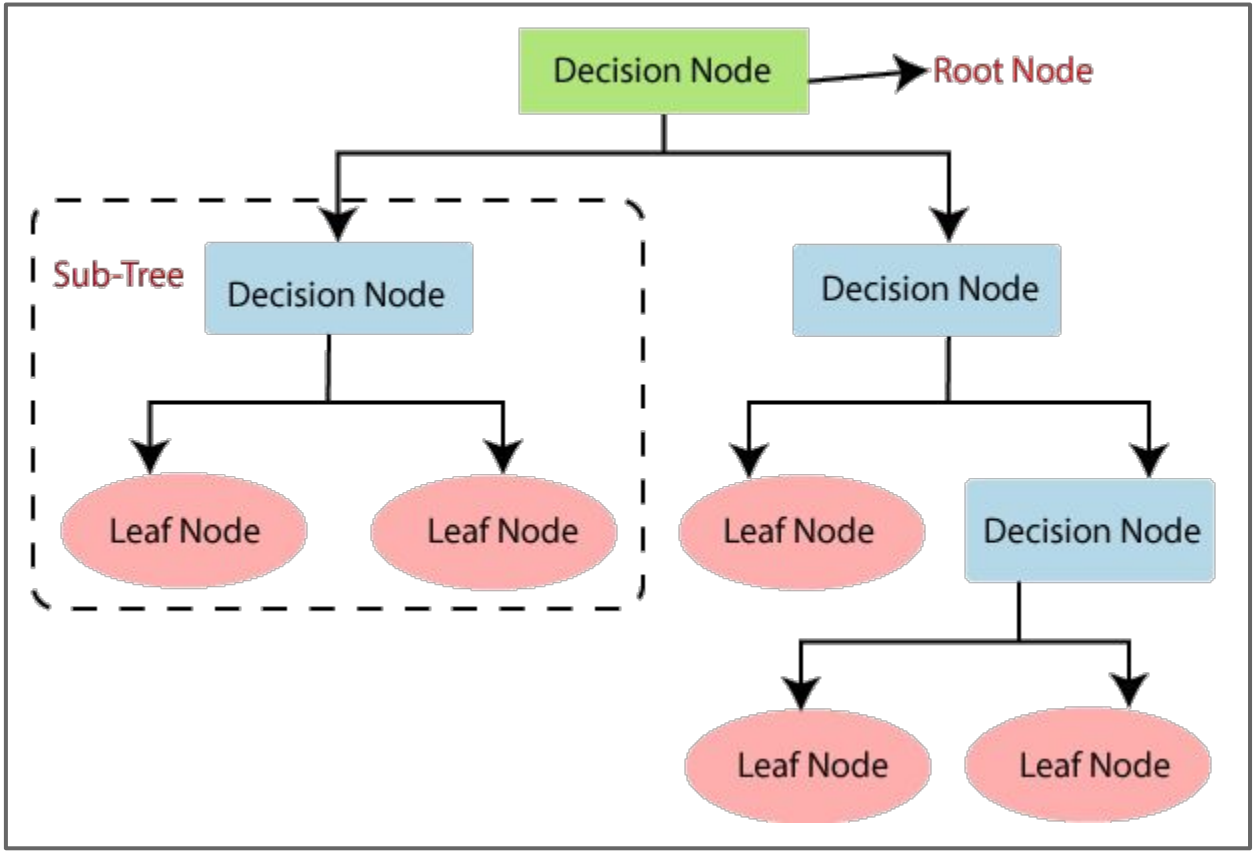
Project

What are decision trees?



Decision trees - Example

Image source



Vocabulary

Image source

WIP

Soient m_L et m_R les noeuds fils gauche et droit du noeud m, et soient π_L et π_R les proportions de données partant respectivement dans ces noeuds fils.

On évalue la qualité d'une partition s par le critère :

$$\Delta\phi(s, m) = \underbrace{\phi(p_m)}_{\text{impureté avant partition}} - \underbrace{(\pi_L\phi(p_{m_L}) + \pi_R\phi(p_{m_R}))}_{\text{impureté après partition}}$$

Pour maximiser ce critère, on peut :

- Essayer toutes les partitions à chaque noeud (coûteux).
- Sous-échantilloner (*i.e.* se limiter au test de certains seuils).

Strengths and weaknesses of decision trees

Ensemble methods

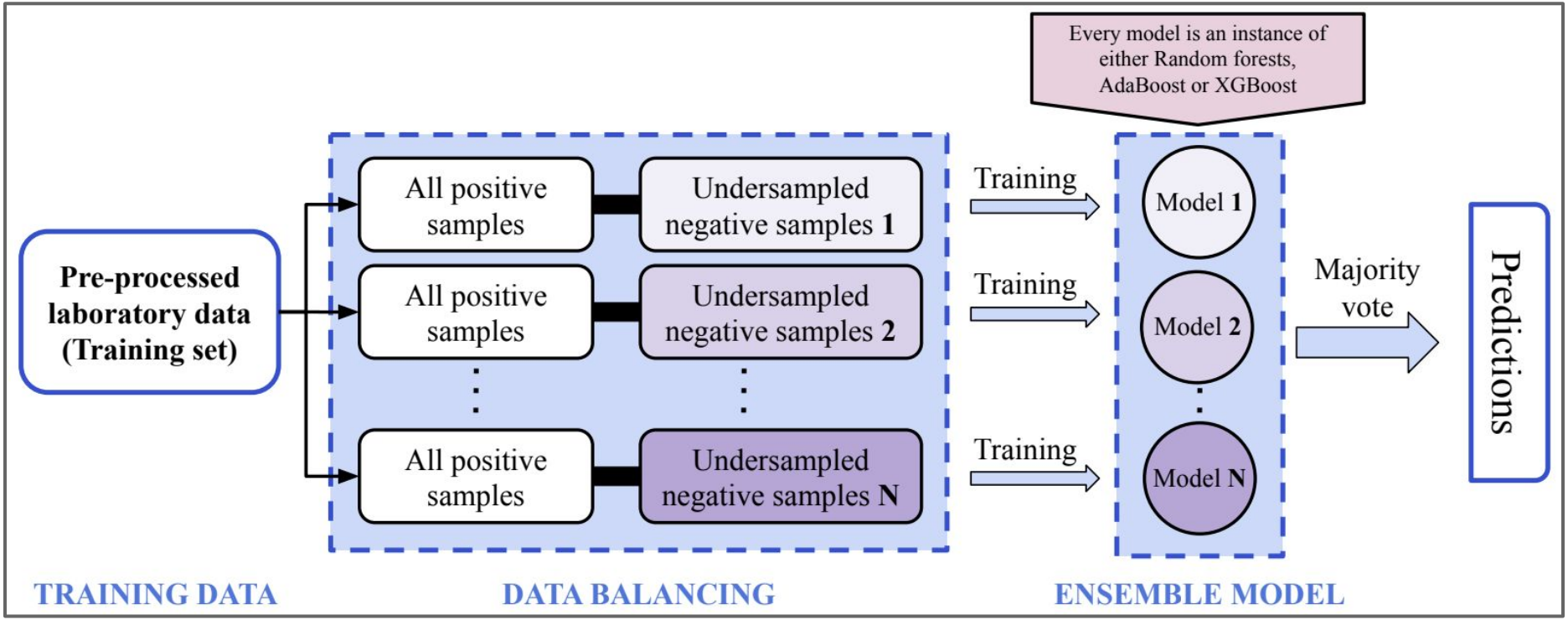
Vocabulary on decision trees

Entropy

-

Bagging

Boosting - principle



EasyEnsemble (example application to laboratory data)

Strengths and weaknesses of of ensemble methods

Practical work

The notebook contains all the necessary instructions

Debrief

Debrief

What did we learn today?

What could we have done better?

What are we doing next time?