Decision Trees & Ensemble Models

Week 7 / Day 3 / 2h

Decision Trees

① Pros & cons
② Intuition behind algo
③ Training algorithms & concepts
④ Training step by step
⑤ Jupyter examples

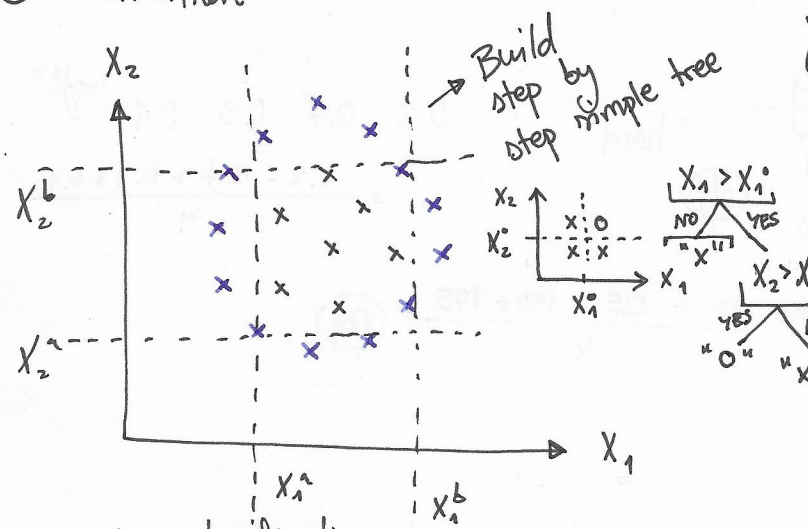
① · pros ✓

+ simple
+ can be visualized
+ computationally efficient
+ no need for extensive preproc.

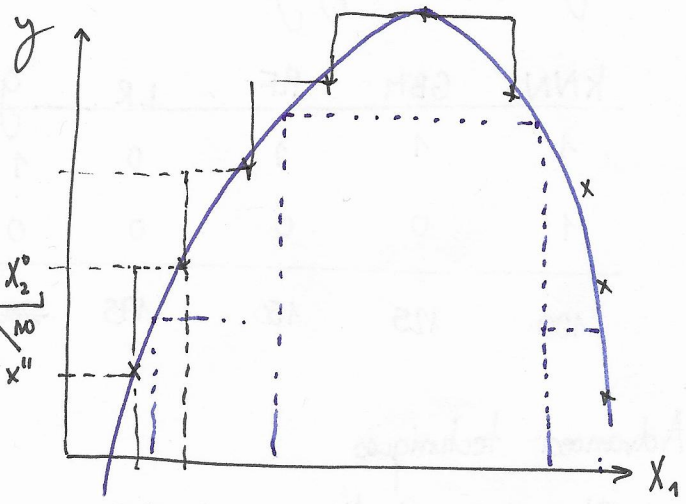
· cons ✗

+ sensitive to small variations in data
+ sensitive to highly imbalanced data

② · intuition



example : classification



example : regression

③ · training algos & concepts

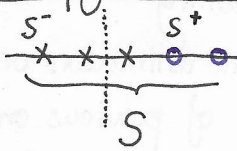
+ impurity ⟷ gini (mention only)
+ information gain ⟷ entropy

} measure how different are samples in groups.

Try to create homogeneous groups : ID3, C4.5, C5.0, CART

⑤ · Jupyter examples for decision tree.

④ training example (iterative dichotomizer)

b) 

$f_{ID3}^S = \frac{2}{5}$ ○○ ××× $H(S) = 0.292$

c) Test splits for all features and thresholds:

$S^+_{x_{(j)}>t}$, $S^-_{x_{(j)}\leq t}$ $H(S^+) = -\frac{2}{3}\cdot\log(\frac{2}{3}) - \frac{1}{3}\cdot\log(\frac{1}{3}) =$

$H(S^+, S^-) = 3/5 * 0.27 + 2/5 \cdot 0 = 0.162$

a) $L(\hat{y}, y) = -[y \cdot \log(\hat{y}) + (1-y)\cdot\log(1-\hat{y})]$

$L(1,1) = -[1\cdot\log(1) + 0\cdot\_]$; $L(\sim 0,1) = -[1\cdot\log(\sim 0) + 0\cdot\sim]$

log-likelihood or cross-entropy

$\hat{y} = f_{ID3}^S$ entropy : $H(S) = -f_{ID3}^S \cdot \ln f_{ID3}^S - (1-f_{ID3}^S)\cdot\ln(1-f_{ID3}^S)$

0.27 $H(S^-) = 0$ d) Choose best split :

Ensemble Models

① Concepts

"meta model"

Combining multiple models (weak learners) to improve overall performance.
Weak learners: simple model with sometimes just over the baseline performance.

- Simple techniques: voting, averaging, weighted → Jupyter
- Advanced techniques: stacking, blending // bagging, boosting
  - bagging, boosting → Tree based ensembles

② Simple techniques

- Classification: voting (soft, hard)
- Regression: averaging — with weights??

KNN	GBM	RF	LR	y
1	1	1	0	1
1	0	0	0	0
100	125	150	175	~~200~~

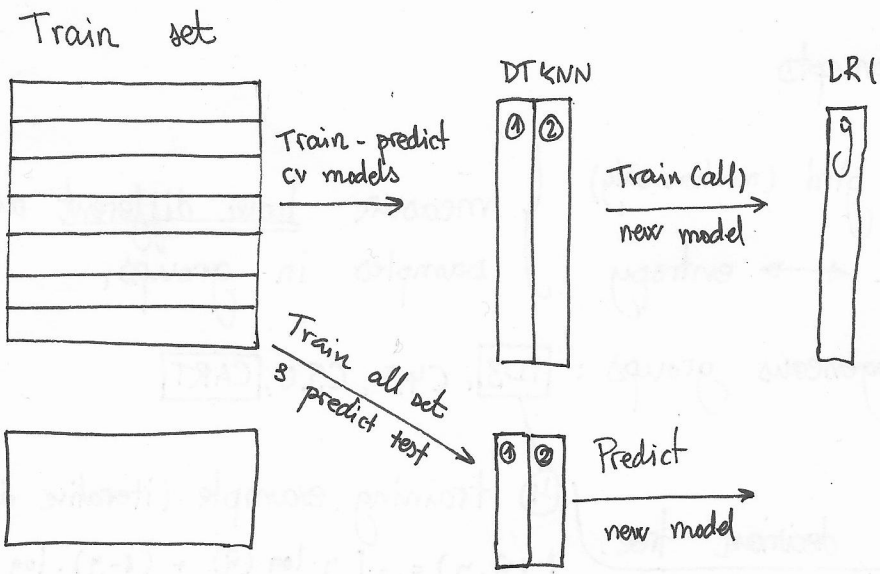
"hard" (clf)

"soft" (clf): 0.2 0.7 0.3 0.1 → $= \frac{0.2 + 0.7 + 0.3 + 0.1}{4}$

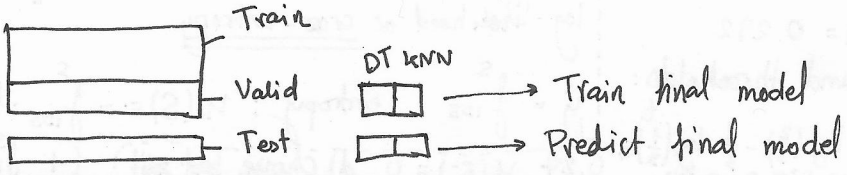
(reg): $\frac{100 + 125 + 150 + 175}{4}$

③ Advanced techniques

- ~~Blending~~ Stacking (CV) 3.1



- Blending (Holdout) 3.2



3.3 Bagging

- Train n estimators in random subsamples, of original data, with replacement. RF, Extra Trees

3.4 Boosting

- Train estimators over errors of previous ones. Presentation examples.