



CIS 635 Knowledge Discovery & Data Mining

ML Models: Decision Tree, Ensemble Learning



Decision Tree

- Another non-parametric model
 - Recall k-NN, its an in memory model; right?



Decision Tree

- Another non-parametric model
 - Recall k-NN, its an in memory model; right?
- Decision Tree is our second example

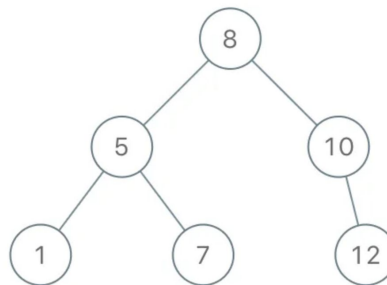


Decision Tree

- Another non-parametric model
 - Recall k-NN, its an in memory model; right?
- Decision Tree is our second example
 - Those are with CS background are already aware of BST
 - Whiteboarding

Decision Tree

- Another non-parametric model
 - Recall k-NN, its an in memory model; right?
- Decision Tree is our second example
 - Those are with CS background are already aware of BST
 - Whiteboarding



8, 5, 10, 1, 12, 7



Decision Tree

- Let's start with the Classification Task
- Whiteboarding

nb of legs	weight (lb)	animal
4	2.1	Bunny
4	7	Cat
4	1.7	Bunny
4	9	Cat
4	2.75	Bunny



Decision Tree

- Let's start with the Classification Task
- Whiteboarding

nb of legs	weight (lb)	animal
4	2.1	Bunny
4	7	Cat
4	1.7	Bunny
4	9	Cat
4	2.75	Bunny
2	2.5	Chicken
2	3	Chicken

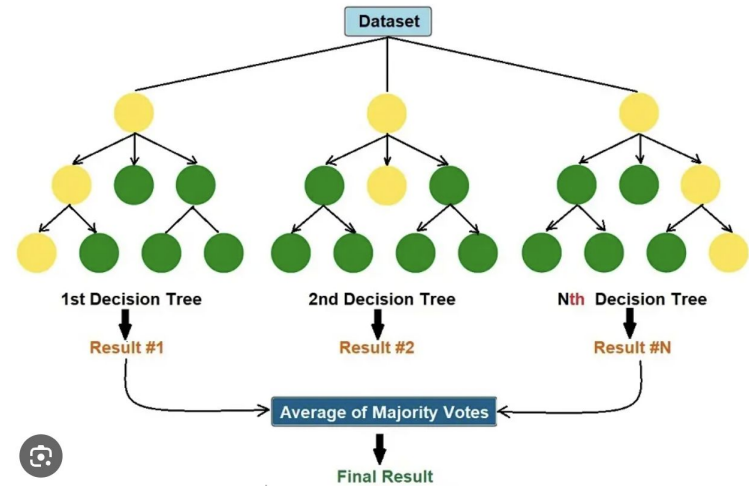


Ensemble Learning/Meta Learning

- Ensemble (more than one model)
 - Bagging
 - Averaging (RF)
 - Reduces Variance
 - Boosting
 - Gradual improvement over weak learners (Adaboost/XGBoost)
 - Reduces Bias??

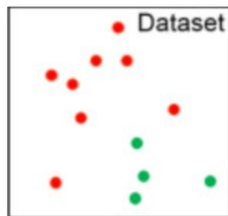
Random Forest

- Instead of one, we have many (but finite) Trees

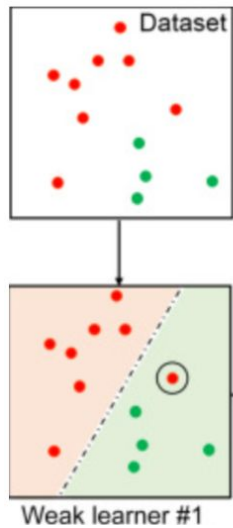


Boosting

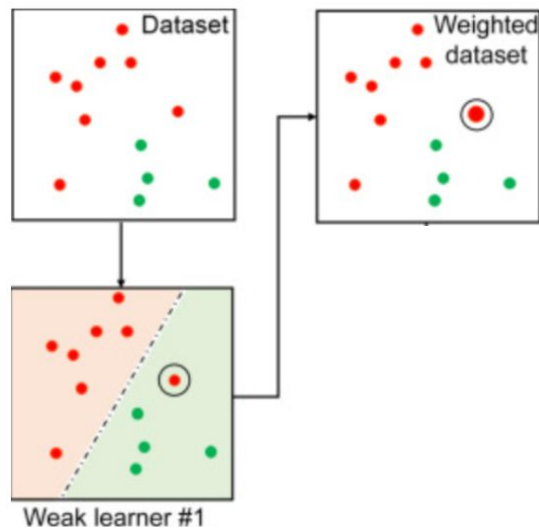
- Combination of some weak models
- Power of many



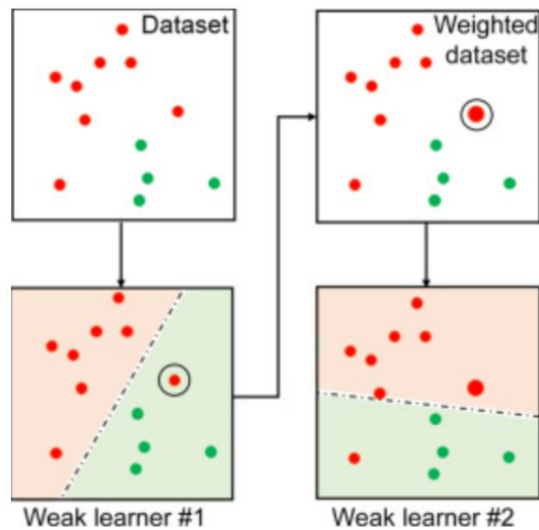
Boosting



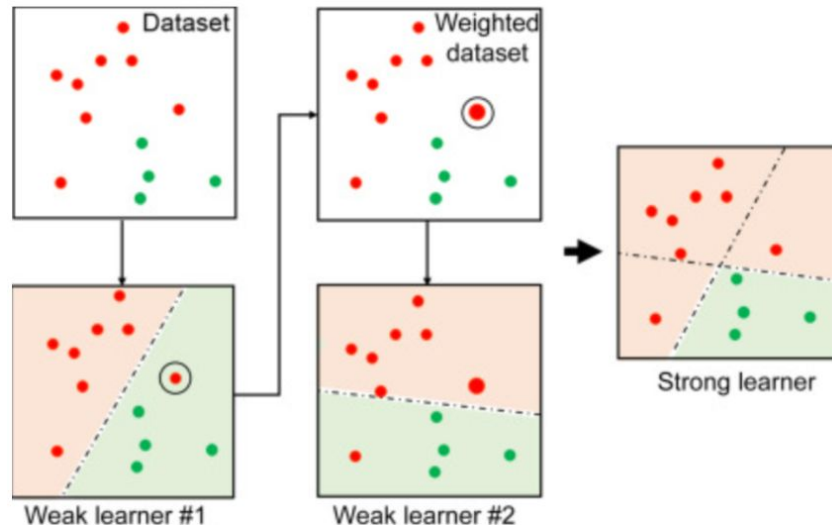
Boosting



Boosting



Boosting





Classification vs Regression

- Simple ideas!
- Whiteboarding



Notebook extension/presentation

- See notebooks section (Blackboard)