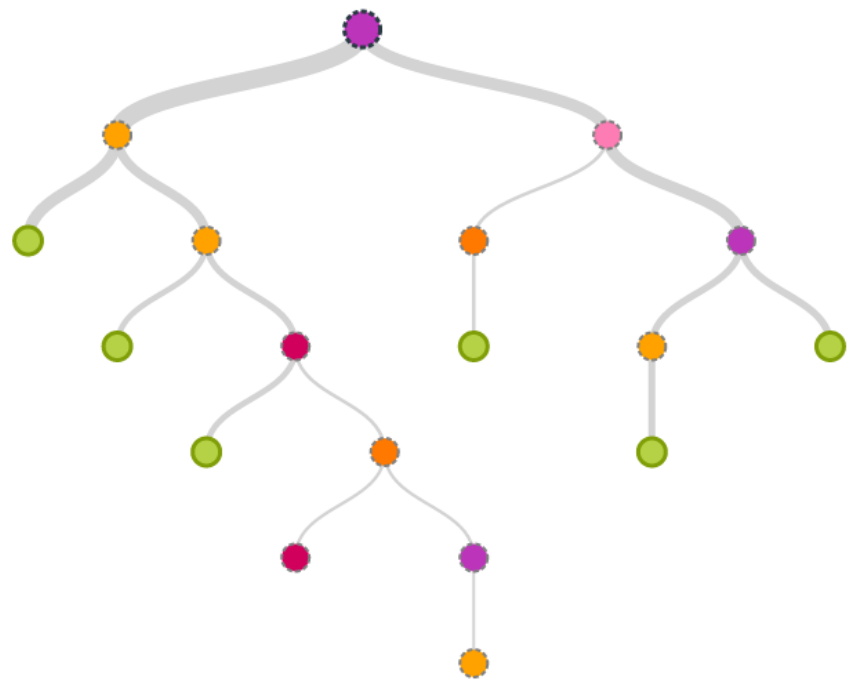


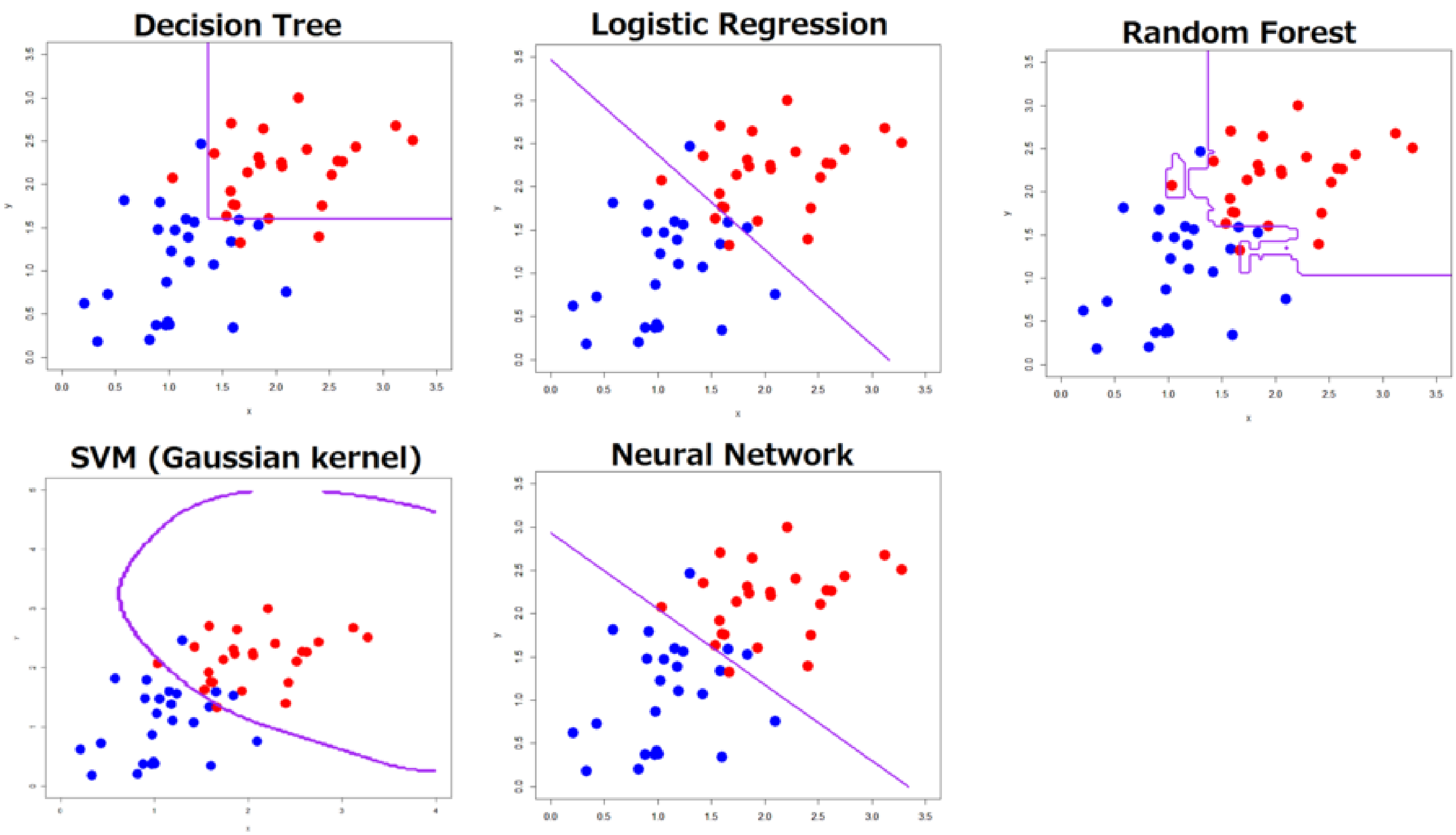
Lecture 05: Decision Tree

Pilsung Kang
School of Industrial Management Engineering
Korea University

AGENDA

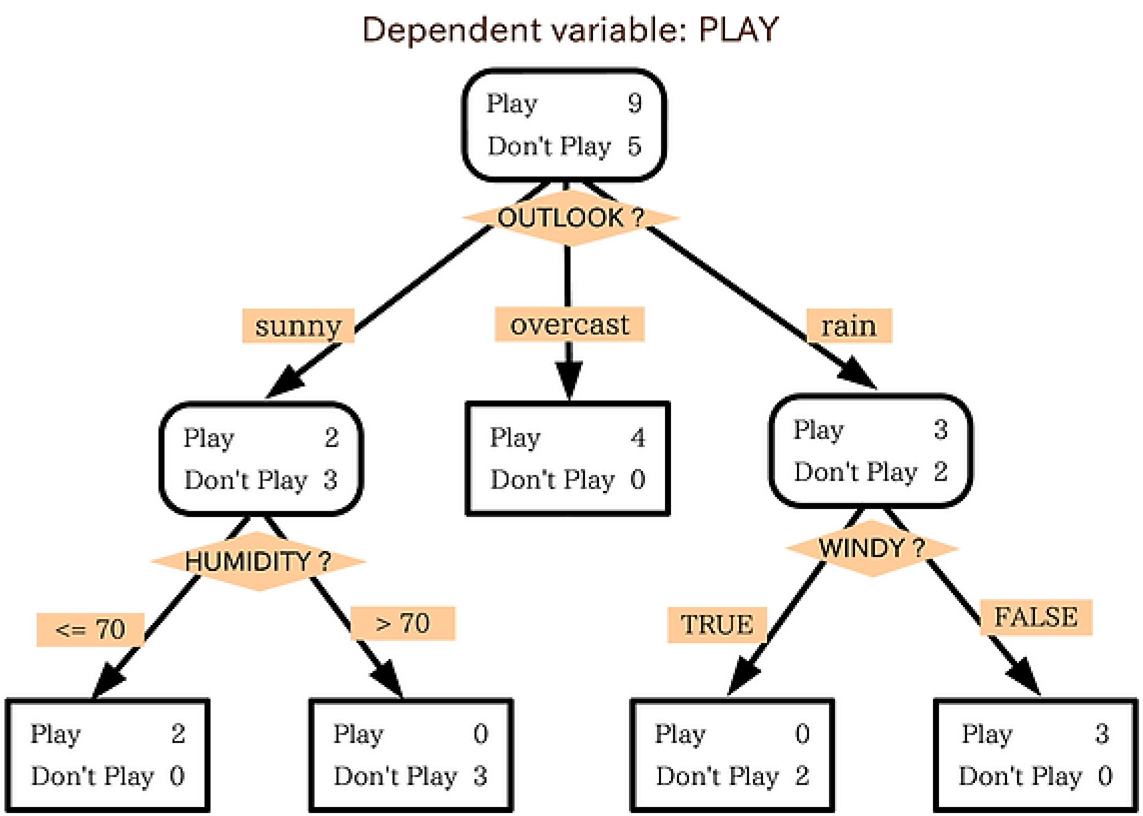
Why Are There So Many Classifiers?

- We cannot guarantee that a single classifier is always better than the others



Classification Tree

- Goal
  - ✓ Classify or predict an outcome based on a set of predictors.
  - ✓ The output is a set of rules.



Rule example

If outlook is sunny

and if humidity > 70

then he does not play

or

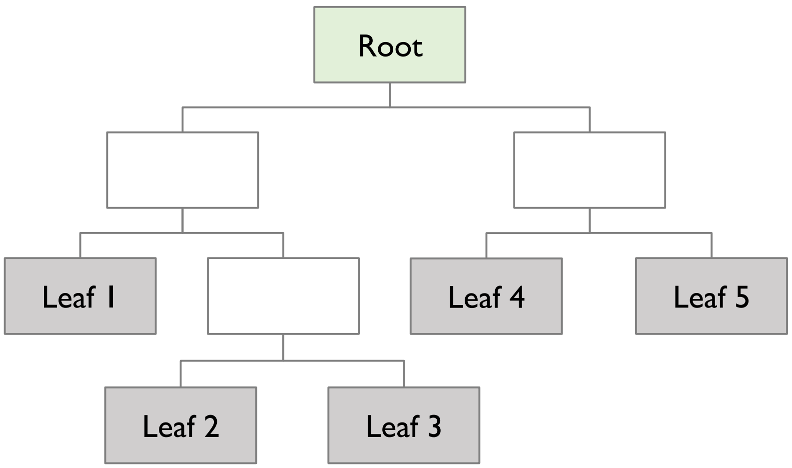
If outlook is rainy

and it is not windy

then he does play

Classification Tree

- Terminologies
  - ✓ Parent node: node before split
  - ✓ Child node: node after split
  - ✓ Split criterion: a certain variable value used for split a node
  - ✓ Root node: node that only has child nodes but no parent node
  - ✓ Leaf nodes: nodes that only have a parent node but no child nodes

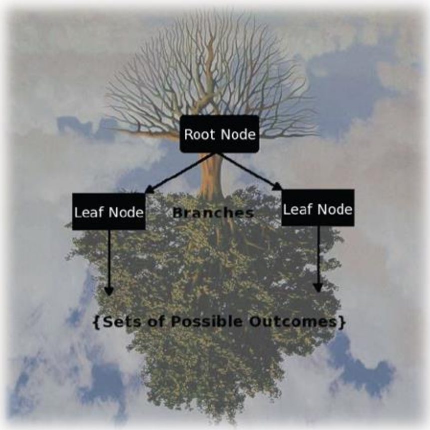


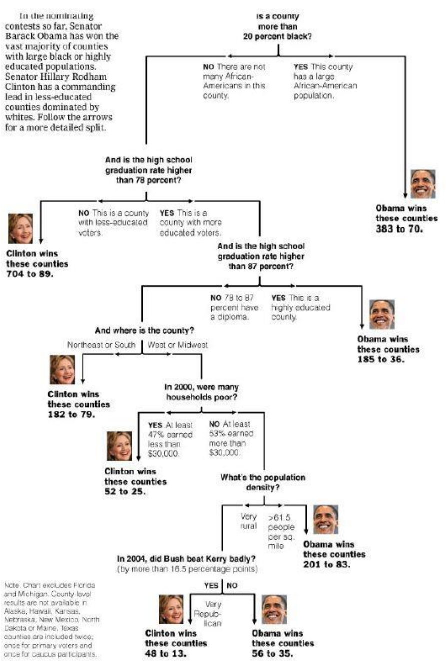
Classification Tree

- Why CART?
  - ✓ Simple to understand and interpret.
  - ✓ Requires little data preparation (normalization, missing value treatments, etc.)
  - ✓ Able to handle both numerical and categorical data.

- Key Ideas
  - ✓ Recursive Partitioning
    - Repeatedly split the records into two parts so as to achieve maximum homogeneity within the new parts.
  - ✓ Pruning the Tree
    - Simplify the tree by pruning peripheral branches to avoid over-fitting.

Classification Tree





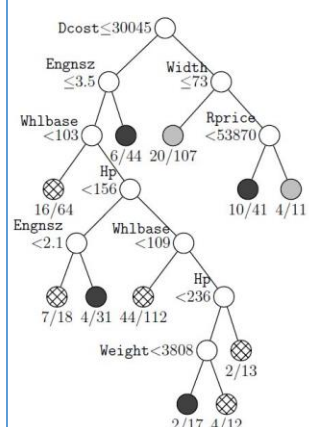
Classification and Regression Tree (CART)

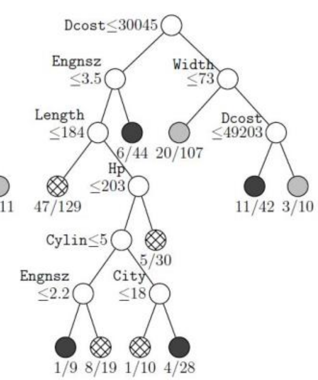
- Generate a set of rules by recursively partitioning the entire datasets to increase the purity of the partitioned area (Breiman, 1984)
- Being able to explain the reason of the prediction result by following the rules to the target leaf node
- Can handle categorical and numerical variables simultaneously

Recursive Partitioning

Pruning

- Partition the data in a parent node into two child nodes using a certain value of a certain variable
- Select the split point to maximize the purity of the child nodes
- Gini-index (for categorical variable) and the variance (for numerical variable) are used to measure the impurity of a node



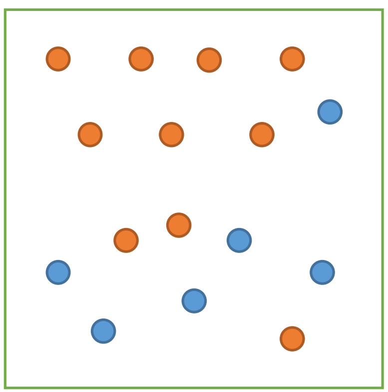


Classification Tree: Impurity Measure

- Measuring Impurity 1: Gini Index
  - ✓ Gini Index for rectangle A containing m records

$$I(A) = 1 - \sum_{k=1}^{m} p_k^2$$

  - p = proportion of cases in rectangle A that belong to class k.



$$I(A) = 1 - \sum_{k=1}^{m} p_k^2$$

$$= 1 - \left(\frac{6}{16}\right)^2 - \left(\frac{10}{16}\right)^2$$

$$\approx 0.47$$

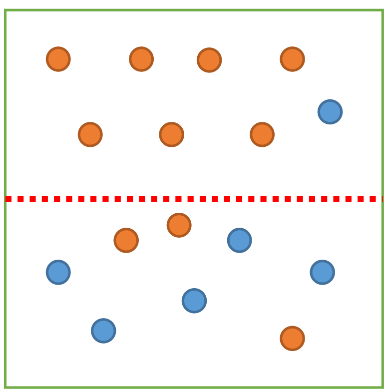
  - I(A) = 0 when all cases belong to the same class.
  - Max value when all classes are equal represented (=0.5 in binary case)

Classification Tree: Impurity Measure

- Measuring Impurity 1: Gini Index
  - ✓ When their more than two rectangles

$$I(A) = \sum_{i=1}^{d}\left(R_i\left(1-\sum_{k=1}^{m} p_{ik}^2\right)\right)$$

- R_i = proportion of cases in rectangle Ri among the training data.



$$I(A)$$

$$= 0.5\times\left(1-\left(\frac{7}{8}\right)^2-\left(\frac{1}{8}\right)^2\right)+0.5\times\left(1-\left(\frac{3}{8}\right)^2-\left(\frac{5}{8}\right)^2\right)$$

$$= 0.34$$

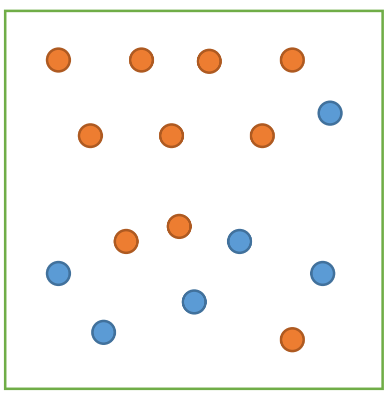
- "Information gain" after splitting: 0.47-0.34=0.13

Classification Tree: Impurity Measure

- Measuring Impurity 2: Deviance

$$D_i = -2 \sum_k n_{ik} log(p_{ik})$$

✓ i: node index, k: class index, p_{ik}: probability of class k in node I



$$D_i = -2 \times \left(10 \times log\left(\frac{10}{16}\right) + 6 \times log\left(\frac{6}{16}\right) \right)$$

$$= 21.17$$

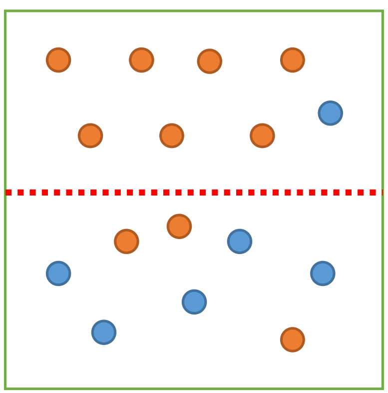
✓ Deviance = 0 if and only if all nodes contain instances from the same class

Classification Tree: Impurity Measure

- Measuring Impurity 2: Deviance

$$D_i = -2 \sum_k n_{ik} log(p_{ik})$$

✓ i: node index, k: class index, p_{ik}: probability of class k in node I



$$D_1 = -2 \times \left(7 \times log\left(\frac{7}{8}\right) + 1 \times log\left(\frac{1}{8}\right)\right) = 6.03$$

$$D_2 = -2 \times \left(3 \times log\left(\frac{3}{8}\right) + 5 \times log\left(\frac{5}{8}\right)\right) = 10.59$$

$$D_1 + D_2 = 16.62$$

✓ Information gain = 21.17 – 16.62 = 4.55

Classification Tree: Recursive Partitioning

- Example: Riding Mowers
  - ✓ Goal: Classify 24 households as owning or not owning riding mowers
  - ✓ Predictors: Income, Lot size

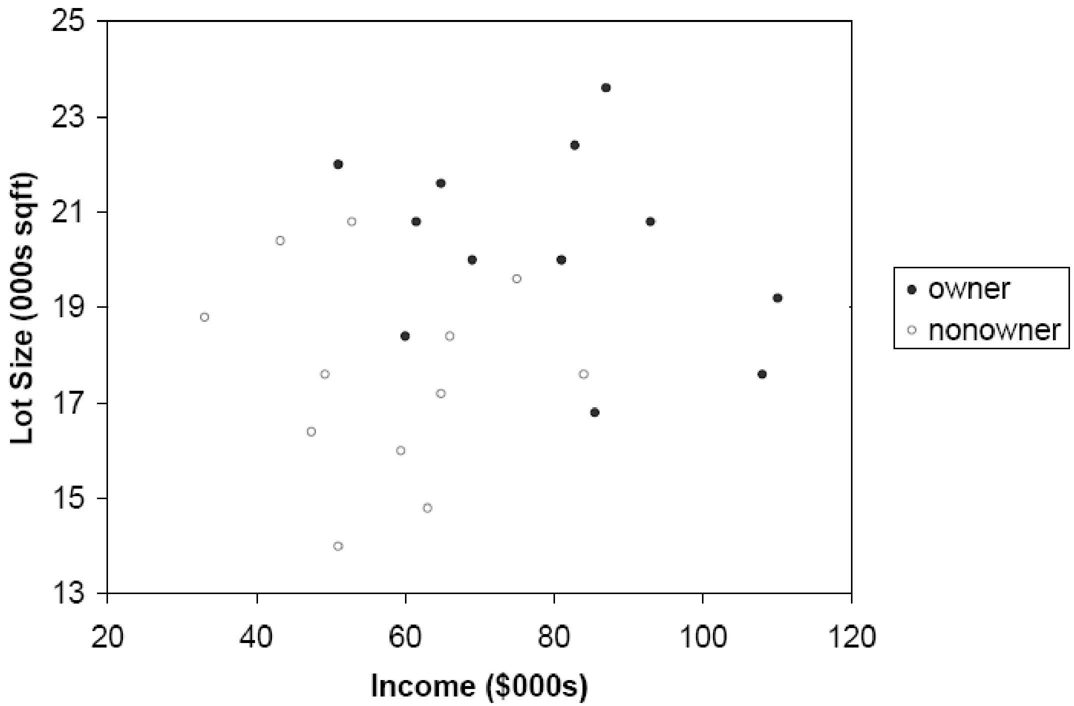
Income	Lot size	Ownership	Income	Lot size	Ownership
60.0	18.4	Owner	75.0	19.6	Non-owner
85.5	16.8	Owner	52.8	20.8	Non-owner
64.8	21.6	Owner	64.8	17.2	Non-owner
61.5	20.8	Owner	43.2	20.4	Non-owner
87.0	23.6	Owner	84.0	17.6	Non-owner
110.1	19.2	Owner	49.2	17.6	Non-owner
108.0	17.6	Owner	59.4	16.0	Non-owner
82.8	22.4	Owner	66.0	18.4	Non-owner
69.0	20.0	Owner	47.4	16.4	Non-owner
93.0	20.8	Owner	33.0	18.8	Non-owner
51.0	22.0	Owner	51.0	14.0	Non-owner
81.0	20.0	Owner	63.0	14.8	Non-owner

Classification Tree: Recursive Partitioning

1

Order records according to one variable

- Order the data with regard to lot size



Income	Lot size	Ownership
51.0	14.0	Non-owner
63.0	14.8	Non-owner
59.4	16.0	Non-owner
47.4	16.4	Non-owner
85.5	16.8	Owner
64.8	17.2	Non-owner
108.0	17.6	Owner
84.0	17.6	Non-owner
49.2	17.6	Non-owner
60.0	18.4	Owner
66.0	18.4	Non-owner
33.0	18.8	Non-owner
110.1	19.2	Owner
75.0	19.6	Non-owner
69.0	20.0	Owner
81.0	20.0	Owner
43.2	20.4	Non-owner
61.5	20.8	Owner
93.0	20.8	Owner
52.8	20.8	Non-owner
64.8	21.6	Owner
51.0	22.0	Owner
82.8	22.4	Owner
87.0	23.6	Owner

Classification Tree: Recursive Partitioning

2

Find midpoints between successive values

- First midpoint = 14.4 (0.5*(14.0+14.8))
- Divide records into those with Lot size > 14.4 and those < 14.4
- Compute the impurity: Gini index
  - ✓ Before splitting:

$$1-\left(\frac{12}{24}\right)^2-\left(\frac{12}{24}\right)^2=0.5$$

  - ✓ After splitting:

$$\frac{1}{24}\left(1-\left(\frac{1}{1}\right)^2\right)+\frac{23}{24}\left(1-\left(\frac{12}{23}\right)^2-\left(\frac{11}{23}\right)^2\right)\approx 0.48$$

  - ✓ Information gain: 0.50-0.48=0.02

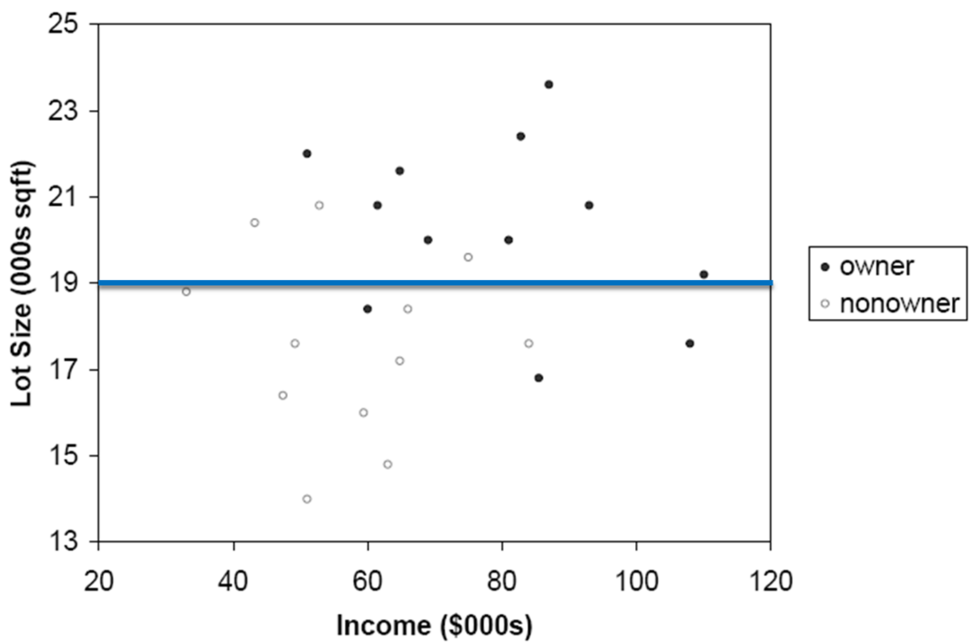
Income	Lot size	Ownership
51.0	14.0	Non-owner
63.0	14.8	Non-owner
59.4	16.0	Non-owner
47.4	16.4	Non-owner
85.5	16.8	Owner
64.8	17.2	Non-owner
108.0	17.6	Owner
84.0	17.6	Non-owner
49.2	17.6	Non-owner
60.0	18.4	Owner
66.0	18.4	Non-owner
33.0	18.8	Non-owner
110.1	19.2	Owner
75.0	19.6	Non-owner
69.0	20.0	Owner
81.0	20.0	Owner
43.2	20.4	Non-owner
61.5	20.8	Owner
93.0	20.8	Owner
52.8	20.8	Non-owner
64.8	21.6	Owner
51.0	22.0	Owner
82.8	22.4	Owner
87.0	23.6	Owner

Classification Tree: Recursive Partitioning

3

Find the best split

- Find the best split which maximize the (Gini or information gain)

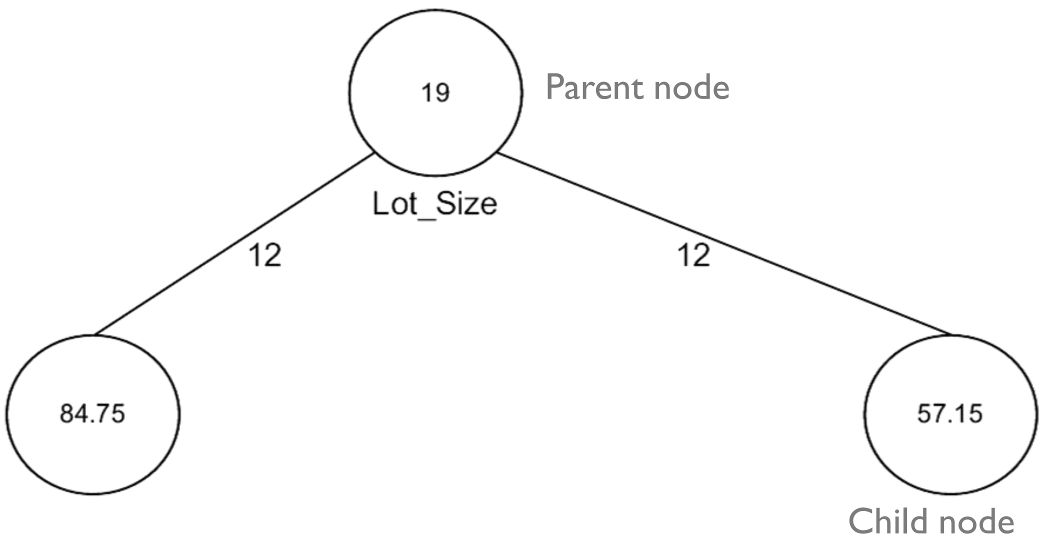


Income	Lot size	Ownership
51.0	14.0	Non-owner
63.0	14.8	Non-owner
59.4	16.0	Non-owner
47.4	16.4	Non-owner
85.5	16.8	Owner
64.8	17.2	Non-owner
108.0	17.6	Owner
84.0	17.6	Non-owner
49.2	17.6	Non-owner
60.0	18.4	Owner
66.0	18.4	Non-owner
33.0	18.8	Non-owner
110.1	19.2	Owner
75.0	19.6	Non-owner
69.0	20.0	Owner
81.0	20.0	Owner
43.2	20.4	Non-owner
61.5	20.8	Owner
93.0	20.8	Owner
52.8	20.8	Non-owner
64.8	21.6	Owner
51.0	22.0	Owner
82.8	22.4	Owner
87.0	23.6	Owner

Classification Tree: Recursive Partitioning

3

Tree structure



- Split point become nodes on tree (circles with split value in center)
- Rectangles represent "leaves" (terminal points, no future splits, classification value noted)
- Numbers on lines between nodes indicate # cases.

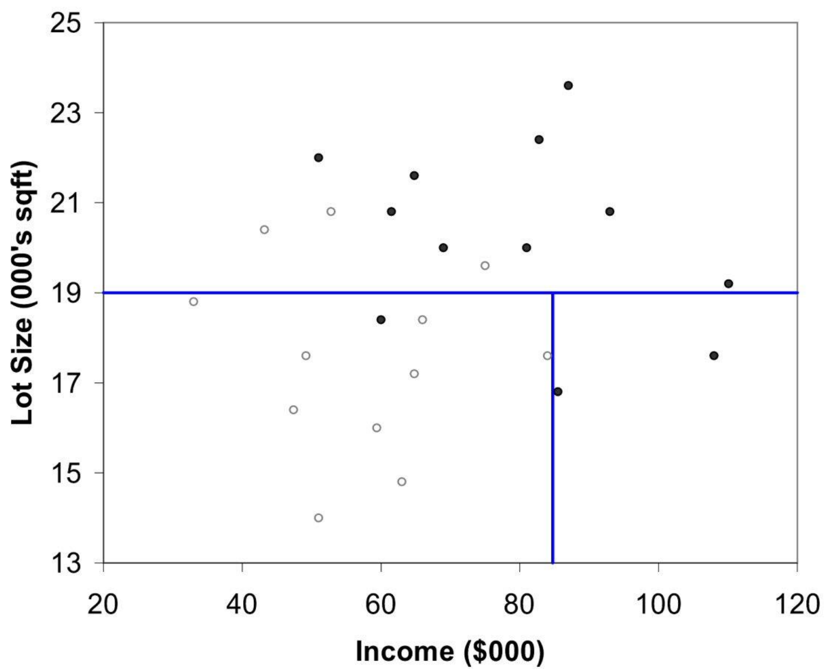
Income	Lot size	Ownership
51.0	14.0	Non-owner
63.0	14.8	Non-owner
59.4	16.0	Non-owner
47.4	16.4	Non-owner
85.5	16.8	Owner
64.8	17.2	Non-owner
108.0	17.6	Owner
84.0	17.6	Non-owner
49.2	17.6	Non-owner
60.0	18.4	Owner
66.0	18.4	Non-owner
33.0	18.8	Non-owner
110.1	19.2	Owner
75.0	19.6	Non-owner
69.0	20.0	Owner
81.0	20.0	Owner
43.2	20.4	Non-owner
61.5	20.8	Owner
93.0	20.8	Owner
52.8	20.8	Non-owner
64.8	21.6	Owner
51.0	22.0	Owner
82.8	22.4	Owner
87.0	23.6	Owner

Classification Tree: Recursive Partitioning

4

Repeat the splitting for each node

- Repeat the splitting until there is no gain.
- E.g., second split = income = 84.75



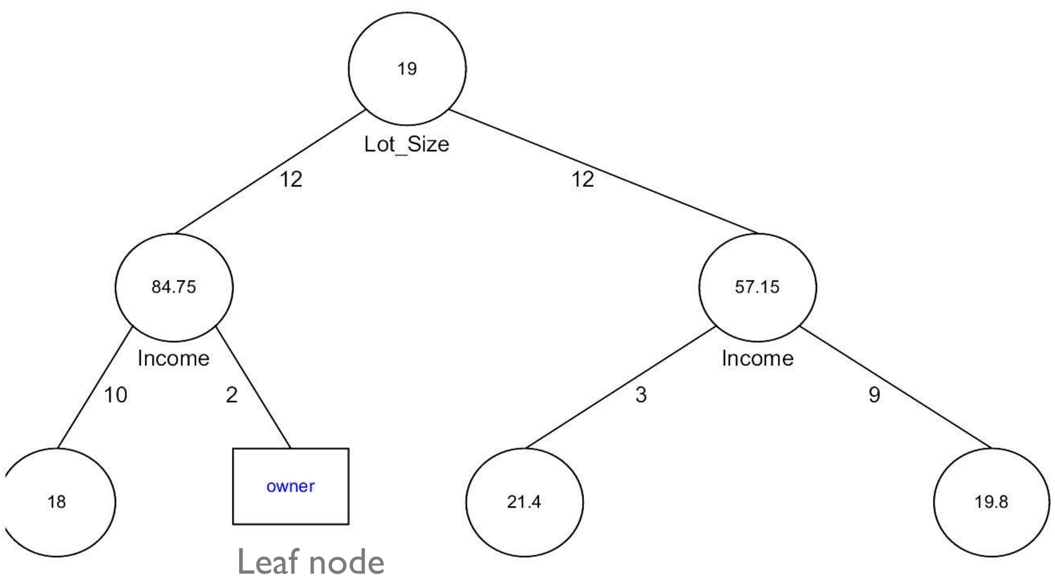
Income	Lot size	Ownership
33.0	18.8	Non-owner
47.4	16.4	Non-owner
49.2	17.6	Non-owner
51.0	14.0	Non-owner
59.4	16.0	Non-owner
60.0	18.4	Owner
63.0	14.8	Non-owner
64.8	17.2	Non-owner
66.0	18.4	Non-owner
84.0	17.6	Non-owner
85.5	16.8	Owner
108.0	17.6	Owner

Classification Tree: Recursive Partitioning

4

Repeat the splitting for each node

- Repeat the splitting until there is no gain.
- E.g., second split = income = 84.75



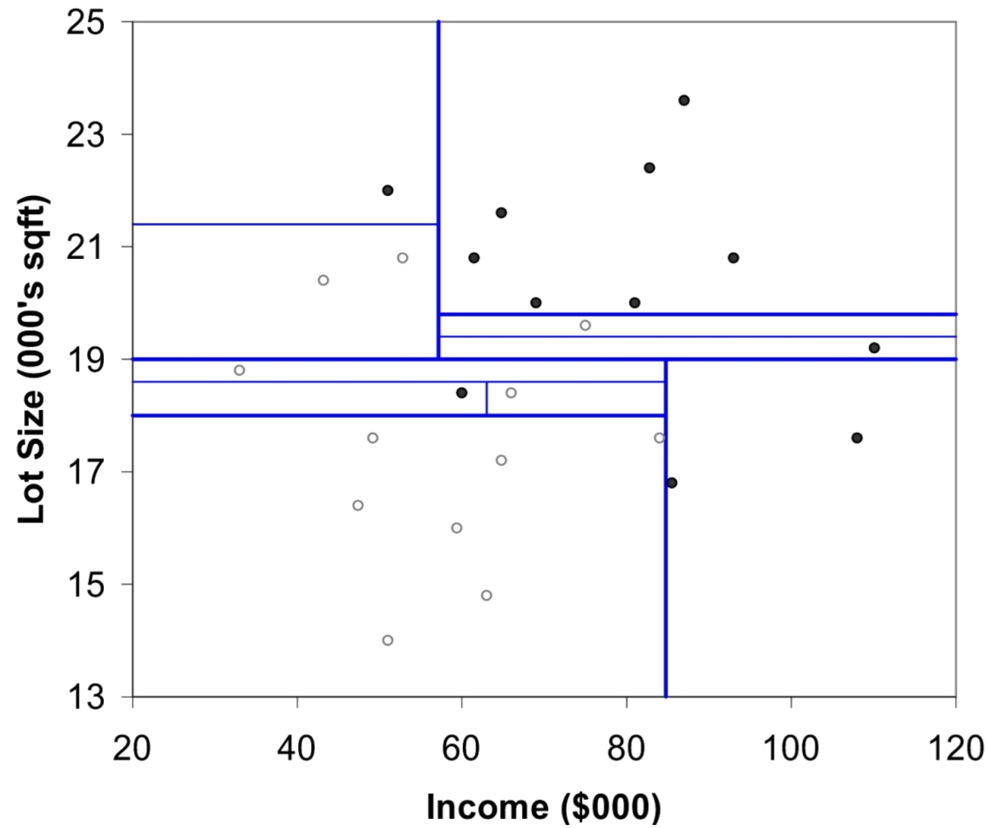
Income	Lot size	Ownership
33.0	18.8	Non-owner
47.4	16.4	Non-owner
49.2	17.6	Non-owner
51.0	14.0	Non-owner
59.4	16.0	Non-owner
60.0	18.4	Owner
63.0	14.8	Non-owner
64.8	17.2	Non-owner
66.0	18.4	Non-owner
84.0	17.6	Non-owner
85.5	16.8	Owner
108.0	17.6	Owner

Classification Tree: Recursive Partitioning

4

Repeat the splitting for each node

- Repeat the splitting until there is no gain.
- Final splitting



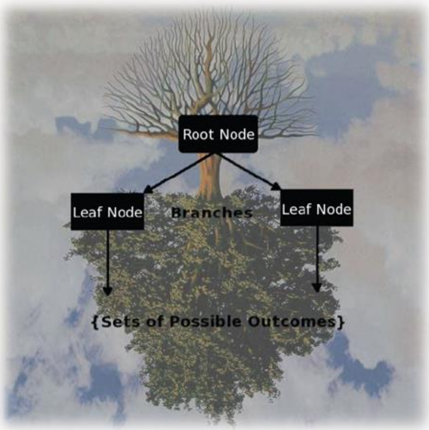
Classification Tree: Recursive Partitioning

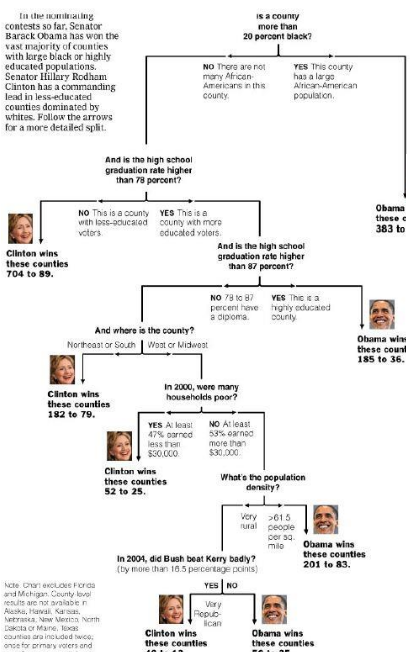
4

Repeat the splitting for each node

- Each leaf node label is determined by "voting" of the records within it, and by the cutoff value.
- Records within each leaf node are from the training data.
- Default cutoff=0.5 means that the leaf node's label is the majority class.
- Cutoff = 0.75 requires majority of 75% of more "1" records in the leaf to label it a "1" node.

Classification Tree: Pruning



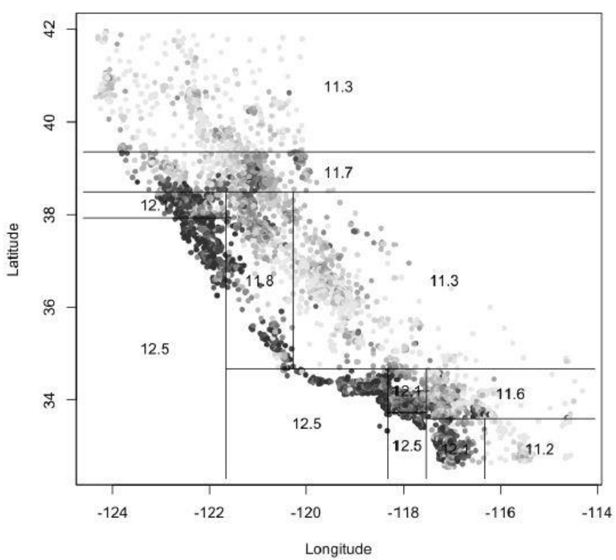


Classification and Regression Tree (CART)

- Generate a set of rules by recursively partitioning the entire datasets to increase the purity of the partitioned area (Breiman, 1984)
- Being able to explain the reason of the prediction result by following the rules to the target leaf node
- Can handle categorical and numerical variables simultaneously

Recursive Partitioning

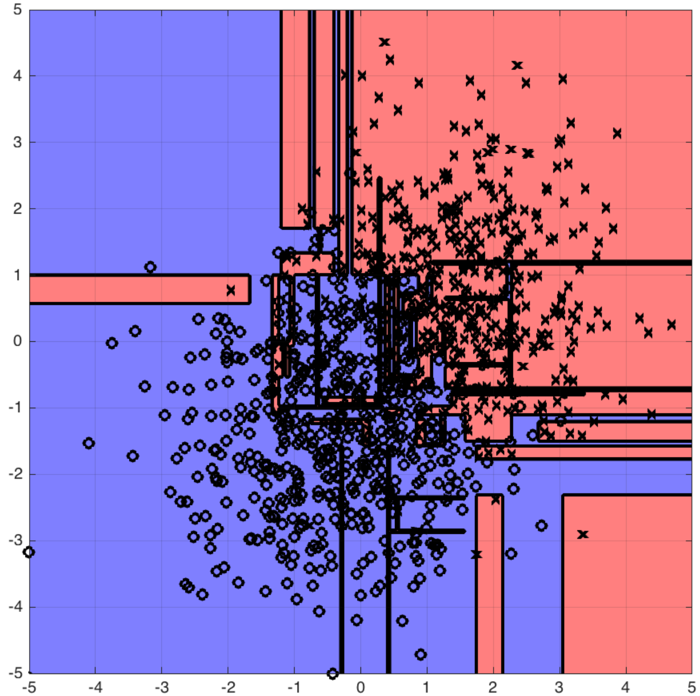
Pruning

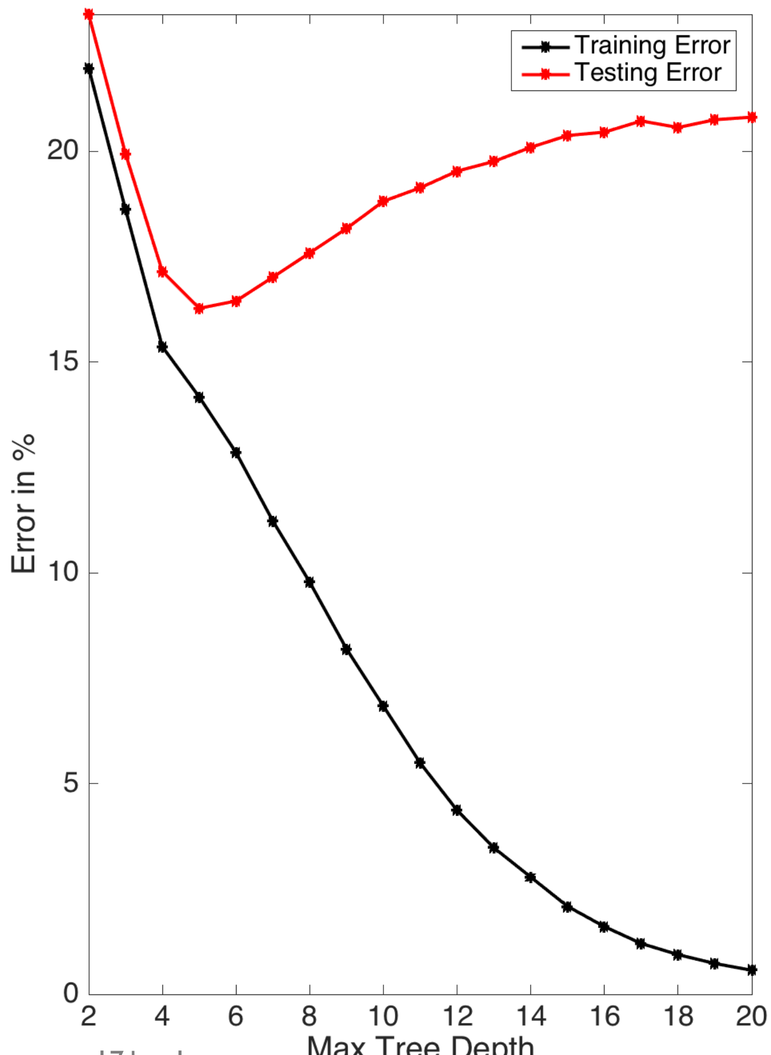


- Aggregate some child node into a parent node to avoid over-fitting
- Pre-pruning: pruning is done during the tree construction
- Post-pruning: Once a full-tree is constructed, nodes are pruned by taking the validation error and tree complexity

Classification Tree: Pruning

- Recursive partitioning is completed when every leaf node has 100% purity

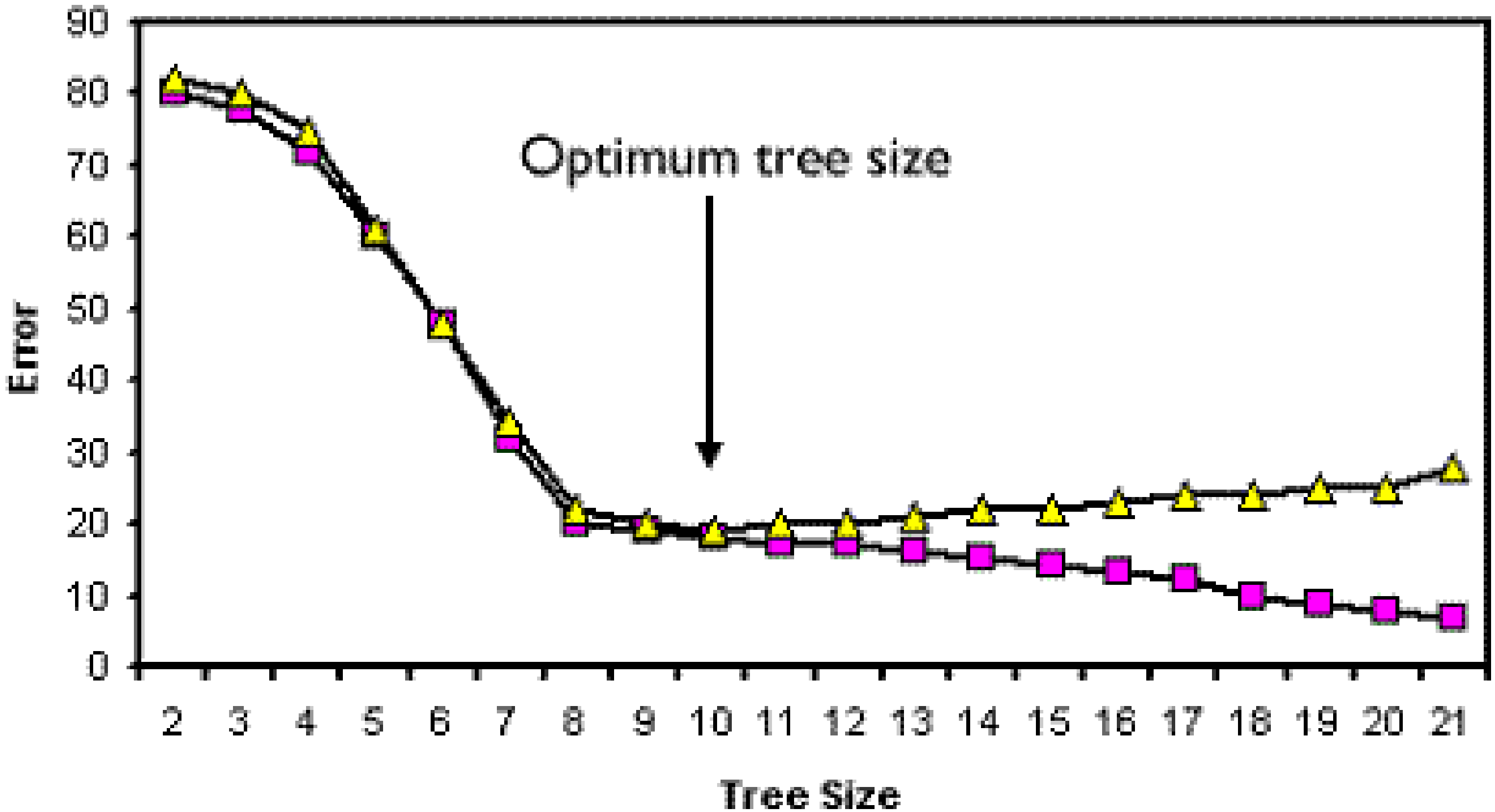




http://www.cs.cornell.edu/courses/cs4780/2017sp/lectures/lecturenote17.html

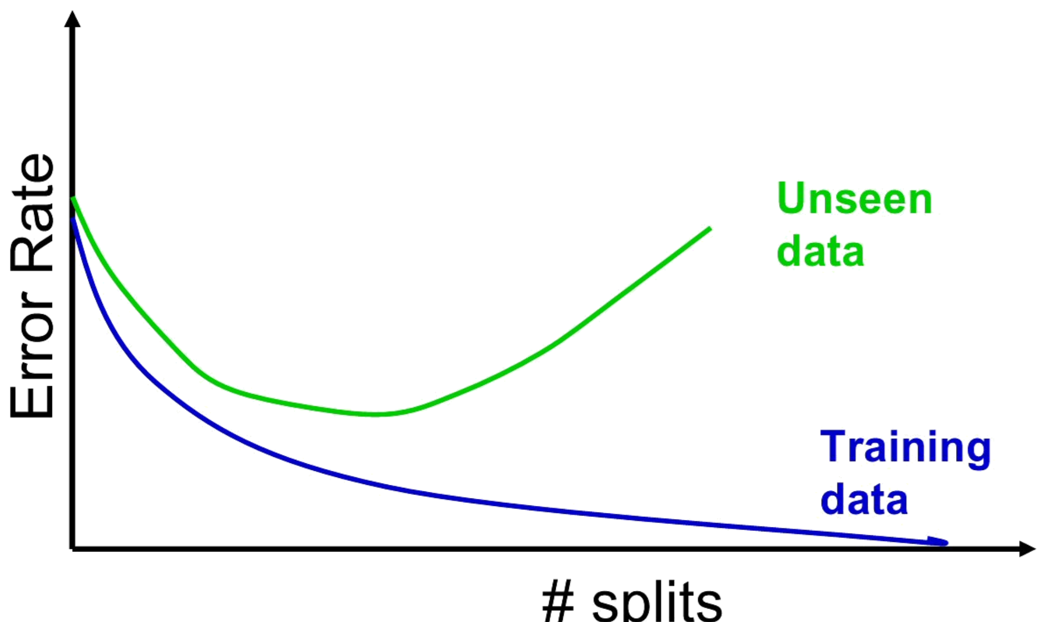
Classification Tree: Pruning

- Full tree which is a result of recursive partitioning has a risk of overfitting, which in turn, results in poor generalization ability
  - ✓ It tends to memorize the training dataset, rather than discovering significant patterns



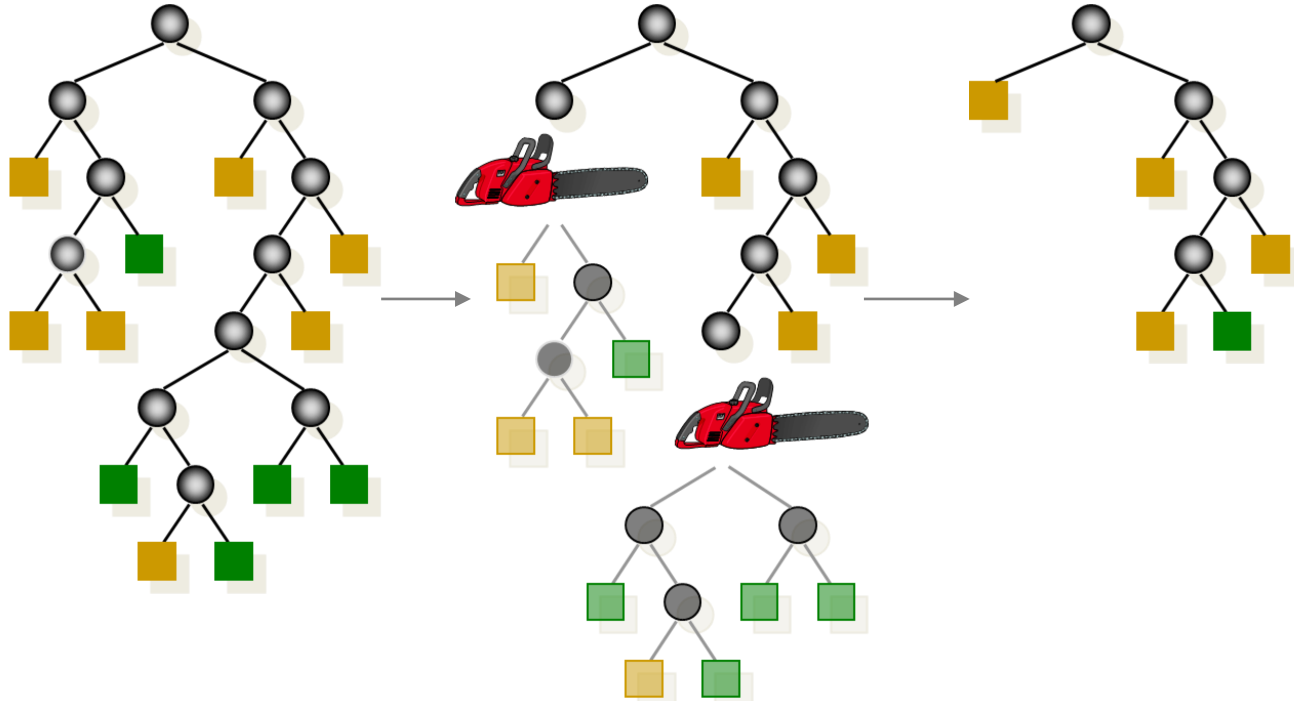
Classification Tree: Pruning

- Overfitting problem
  - ✓ The end of recursive partitioning process is 100% purity in each leaf
  - ✓ It over-fits the data, ending up fitting noise in the data and leading to low predictive accuracy of new data
  - ✓ Past a certain point, the error rate for the validation data starts to increase



Classification Tree: Pruning

- Pruning



✓ CART lets tree grow to full extent, then prunes it back.

✓ Idea is to find that point at which the validation error begins to rise.

✓ Generate successively smaller trees by pruning leaves.

✓ At each pruning stage, multiple trees are possible.

✓ Use "cost complexity" to choose the best tree at that stage.

Classification Tree: Pruning

- Cost complexity

$$CC(T) = Err(T) + \alpha \times L(T)$$

✓ CC(T) = cost complexity of a tree

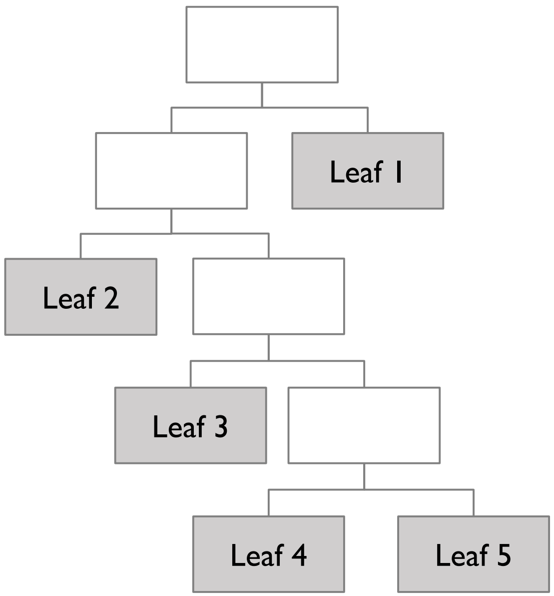
✓ ERR(T) = proportion of misclassified records in the validation data

✓ Alpha = penalty factor attached to the tree size (set by the user)

Classification Tree: Pruning

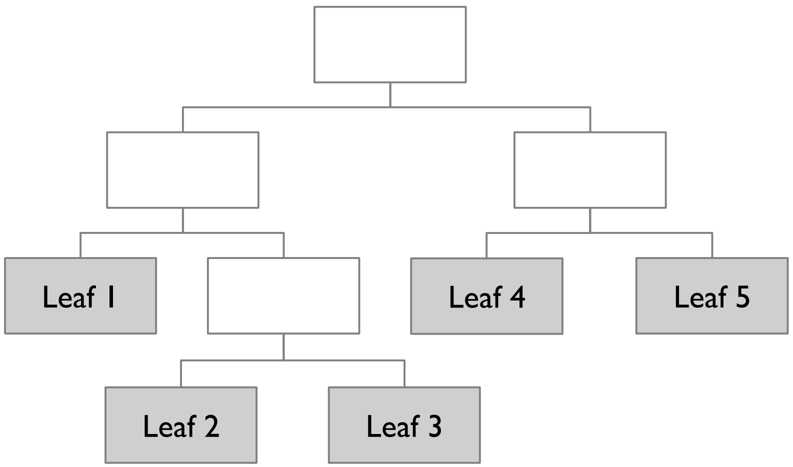
- Cost Complexity Example 1

Tree 1



Validation error = 10%

Tree 2

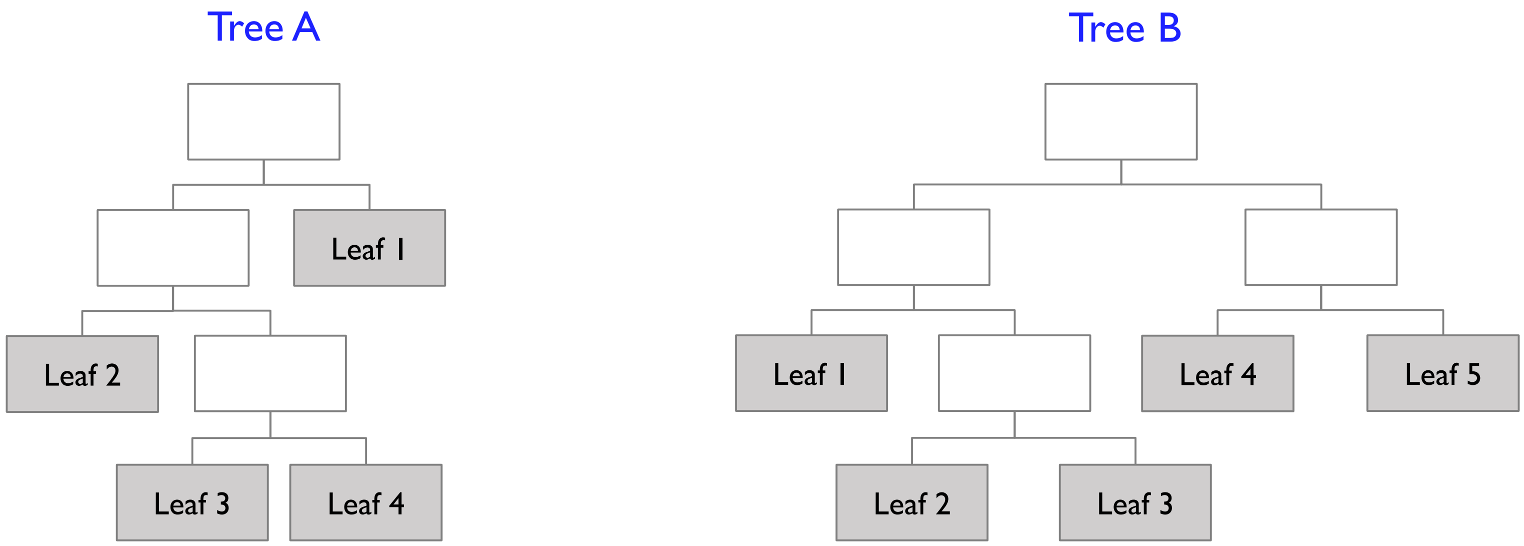


Validation error = 15%

- Two trees have the same number of leaf nodes but Tree 1 yields lower validation error → Tree 1 should be preferred

Classification Tree: Pruning

- Cost Complexity Example 2



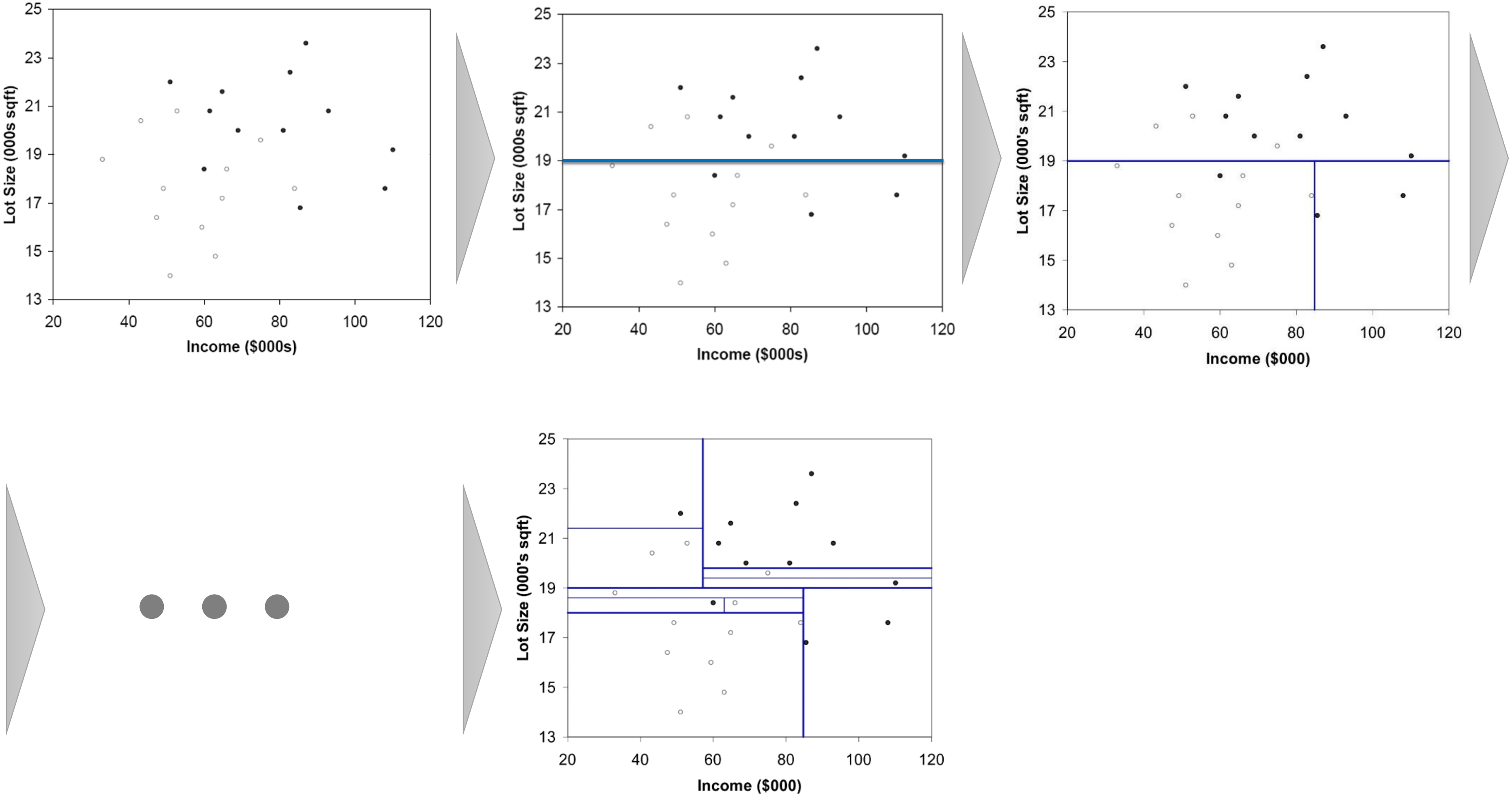
Validation error = 15%

Validation error = 15%

✓ Two trees yield the same validation error, but Tree A has fewer leaf nodes (simpler tree structure) → Tree A should be preferred

Classification Tree

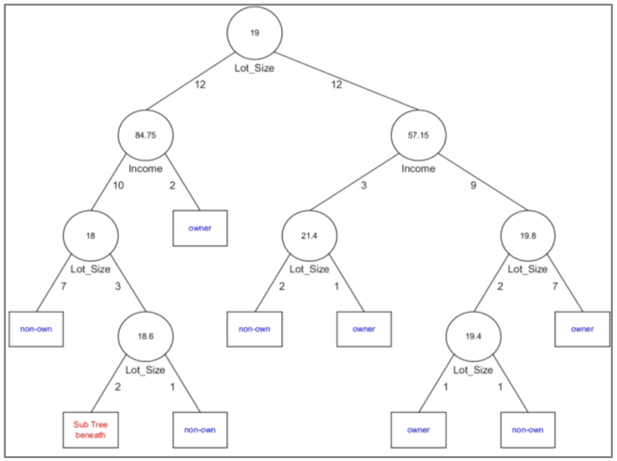
- Full tree constructed by recursive partitioning



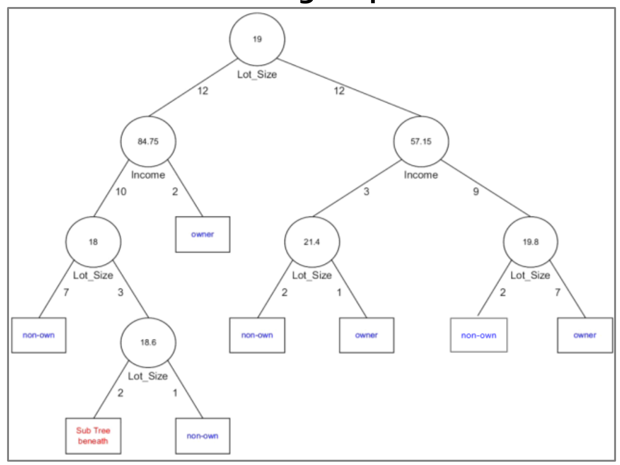
Classification Tree

- Pruning

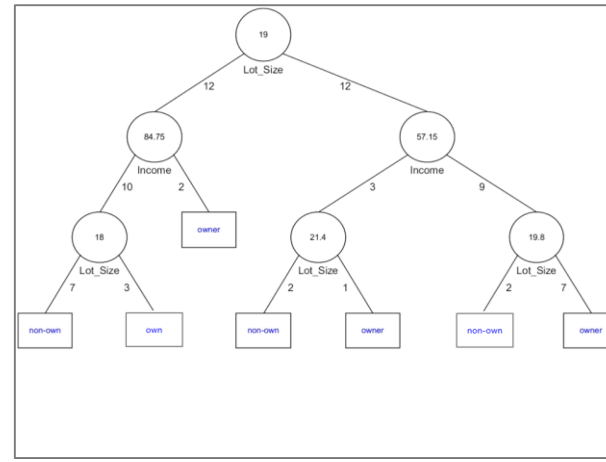
Full Tree

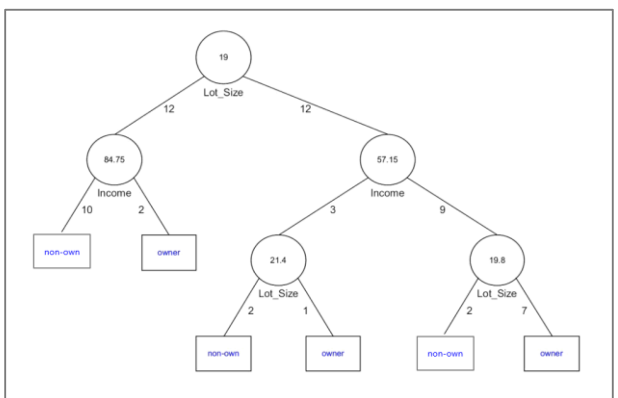


Pruning Step 1

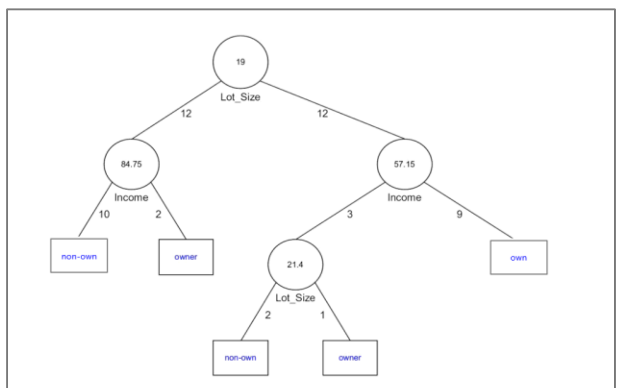


Pruning Step 2

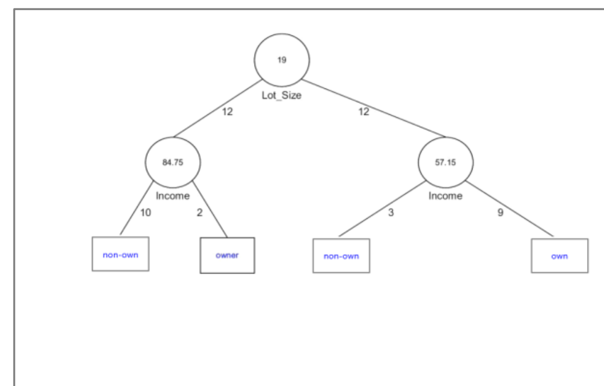




Pruning Step 3



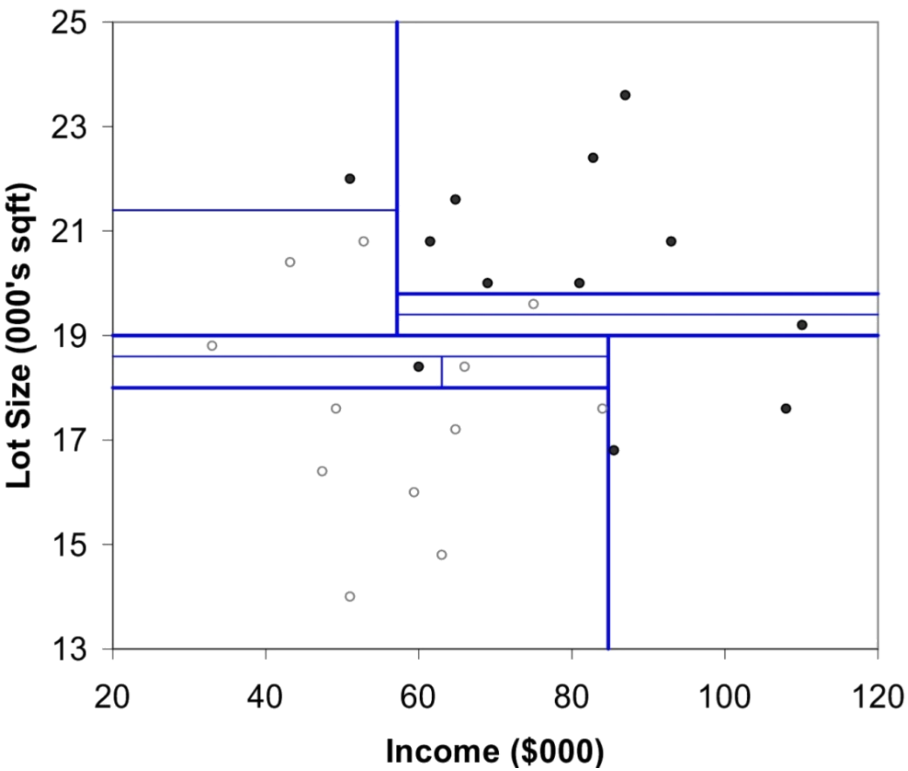
Pruning Step 4

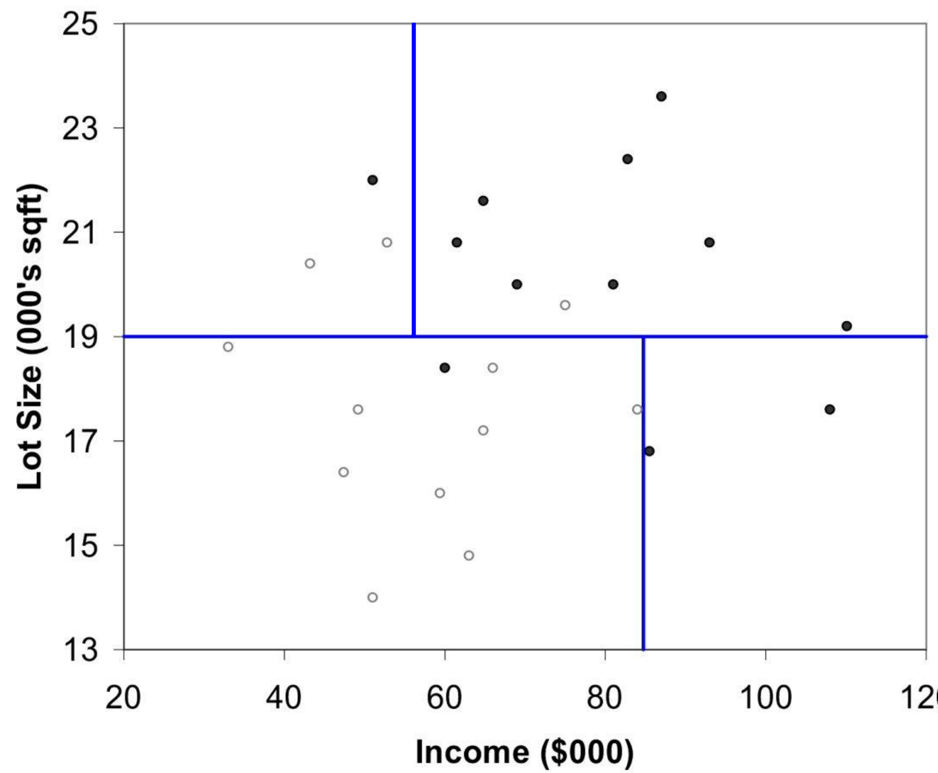


Pruned Tree

Classification Tree

- Full tree vs. Pruned tree





Classification Tree

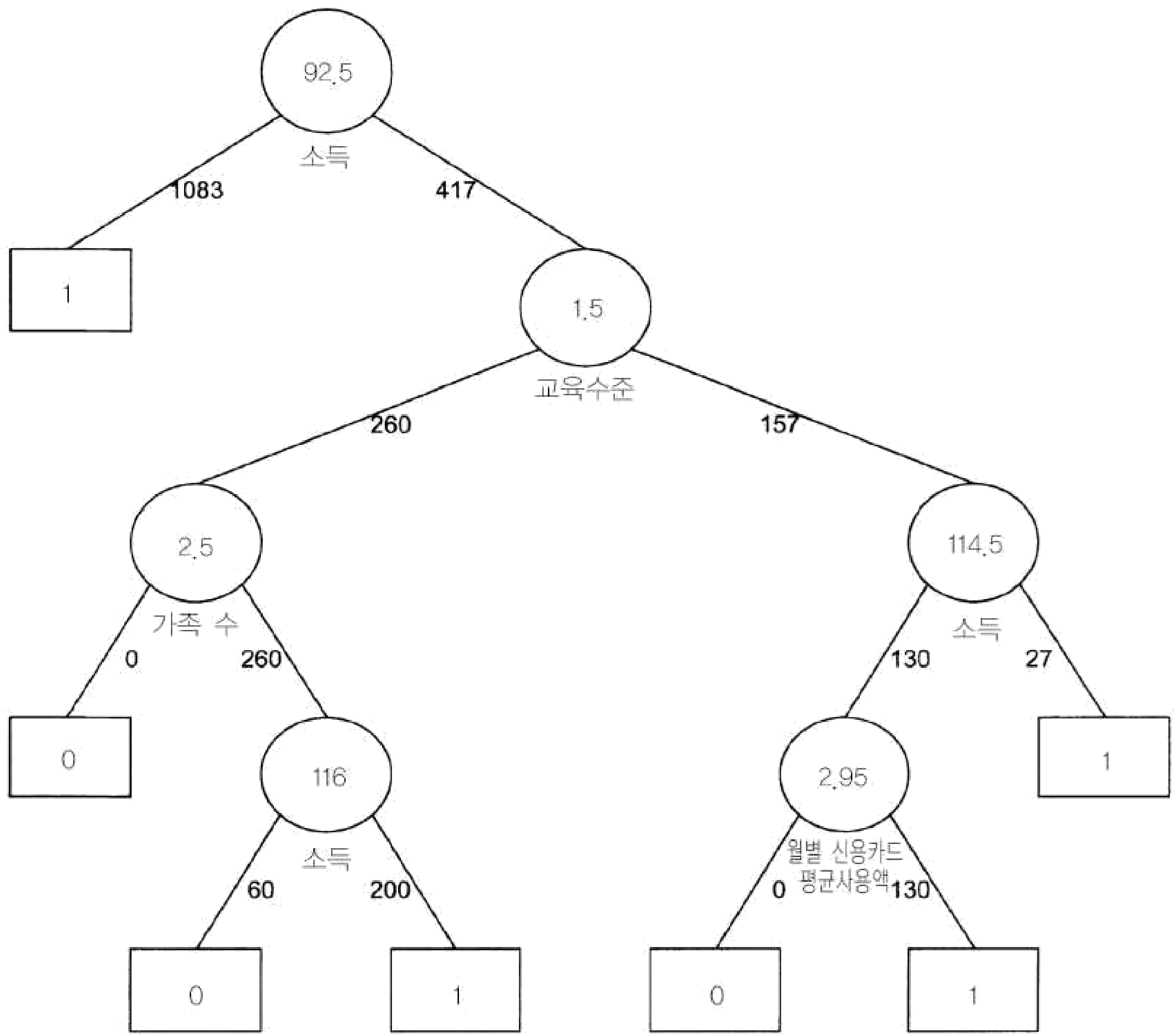
- Example: Universal bank
  - ✓ Goal: to analyze what combination of factors make a customer more likely to accept a personal loan

일련 번호	나이	경력	소득	가족 수	월별 신용카드 평균사용액	교육 수준	담보부 채권	개인 대출	증권 계좌	CD 계좌	온라인 뱅킹	신용 카드
1	25	1	49	4	1.60	UG	0	No	Yes	No	No	No
2	45	19	34	3	1.50	UG	0	No	Yes	No	No	No
3	39	15	11	1	1.00	UG	0	No	No	No	No	No
4	35	9	100	1	2.70	Grad	0	No	No	No	No	No
5	35	8	45	4	1.00	Grad	0	No	No	No	No	Yes
6	37	13	29	4	0.40	Grad	155	No	No	No	Yes	No
7	53	27	72	2	1.50	Grad	0	No	No	No	Yes	No
8	50	24	22	1	0.30	Prof	0	No	No	No	No	Yes
9	35	10	81	3	0.60	Grad	104	No	No	No	Yes	No
10	34	9	180	1	8.90	Prof	0	Yes	No	No	No	No
11	65	39	105	4	2.40	Prof	0	No	No	No	No	No
12	29	5	45	3	0.10	Grad	0	No	No	No	Yes	No
13	48	23	114	2	3.80	Prof	0	No	Yes	No	No	No
14	59	32	40	4	2.50	Grad	0	No	No	No	Yes	No
15	67	41	112	1	2.00	UG	0	No	Yes	No	No	No
16	60	30	22	1	1.50	Prof	0	No	No	No	Yes	Yes
17	38	14	130	4	4.70	Prof	134	Yes	No	No	No	No
18	42	18	81	4	2.40	UG	0	No	No	No	No	No
19	46	21	193	2	8.10	Prof	0	Yes	No	No	No	No
20	55	28	21	1	0.50	Grad	0	No	Yes	No	No	Yes

Classification Tree

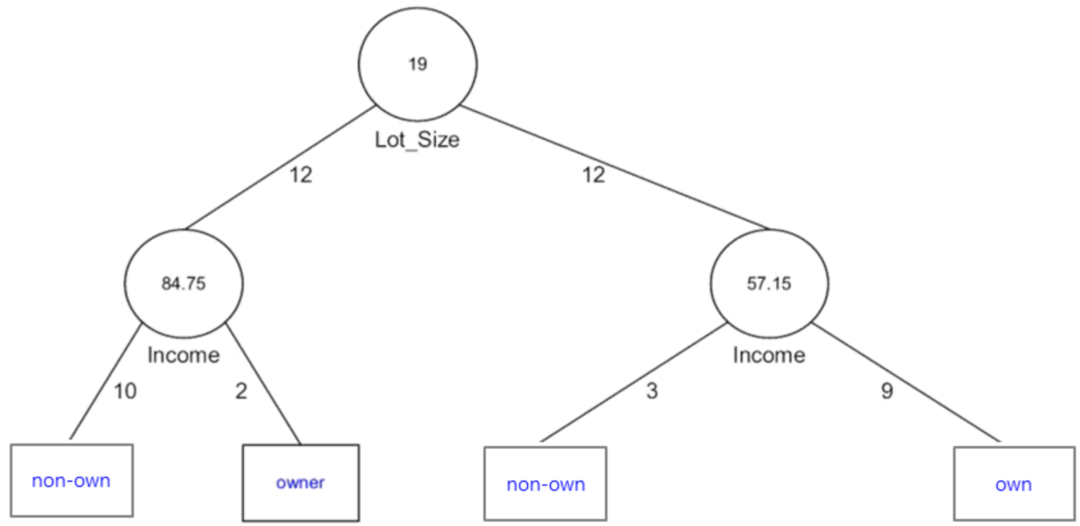
의사결정 마디	학습용 집합의 오차율	평가용 집합의 오차율			
41	0	2.133333			
40	0.04	2.2			
39	0.08	2.2			
38	0.12	2.2			
37	0.16	2.066667			
36	0.2	2.066667			
35	0.2	2.066667			
34	0.24	2.066667			
...	...	...			
13	1.16	1.6			
12	1.2	1.6			
11	1.2	1.466667	최소 오차 나무	표준오차	0.003103929
10	1.6	1.666667			
9	2.2	1.666667			
8	2.2	1.866667			
7	2.24	1.866667			
6	2.24	1.6	<-- 최적의 가지친 나무		
5	4.44	1.8			
4	5.08	2.333333			
3	5.24	3.466667			
2	9.4	9.533333			
1	9.4	9.533333			
0	9.4	9.533333			

Classification Tree



Classification Tree

- Generating the rules from the pruned tree

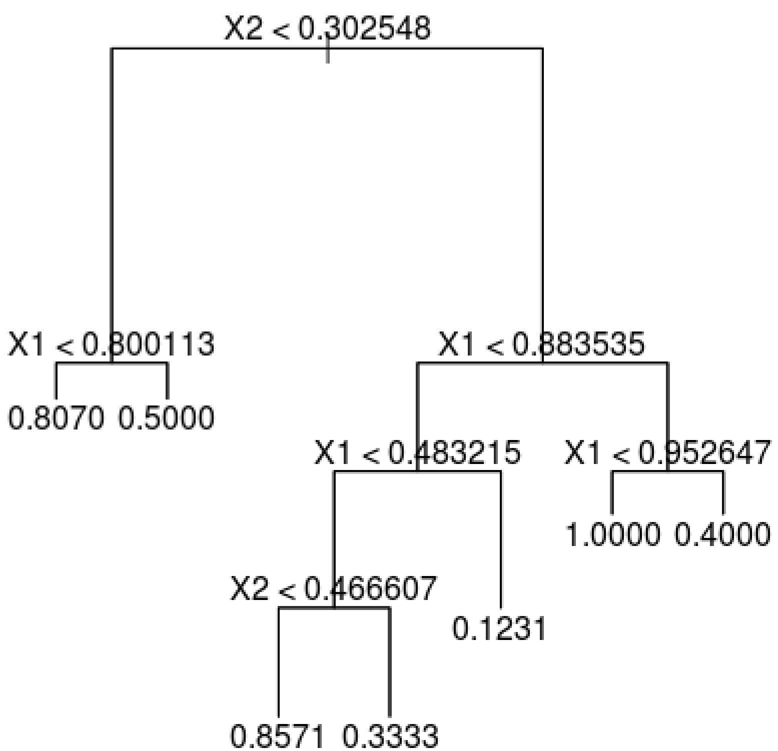


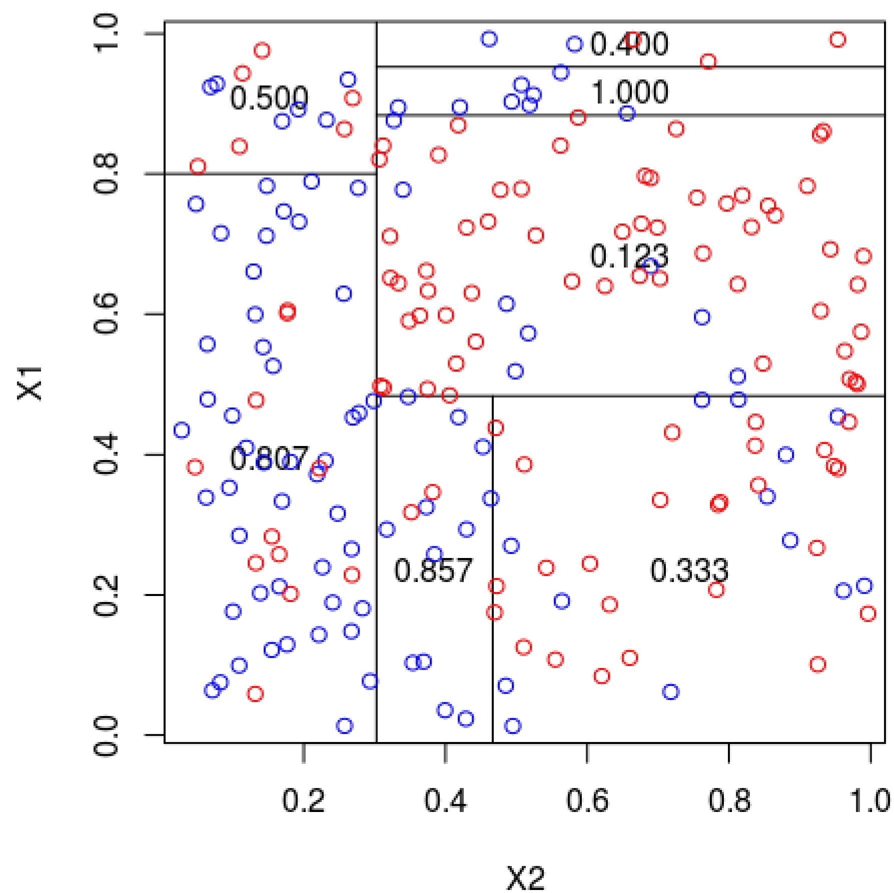
- IF(Lot size < 19) AND IF(Income < 84.75) THEN Owner = No
- IF(Lot size < 19) AND IF(Income > 84.75) THEN Owner = YES
- IF(Lot size > 19) AND IF(Income < 57.15) THEN Owner = NO
- IF(Lot size > 19) AND IF(Income > 57.15) THEN Owner = YES

AGENDA

Regression Tree

- The output of a leaf (terminal) node
  - ✓ The average of the target values of the observations in the node

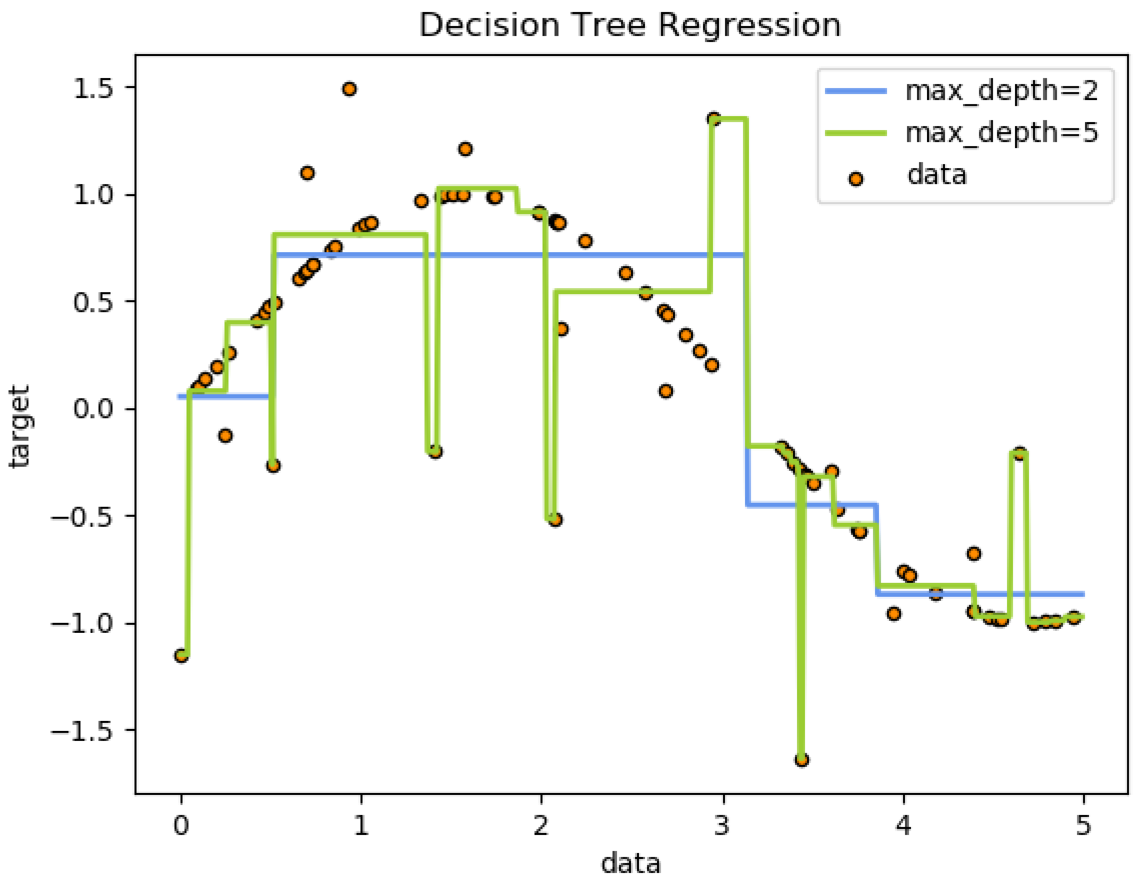




https://www.datacamp.com/community/tutorials/decision-trees-R

Regression Tree

- The output of a leaf (terminal) node
  - ✓ The average of the target values of the observations in the node
  - ✓ Regression tree example

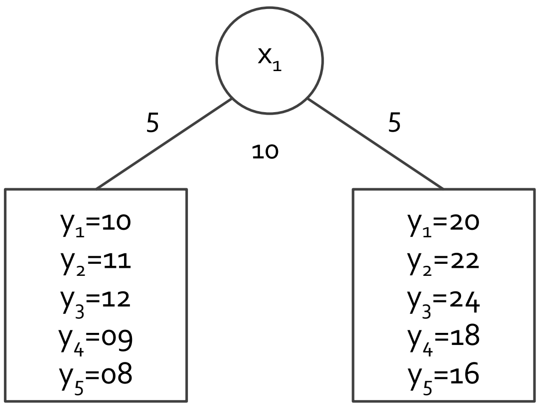


https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html#sphx-glr-auto-examples-tree-plot-tree-regression-py

Regression Tree

Similar process with classification tree except

- Prediction of the node
  - ✓ The average of the outcome variables belonging to the node



- Predicted value of the left leaf node = 10
- Predicted value of the right leaf node = 20

- Impurity
  - ✓ Sum of squared error (SSE: $\sum_{i=1}^{n}(y_i - \hat{y})^2$)
  - ✓ SSE(Parent) = 300, SSE(Left) = 10, SSE(Right) = 40, Gain = 250

Regression Tree

- Predict the selling price of Toyota corolla

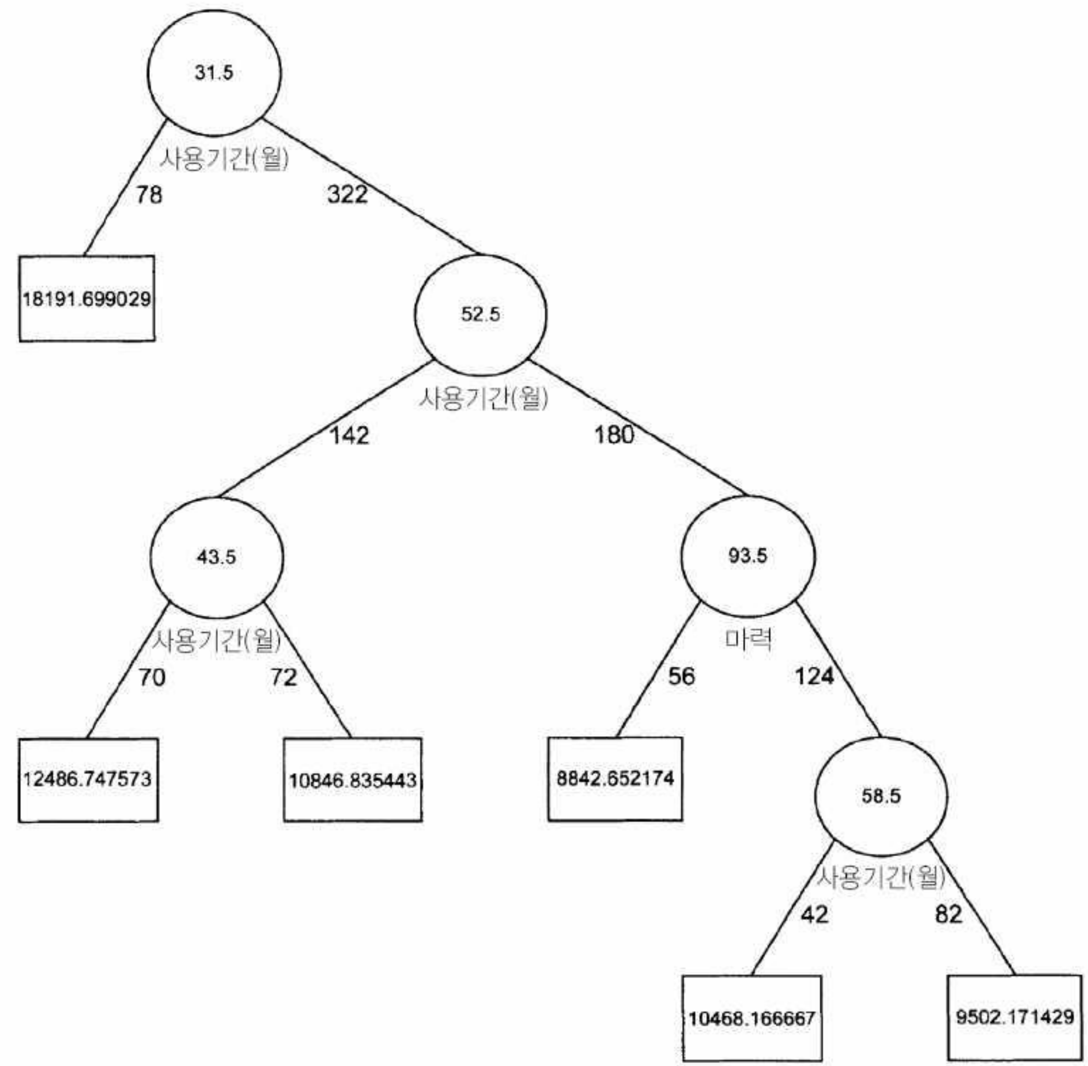

Dependent variable (target)

Independent variables (attributes, features)

Variable	Description
Price	Offer Price in EUROs
Age_08_04	Age in months as in August 2004
KM	Accumulated Kilometers on odometer
Fuel_Type	Fuel Type (Petrol, Diesel, CNG)
HP	Horse Power
Met_Color	Metallic Color? (Yes=1, No=0)
Automatic	Automatic ((Yes=1, No=0)
CC	Cylinder Volume in cubic centimeters
Doors	Number of doors
Quarterly_Tax	Quarterly road tax in EUROs
Weight	Weight in Kilograms

Regression Tree

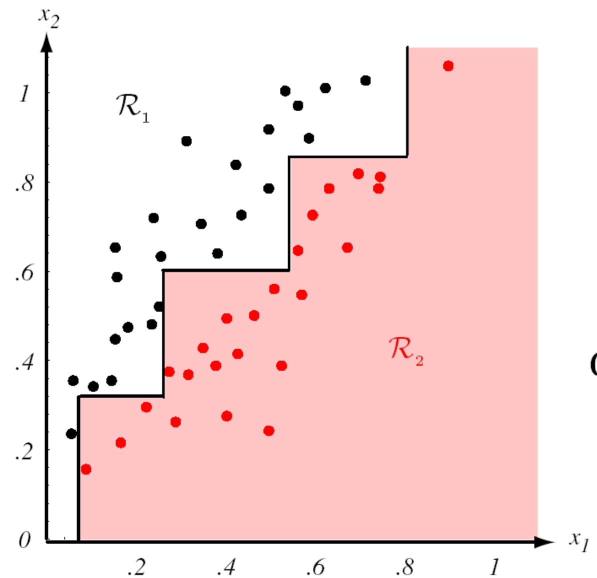
- Pruned Tree

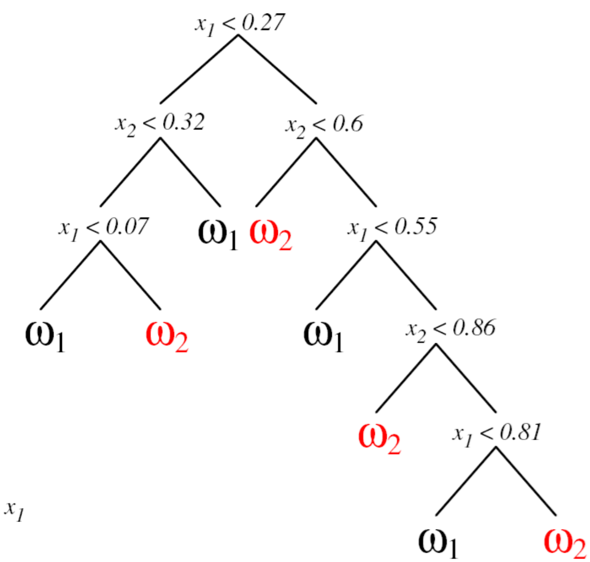


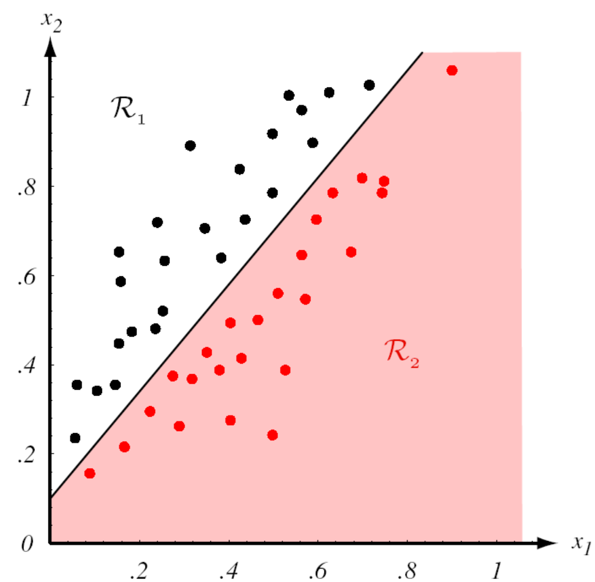
CART: Summary

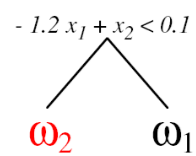
- Advantages
  - ✓ Classification and regression tree (CART) is easy to use and understand
  - ✓ Produce rules that are easy to interpret & implement
  - ✓ Variable selection & reduction is automatic
  - ✓ Do not require the assumptions of statistical models
  - ✓ Can work without extensive handling of missing data
- Disadvantages
  - ✓ May not perform well where there is structure in the data that is not well captured by horizontal or vertical split
  - ✓ Since the process deals with "one variable at a time", no way to capture interactions between variables

CART: Summary










ANY
questions?