Learning Analytics

Dr. Bowen Hui

Computer Science

University of British Columbia Okanagan

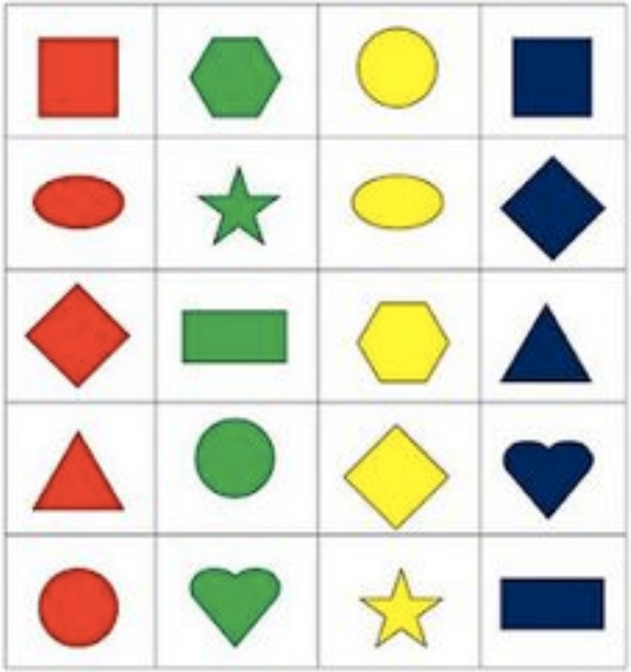


Image taken from www.pinterest.com

Image taken from www.ecoruraltrip.org.br

Categorization vs. Decision Making

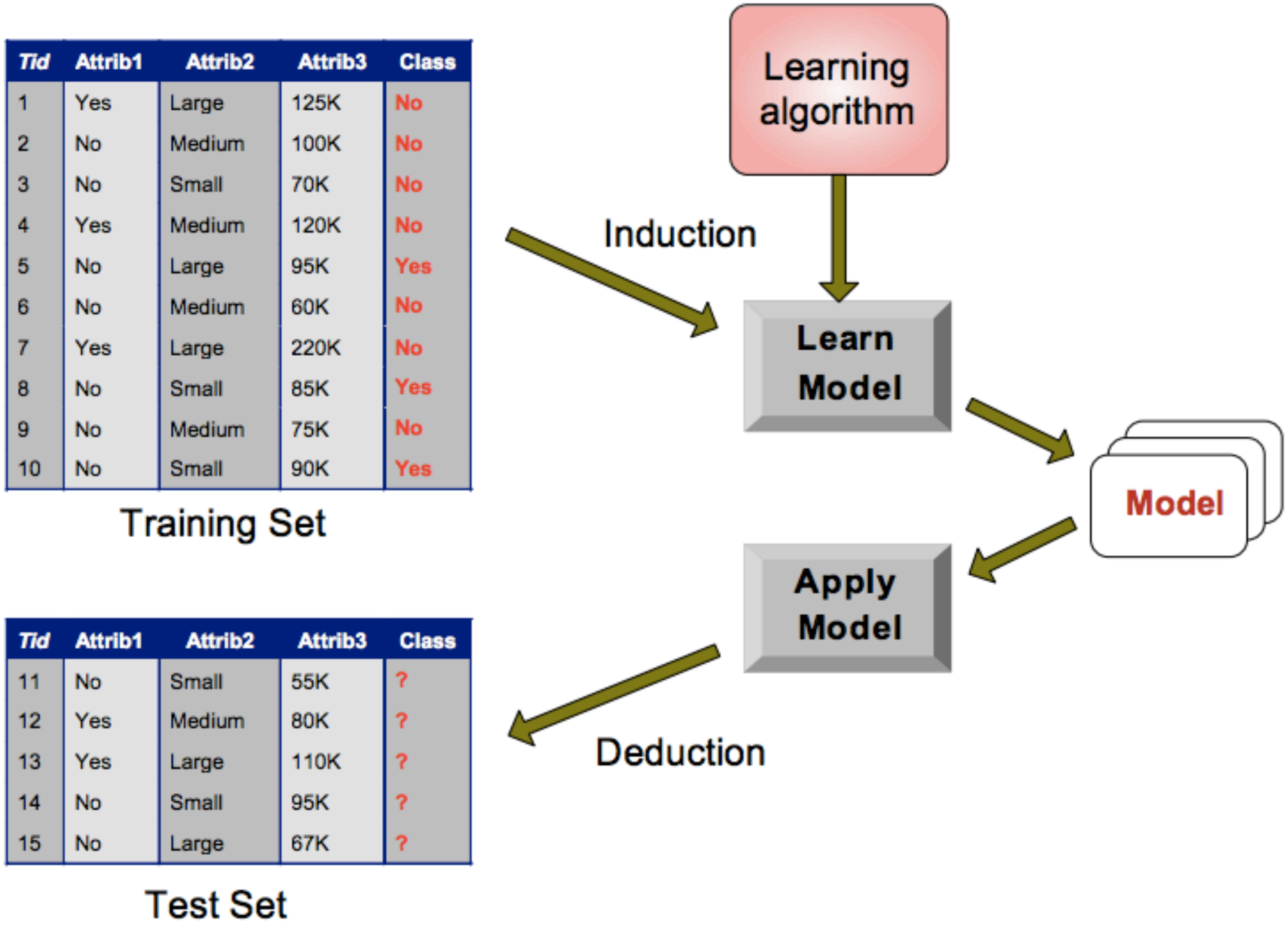
Group Exercise

- Form teams of 2, 3, or 4
- Choose a topic to discuss
- Example topics (or pick others):
  - Which job offer should I accept?
  - Should I date while in school?
  - Should I study over midterm break or relax at home for the week?
- Consider possible choices, possible outcomes, and long term consequences

Machine Learning Terminology

- Categorization just means putting things into different groups in some way
- Classification
  - Supervised learning task
  - All data requires labels of what the “right” answer is
  - Based on labeled dataset, learn the model underlying the data and predict labels for unseen data
- Clustering
  - Unsupervised learning task
  - Data does not have labels
  - Based on unlabeled dataset, discover the model underlying the data to find labels for each group

Classification Process



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

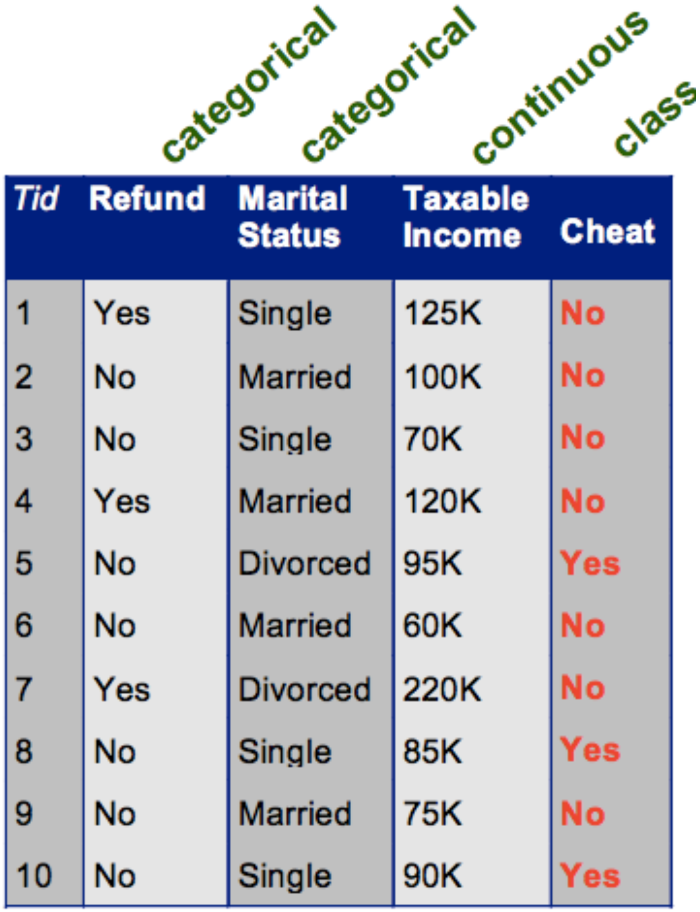
Example Classification Tasks

- Predict handwritten digit (0,..,9)
- Classify credit card transaction as fraudulent or not
- Predict if an email is spam or not
- Classify webpages into topics
- Classify students activities/performance into letter grades

Example: Tax Evasion

Data is usually in CSV format

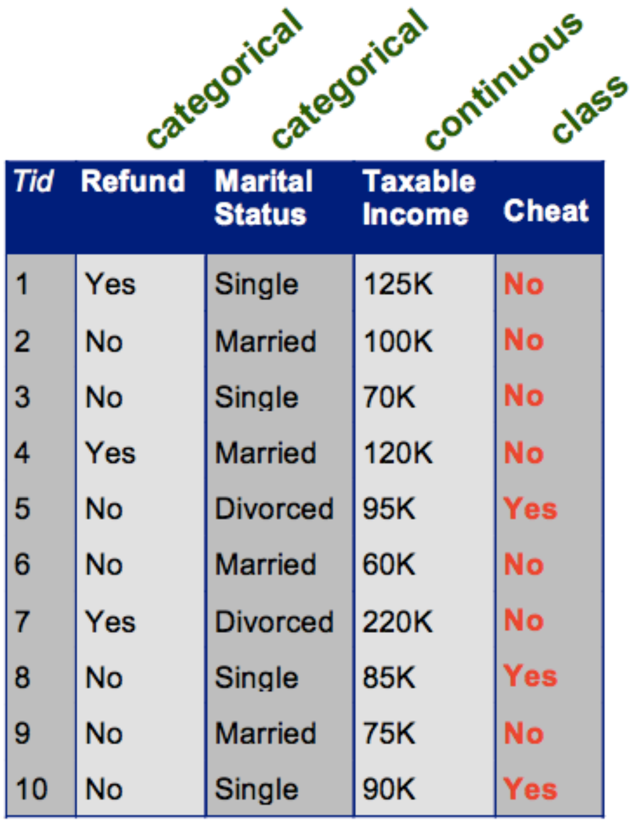
- Each row is a data point
- Each column is an attribute
- Last column is the class label



	categorical	categorical	continuous	class
Tid	**Refund**	**Marital Status**	**Taxable Income**	**Cheat**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

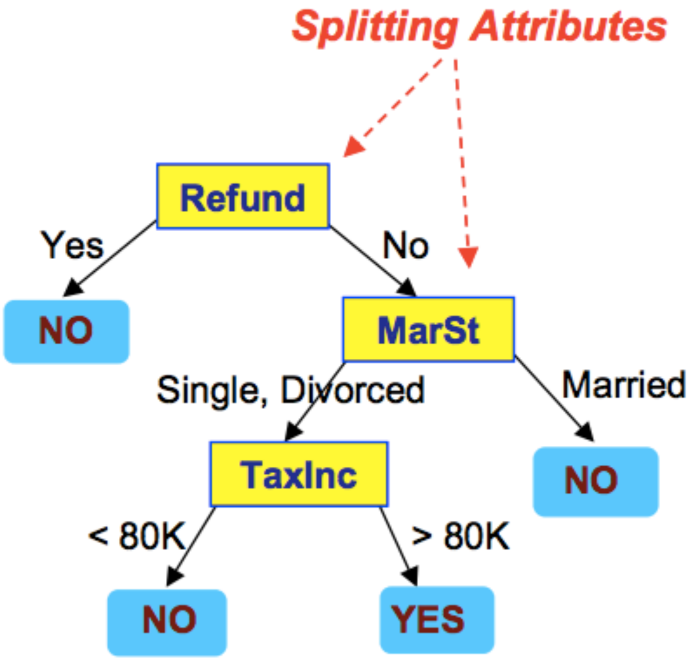
Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Example: Tax Evasion



	categorical	categorical	continuous	class
Tid	**Refund**	**Marital Status**	**Taxable Income**	**Cheat**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

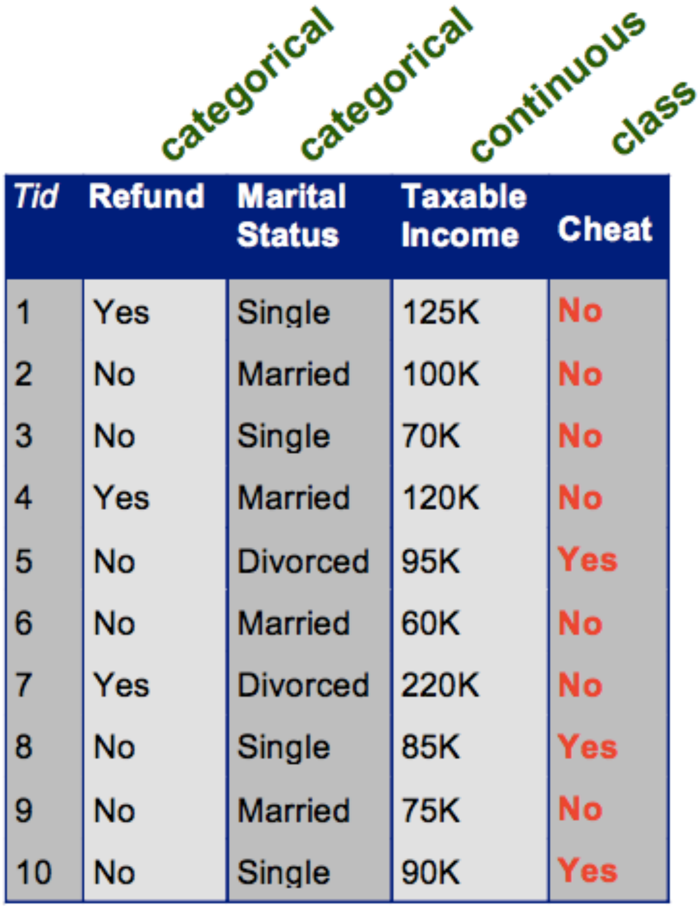
Training Data



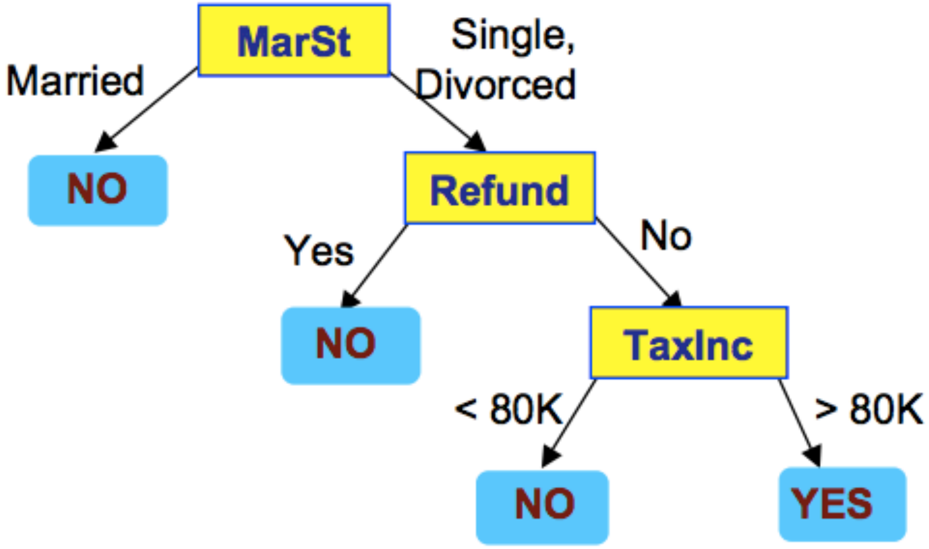
Model: Decision Tree

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Example: Tax Evasion



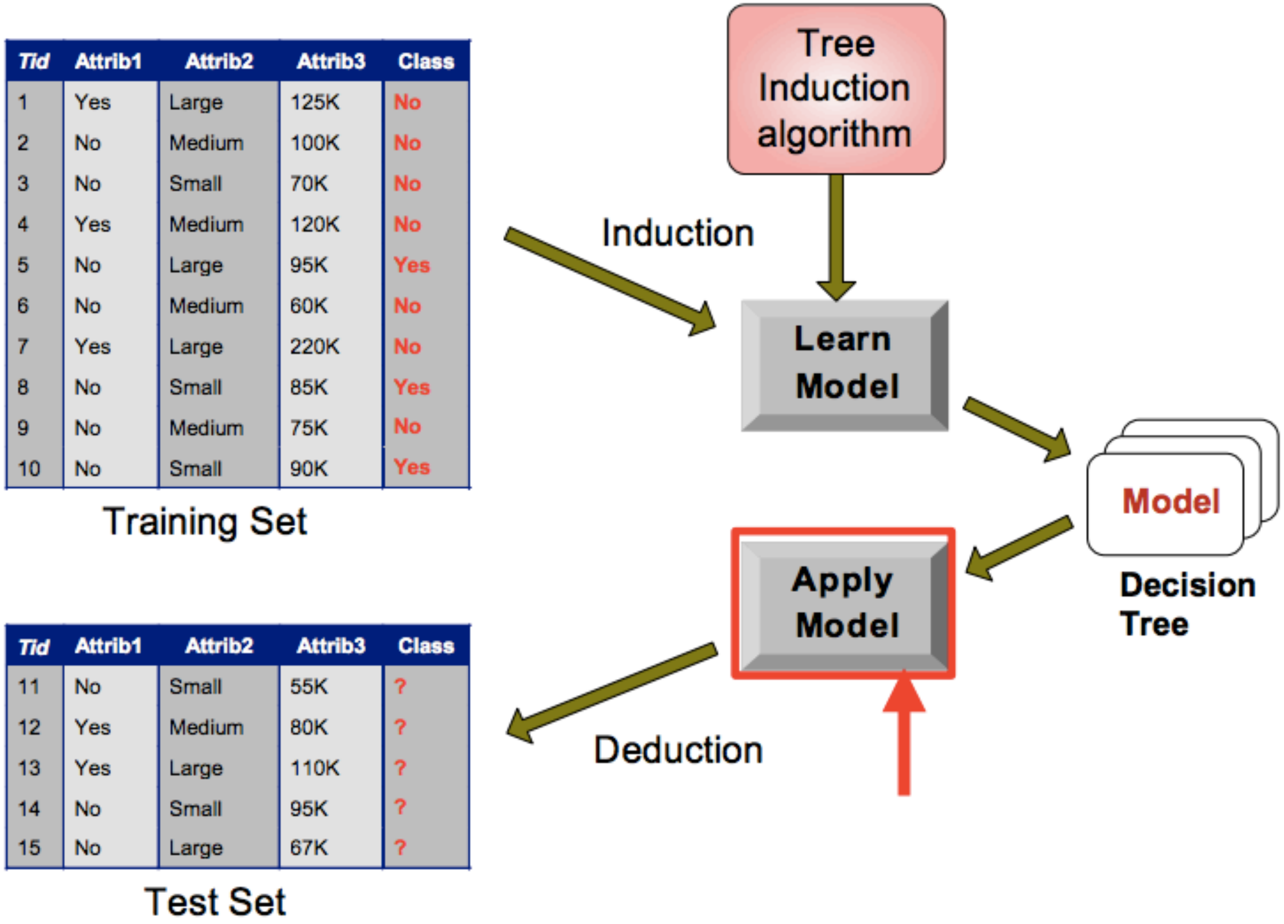
	categorical	categorical	continuous	class
Tid	**Refund**	**Marital Status**	**Taxable Income**	**Cheat**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



There could be more than one tree that fits the same data!

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Example: Tax Evasion



Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

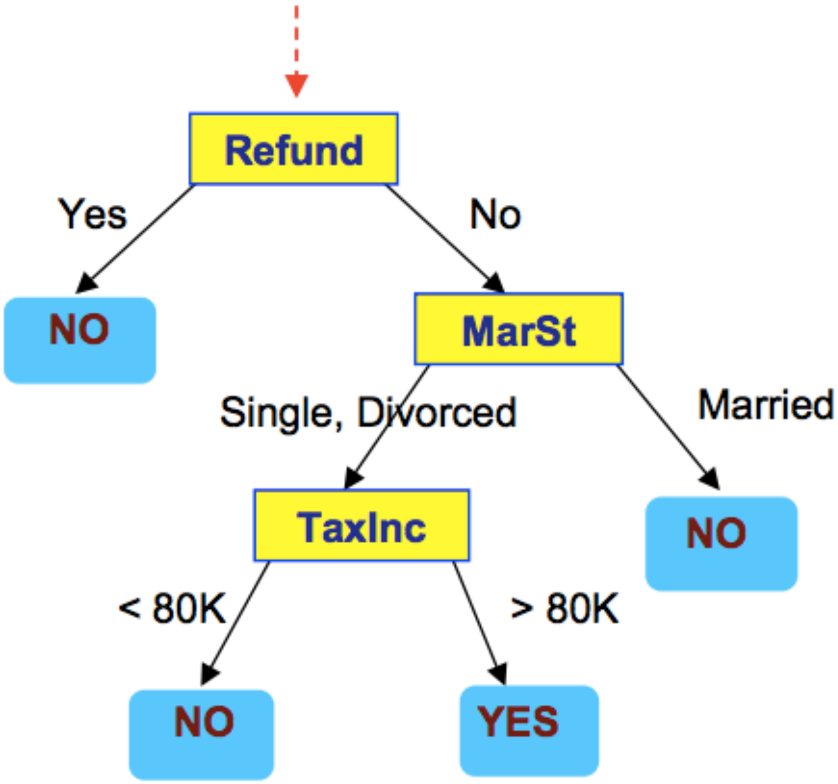
Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Example: Tax Evasion



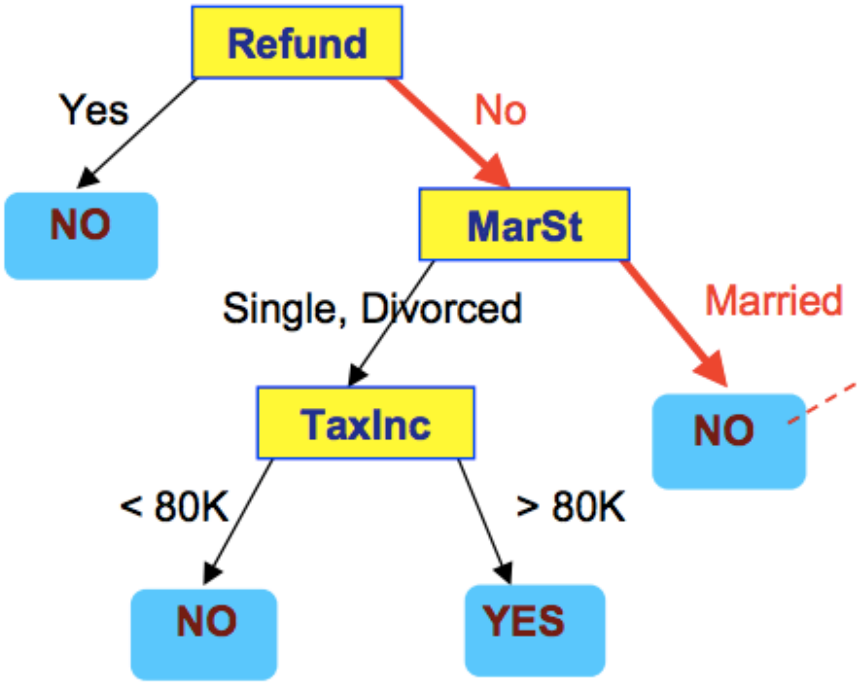
Start from the root of tree.

Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Example: Tax Evasion



Test Data

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

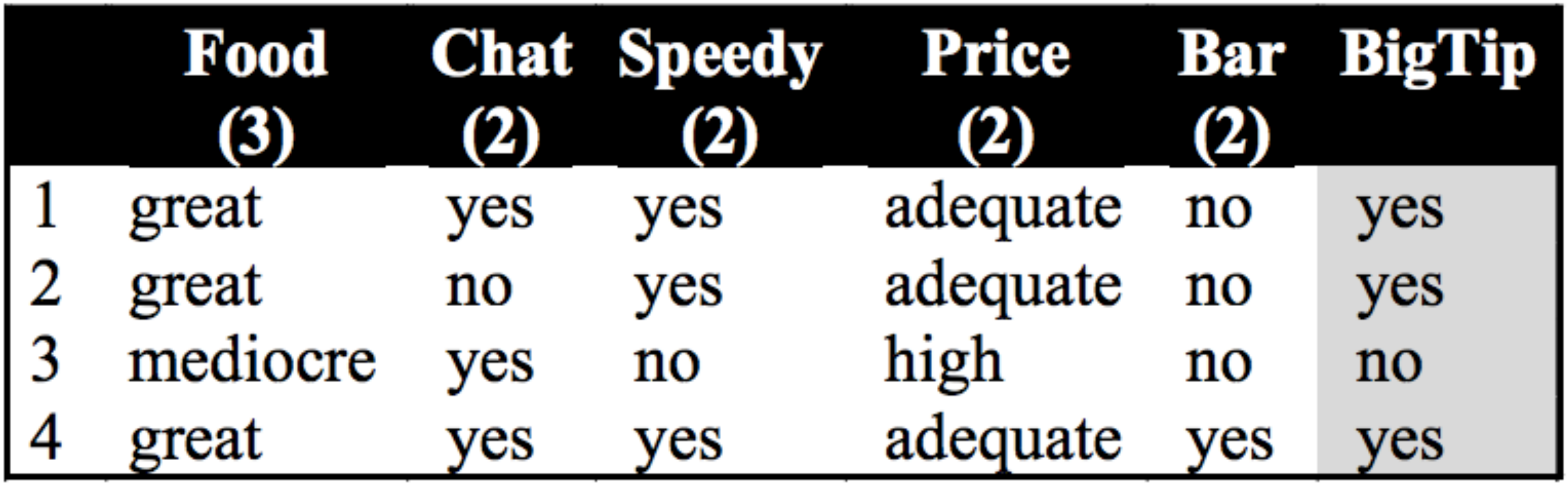
Assign “No” to Cheat

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

How to Build a Decision Tree?

- ID3 (iterative dichotomiser 3)
  - Developed in 1986 by Ross Quinlan
  - Builds multiway trees
- C4.5 (successor to ID3)
  - Improves various algorithmic restrictions
- C5.0 (latest version)
  - Proprietary – uses less memory and smaller trees while being more accurate
- CART (classification and regression trees)
  - Similar to C4.5
  - Constructs binary trees

Example: Restaurant Tips

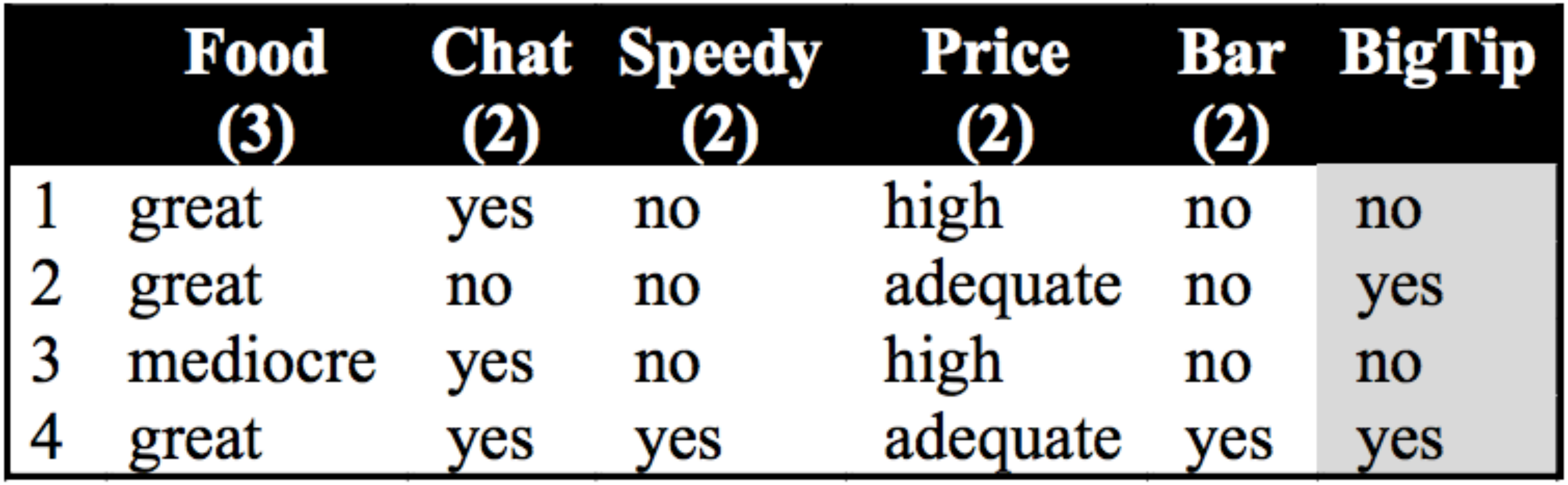


	Food (3)	Chat (2)	Speedy (2)	Price (2)	Bar (2)	BigTip
1	great	yes	yes	adequate	no	yes
2	great	no	yes	adequate	no	yes
3	mediocre	yes	no	high	no	no
4	great	yes	yes	adequate	yes	yes

Image taken from http://www.cs.cornell.edu/courses/cs4700/2011fa/lectures/09_decision_trees.pdf

Can you find a simple DT to explain this data?

Example: Restaurant Tips 2



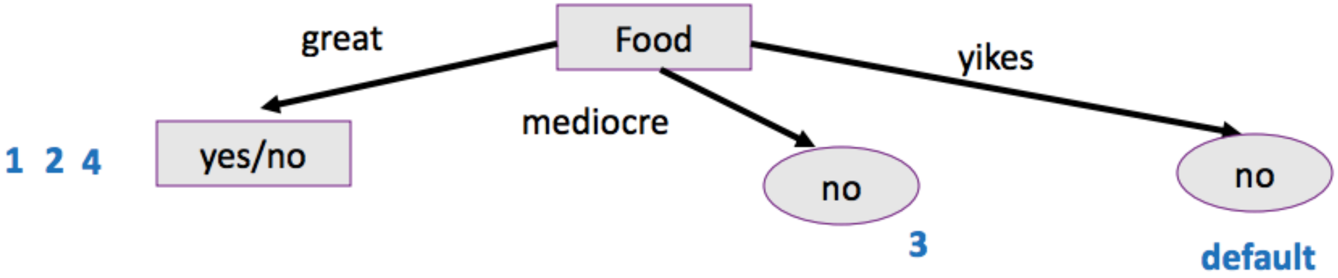
	Food	Chat	Speedy	Price	Bar	BigTip
	(3)	(2)	(2)	(2)	(2)	
1	great	yes	no	high	no	no
2	great	no	no	adequate	no	yes
3	mediocre	yes	no	high	no	no
4	great	yes	yes	adequate	yes	yes

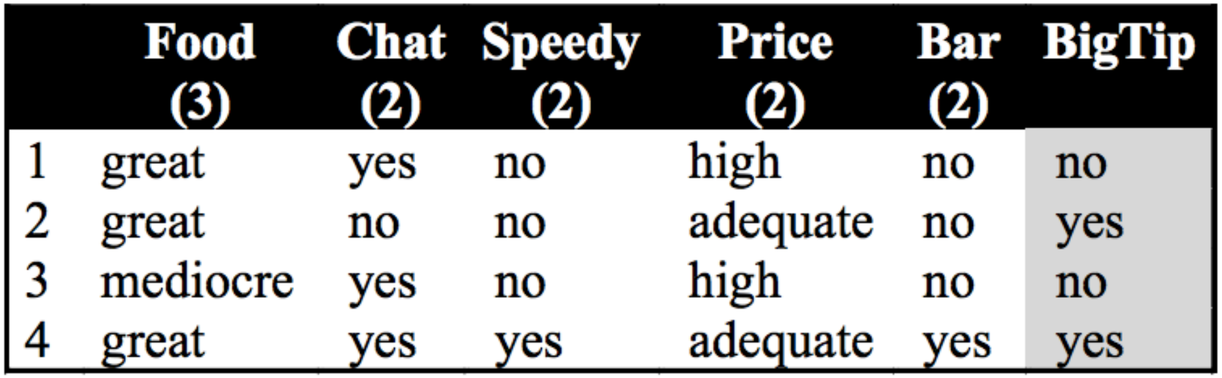
Image taken from http://www.cs.cornell.edu/courses/cs4700/2011fa/lectures/09_decision_trees.pdf

How to derive a tree?

Which attribute to split on? What’s left? When to stop?

Example: Restaurant Tips 2

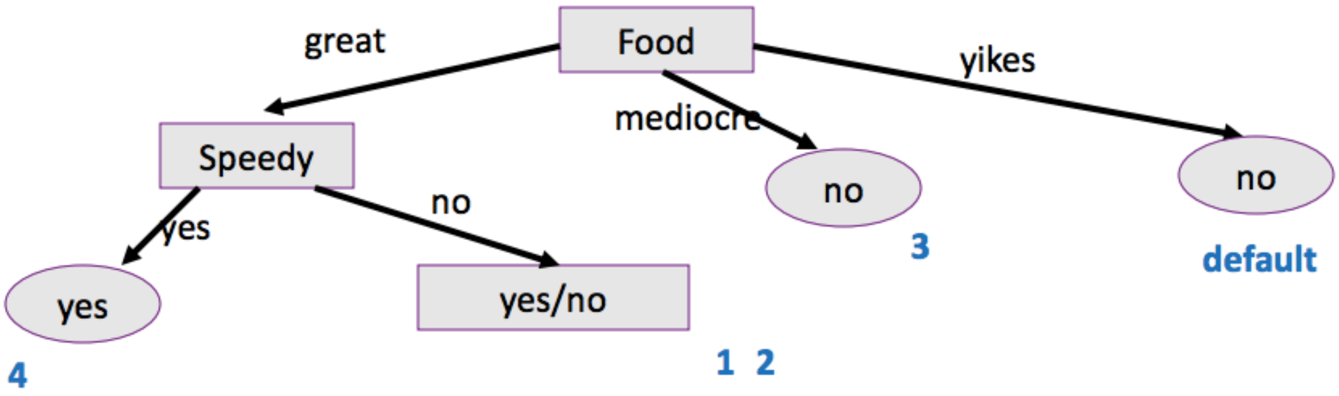


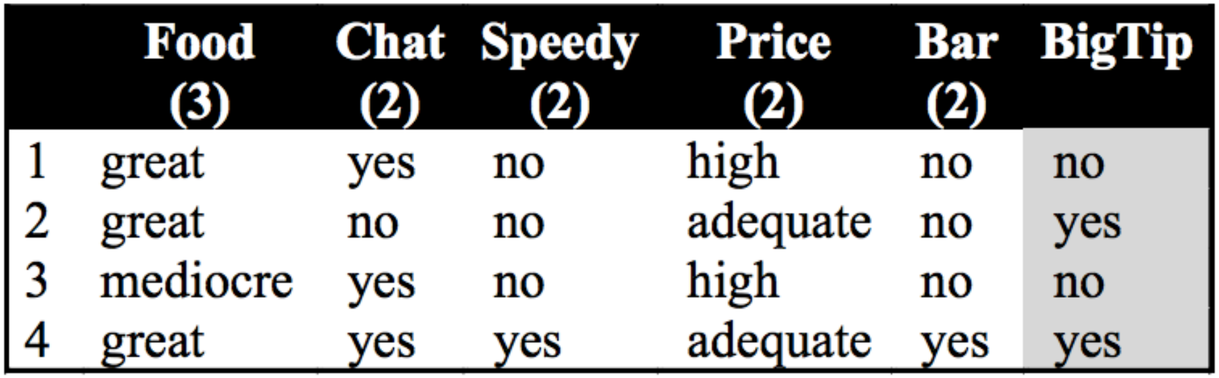


	Food	Chat	Speedy	Price	Bar	BigTip
	(3)	(2)	(2)	(2)	(2)	
1	great	yes	no	high	no	no
2	great	no	no	adequate	no	yes
3	mediocre	yes	no	high	no	no
4	great	yes	yes	adequate	yes	yes

Image taken from http://www.cs.cornell.edu/courses/cs4700/2011fa/lectures/09_decision_trees.pdf

Example: Restaurant Tips 2

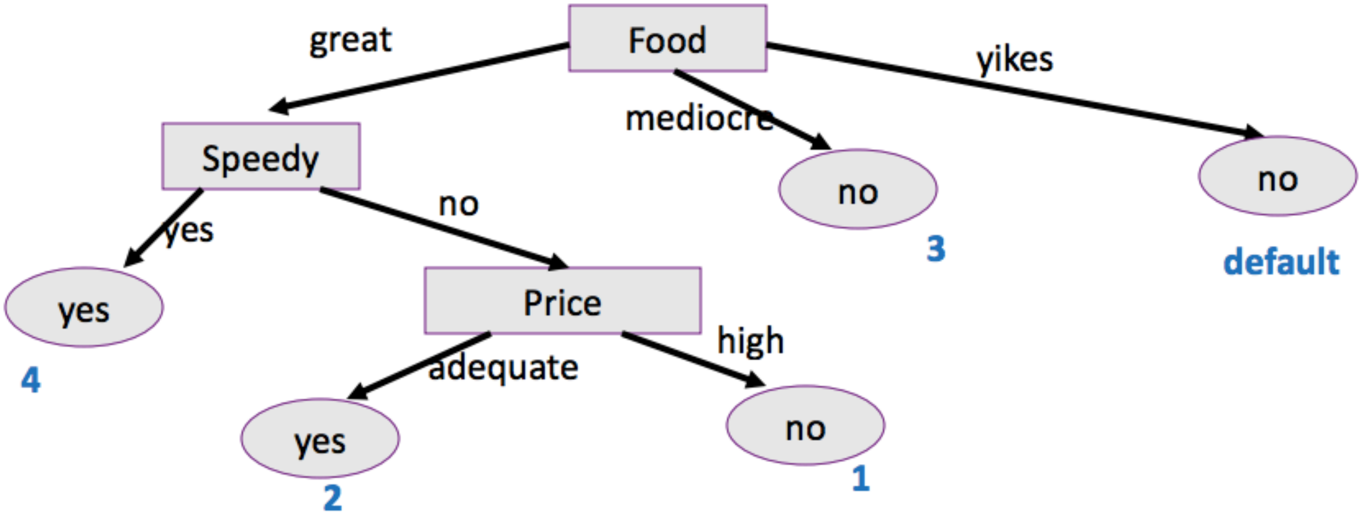




	Food (3)	Chat (2)	Speedy (2)	Price (2)	Bar (2)	BigTip
1	great	yes	no	high	no	no
2	great	no	no	adequate	no	yes
3	mediocre	yes	no	high	no	no
4	great	yes	yes	adequate	yes	yes

Image taken from http://www.cs.cornell.edu/courses/cs4700/2011fa/lectures/09_decision_trees.pdf

Example: Restaurant Tips 2



	Food (3)	Chat (2)	Speedy (2)	Price (2)	Bar (2)	BigTip
1	great	yes	no	high	no	no
2	great	no	no	adequate	no	yes
3	mediocre	yes	no	high	no	no
4	great	yes	yes	adequate	yes	yes

Image taken from http://www.cs.cornell.edu/courses/cs4700/2011fa/lectures/09_decision_trees.pdf

General Induction Algorithm

- Pick an attribute and find the “best” split

- Repeat until subtree data belongs to same class

General Induction Algorithm

- Pick an attribute and find the “best” split
  - For each attribute:
    - Find the split that yields the largest information gain
- Repeat until subtree data belongs to same class

Intuition Behind Finding a Split

- 20 question game:
  - I choose a number between 1 and 1000
  - You ask a series of yes/no questions
  - Which question would you rather ask?
    - Is the number 500?
    - Is the number prime?
    - Is the number smaller than 500?

- Find a question that is most informative

General Induction Algorithm

- Pick an attribute and find the “best” split
  - If no attributes left, return the most common label
  - For each attribute:
    - Find the split that yields the largest information gain
- Repeat until subtree data belongs to same class

General Induction Algorithm

- Pick an attribute and find the “best” split
  - If no attributes left, return the most common label
  - For each attribute:
    - Find the split that yields the largest information gain
- Repeat until subtree data belongs to same class
- Resulting tree may require pruning
  - Returns a smaller tree
  - Better able to generalize to unseen data

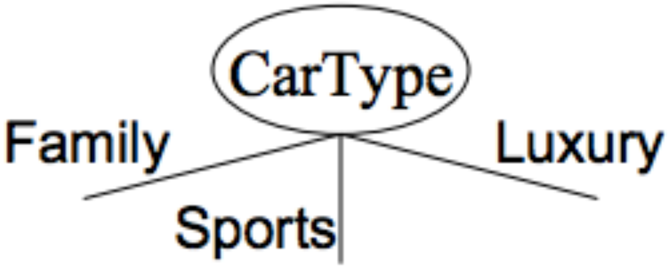
Specifying the Split

- Depends on attribute type:
  - Nominal (discrete variable, no order)
  - Ordinal (discrete variable, with order)
  - Continuous

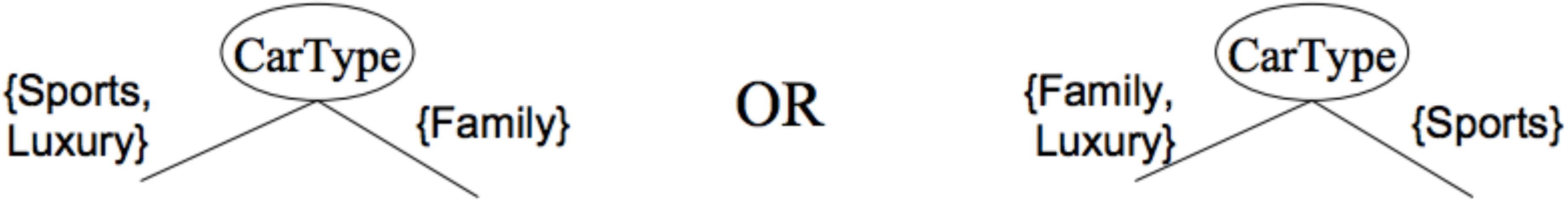
- Depends on number of splits:
  - Binary split
  - Multiway split

Splitting on Nominal Attributes

- **Multiway split:** Use as many partitions as distinct values

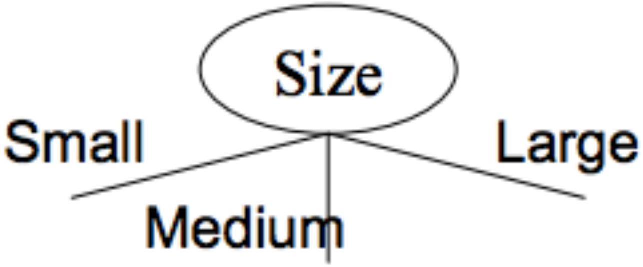


- **Binary split:** Find best partition of subset of values

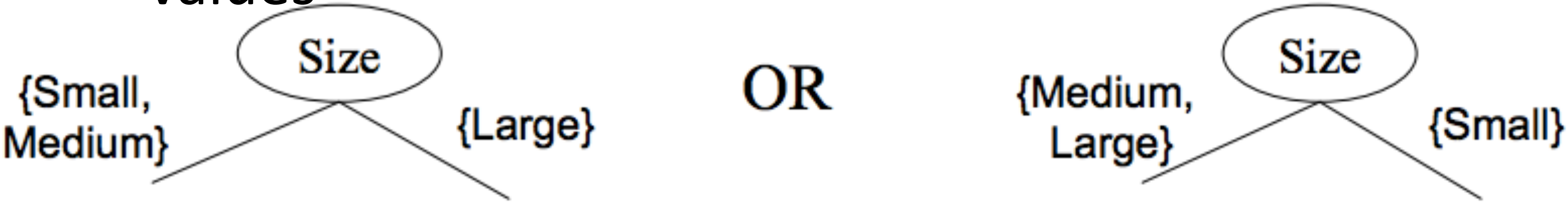


Splitting on Ordinal Attributes

- **Multiway split:** Use as many partitions as distinct values

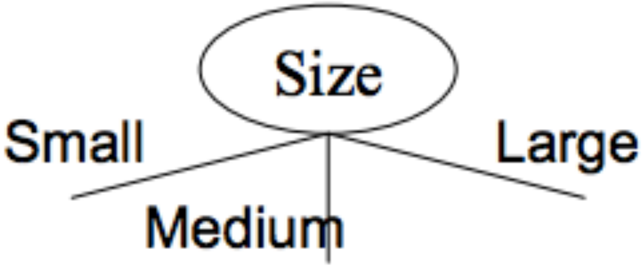


- **Binary split:** Find best partition of subset of values

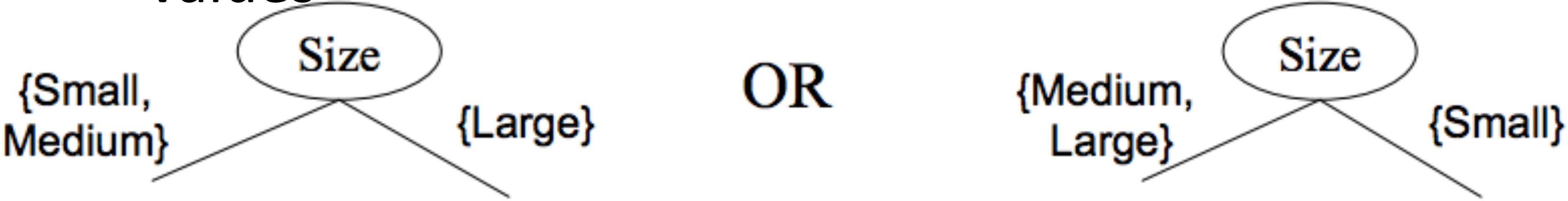


Splitting on Ordinal Attributes

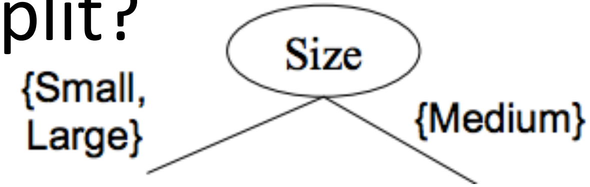
- **Multiway split:** Use as many partitions as distinct values



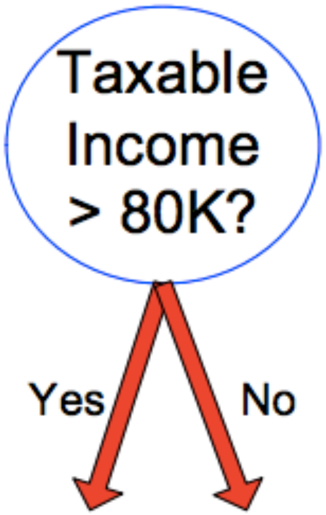
- **Binary split:** Find best partition of subset of values



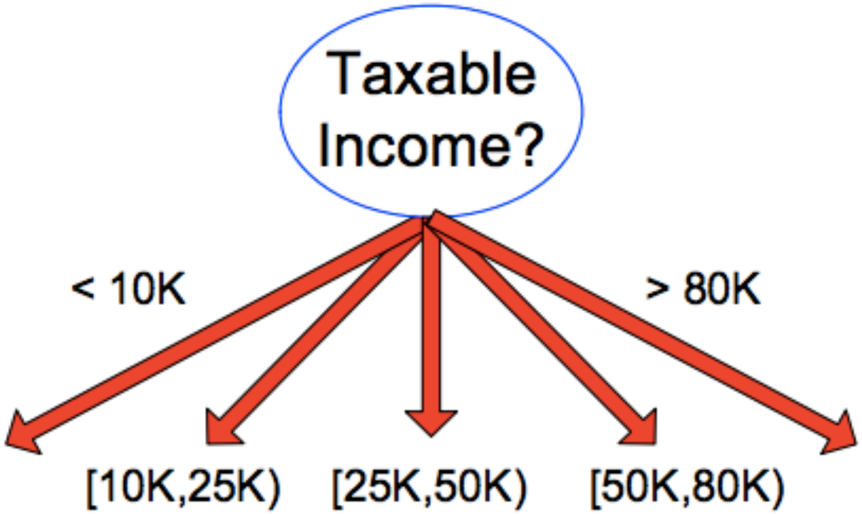
- What about this split?



Splitting on Continuous Attributes



(i) Binary split



(ii) Multi-way split

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Splitting on Continuous Attributes

- Discretization to form an ordinal categorical attribute
  - Static – discretize once at start of algorithm
  - Dynamic – ranges can be found using various methods (e.g. equal interval bucketing, percentiles, clustering)

- Binary decision: (A < v) or (A >= v)
  - Consider all possible splits and find best cut
  - More computationally intensive

Picking the Best Attribute to Split

- Occam's razor
  – All else being equal, choose the simplest explanation

- Decision tree induction
  – Find the smallest tree that classifies the training data correctly
  – Problem: Finding an optimal tree is NP-hard
  – Solution: Use a heuristic approach (e.g. greedy algorithm)

Visualizing the Change in Split

What is the information gained after a split?

Before Splitting: 10 records of class 0,
10 records of class 1

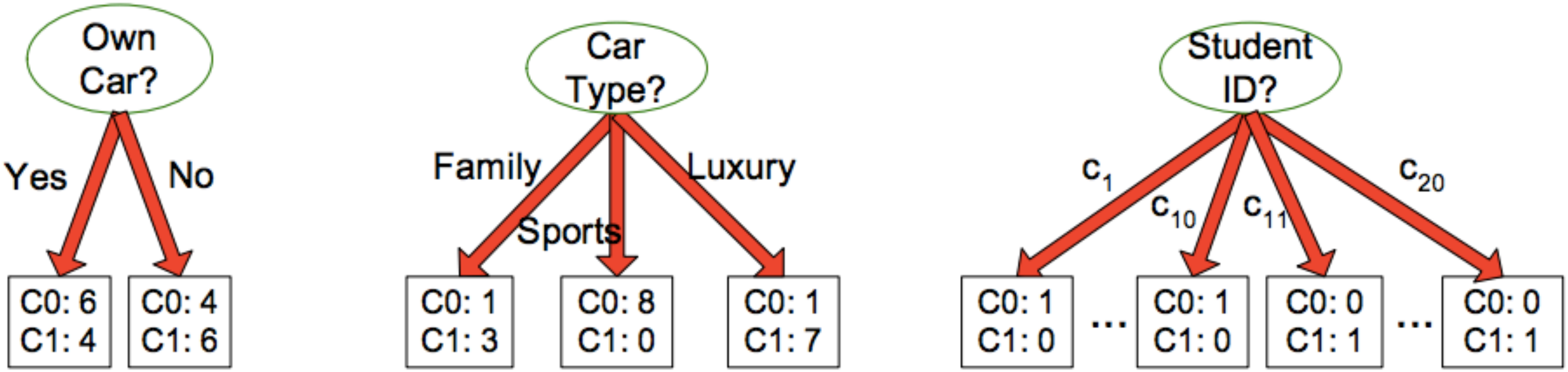


Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

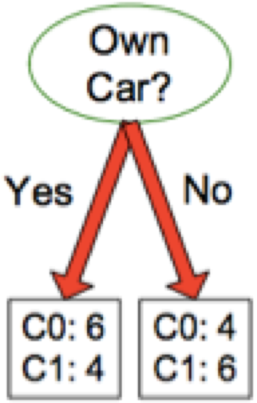
Visualizing the Change in Split (cont.)

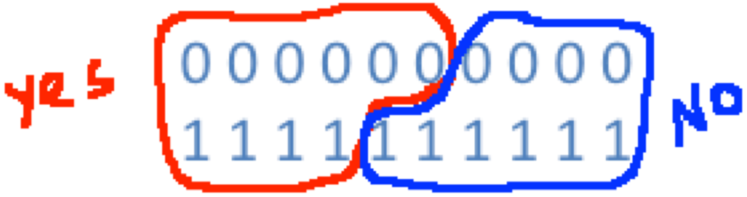
Before Splitting:

0 0 0 0 0 0 0 0 0 0

1 1 1 1 1 1 1 1 1 1

Split by this:





yes

0 0 0 0 0 0 0 0 0 0

1 1 1 1 1 1 1 1 1 1

No

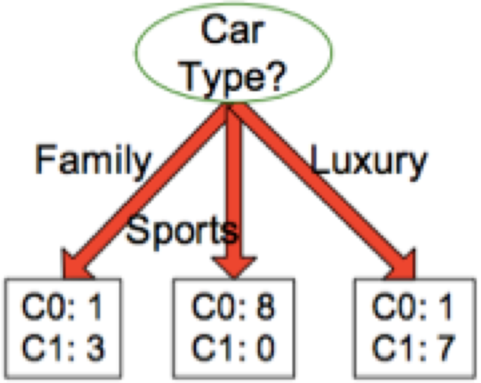
Visualizing the Change in Split (cont.)

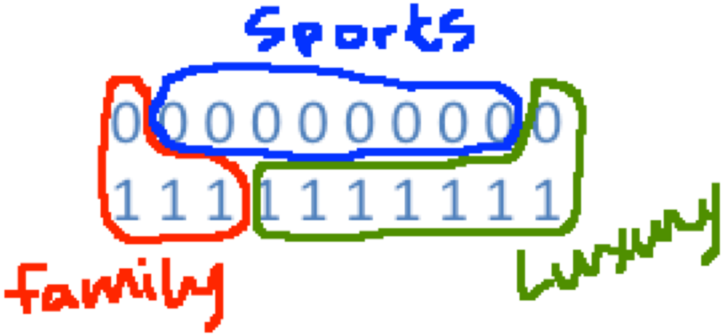
Before Splitting:

0 0 0 0 0 0 0 0 0 0

1 1 1 1 1 1 1 1 1 1

Split by this:





0 0 0 0 0 0 0 0 0 0

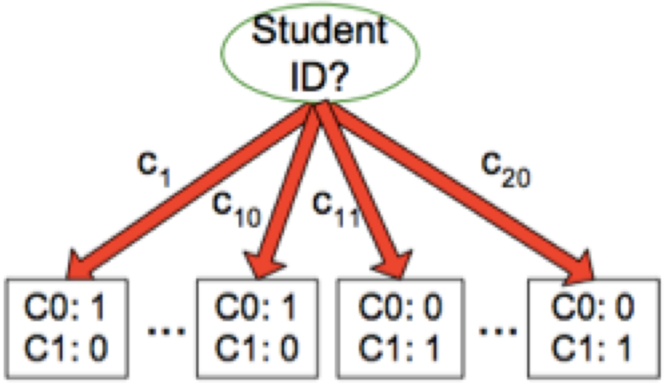
1 1 1 1 1 1 1 1 1 1

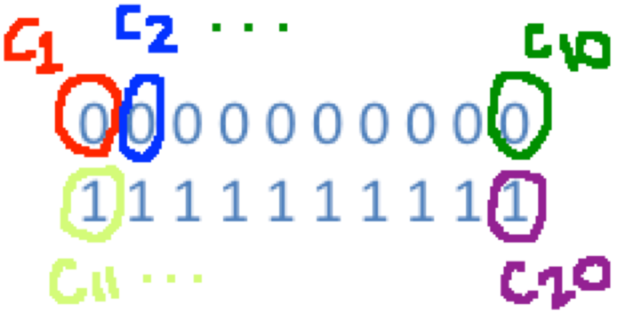
Visualizing the Change in Split (cont.)

Before Splitting:

0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1

Split by this:



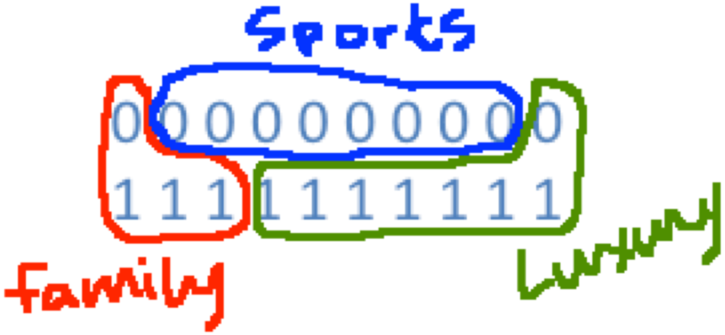


Visualizing the Change in Split (cont.)

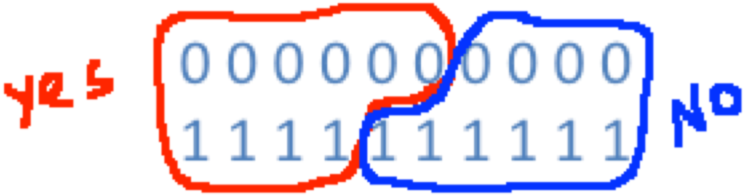
Before Splitting: 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1

Which split is better?

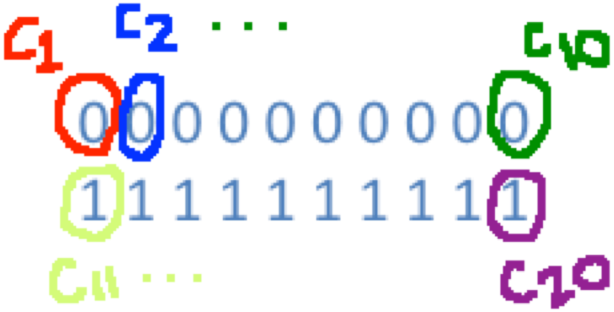
Car Type?


0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1

Own car?


0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1

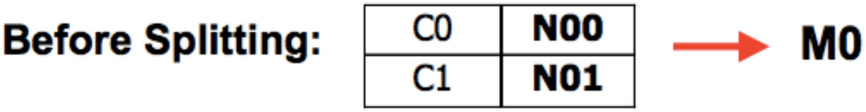
Student ID?


0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1

A Measure of Node Impurity

- Goal is to get homogeneous node
  - All the labels in the subtree are the same
  - Minimize impurity
- Examples:
  - Low degree of impurity
    C0: 9
    C1: 1
  - High degree of impurity
    C0: 5
    C1: 5
- Need a quantitative measure of impurity

Measuring Information Gain Before and After the Split



Before Splitting:

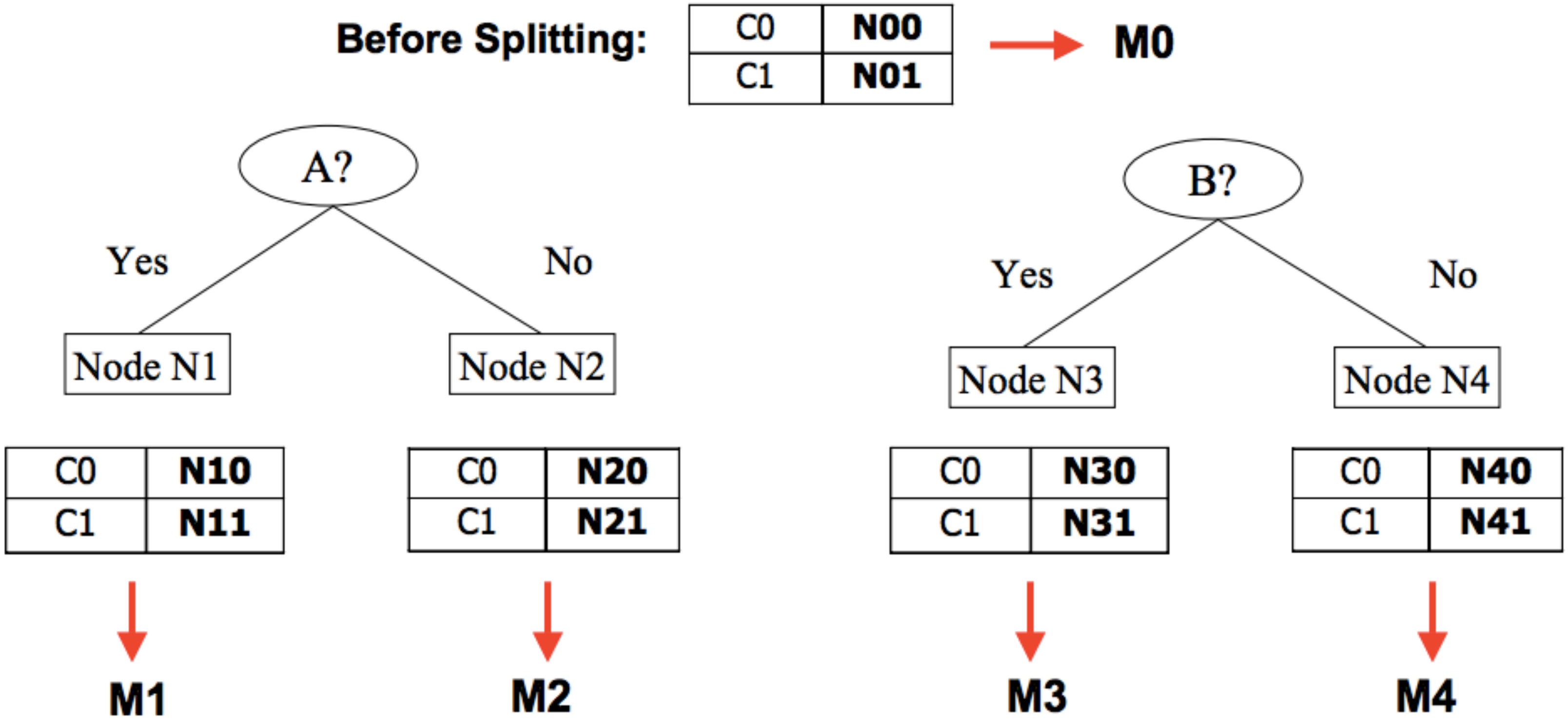
C0	N00
C1	N01

→ M0

Let M0 be the measure of impurity at the current node

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

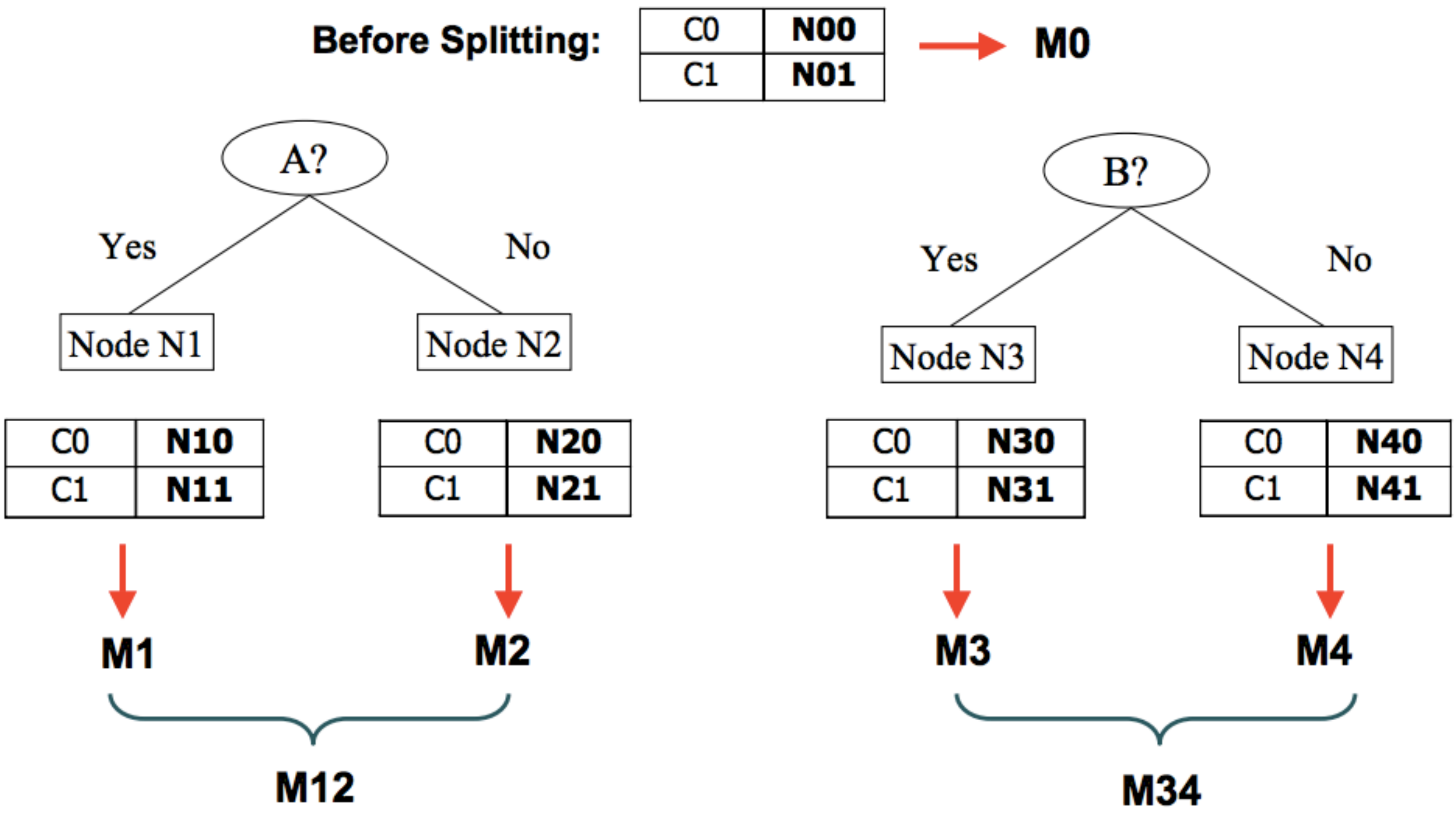
Measuring Information Gain Before and After the Split



Let M1 to M4 be the impurities at each of those nodes

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Measuring Information Gain Before and After the Split



Before Splitting:

C0	N00
C1	N01

→ M0

A? Yes → Node N1, No → Node N2

B? Yes → Node N3, No → Node N4

Node N1

C0	N10
C1	N11

Node N2

C0	N20
C1	N21

Node N3

C0	N30
C1	N31

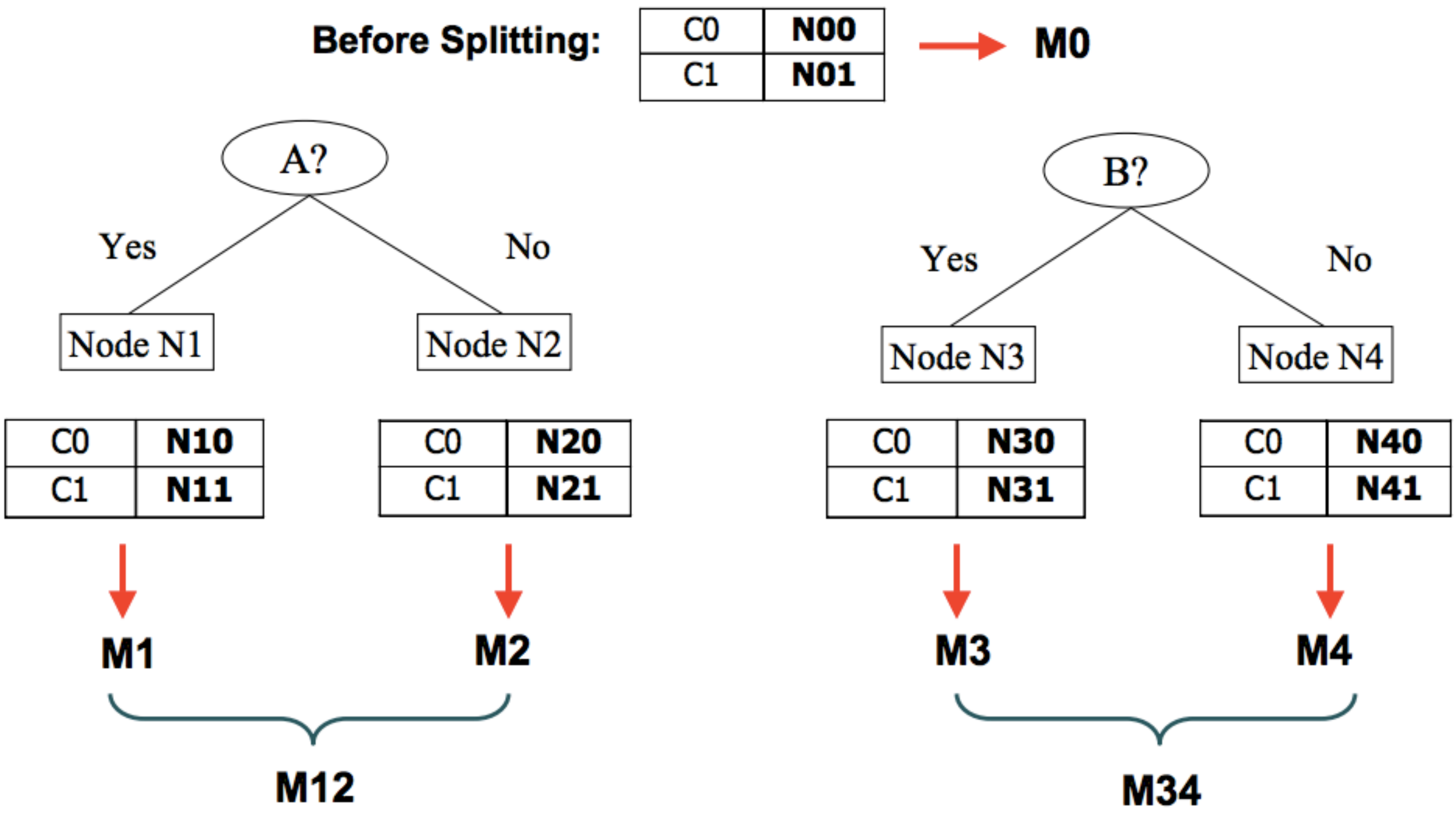
Node N4

C0	N40
C1	N41

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Summarize the impurities in M12 across all the nodes, then M34

Measuring Information Gain Before and After the Split



C0	N00
C1	N01

C0	N10
C1	N11

C0	N20
C1	N21

C0	N30
C1	N31

C0	N40
C1	N41

Gain = M0 – M12 vs M0 – M34

Image taken from https://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap4_basic_classification.pdf

Compare the gain between the two splits

Impurity Measure: Gini Diversity Index

- Default measure used in Matlab fitctree
- Given node n, class c:

$$\boxed{\mathrm{GINI}(n) = 1 - \sum_{c} \mathrm{pr}(c|n)^2}$$

  where pr(c|n) is how likely c is at node n

- Ranges in: [0, 1-1/# classes]
  - 0 when node is homogeneous
  - 1-1/# classes when classes are equally distributed

Impurity Measure: Gini Diversity Index

- Default measure used in Matlab fitctree
- Given node n, class c:

$$\boxed{\mathrm{GINI}(n) = 1 - \sum_{c} \mathrm{pr}(c|n)^2}$$

where pr(c|n) is how likely c is at node n

- pr(c|n) is defined by frequency counts e.g. 0 out of 6 possible instances

Impurity Measure: Gini Diversity Index

- Default measure used in Matlab fitctree
- Given node n, class c:

$$\boxed{\mathrm{GINI}(n) = 1 - \sum_{c} pr(c|n)^2}$$

where pr(c|n) is how likely c is at node n

- Example:
  C0: 0
  C1: 6
  Gini = $1 – (0/6)^2 – (6/6)^2 = 1 – 0 – 1 = 0$

Impurity Measure: Gini Diversity Index

- Default measure used in Matlab fitctree
- Given node n, class c:

$$\boxed{\mathrm{GINI}(n) = 1 - \sum_{c} \mathrm{pr}(c|n)^2}$$

where pr(c|n) is how likely c is at node n

- Example:
C0: 3
C1: 3
Gini = $1 – (3/6)^2 – (3/6)^2 = 1 – 0.25 – 0.25 = 0.5$

Impurity Measure: Gini Diversity Index

- Default measure used in Matlab fitctree
- Given node n, class c:

$$\boxed{\mathrm{GINI}(n) = 1 - \sum_{c} pr(c|n)^2}$$

  where pr(c|n) is how likely c is at node n

- Example:

  C0: 1

  C1: 5

  Gini = $1 – (1/6)^2 – (5/6)^2 = 1 – 0.0278 – 0.694 = 0.278$

Other Impurity Measures

- Common impurity measures
  - Gini diversity index
    - Measures relative frequency of a class at a node
  - Entropy
    - Measures homogeneity of a node
    - Similar to Gini index
  - Misclassification error
    - Measures misclassification error made by a node

Preliminary Case Study: COSC 111

- How to help students do better in the course?
- Course structure (3-hour night class):
  - Exams: MT1, MT2, Final Exam
  - Weekly Labs
  - Weekly Quizzes (unlimited re-takes allowed)
  - In-Class Activities
  - Monthly Assignments

Additional Data Gathered

- Total Sessions
- Total Pageviews
- Total Time on Page
- Discussion Forum Activity
- Number of Replies
- Hours before Submission Data
- Time Spent on Slides
- Pageviews on Slides
- Time Spent on Discussion Forum
- Number of Online Sessions

Decision Tree Induction in Matlab

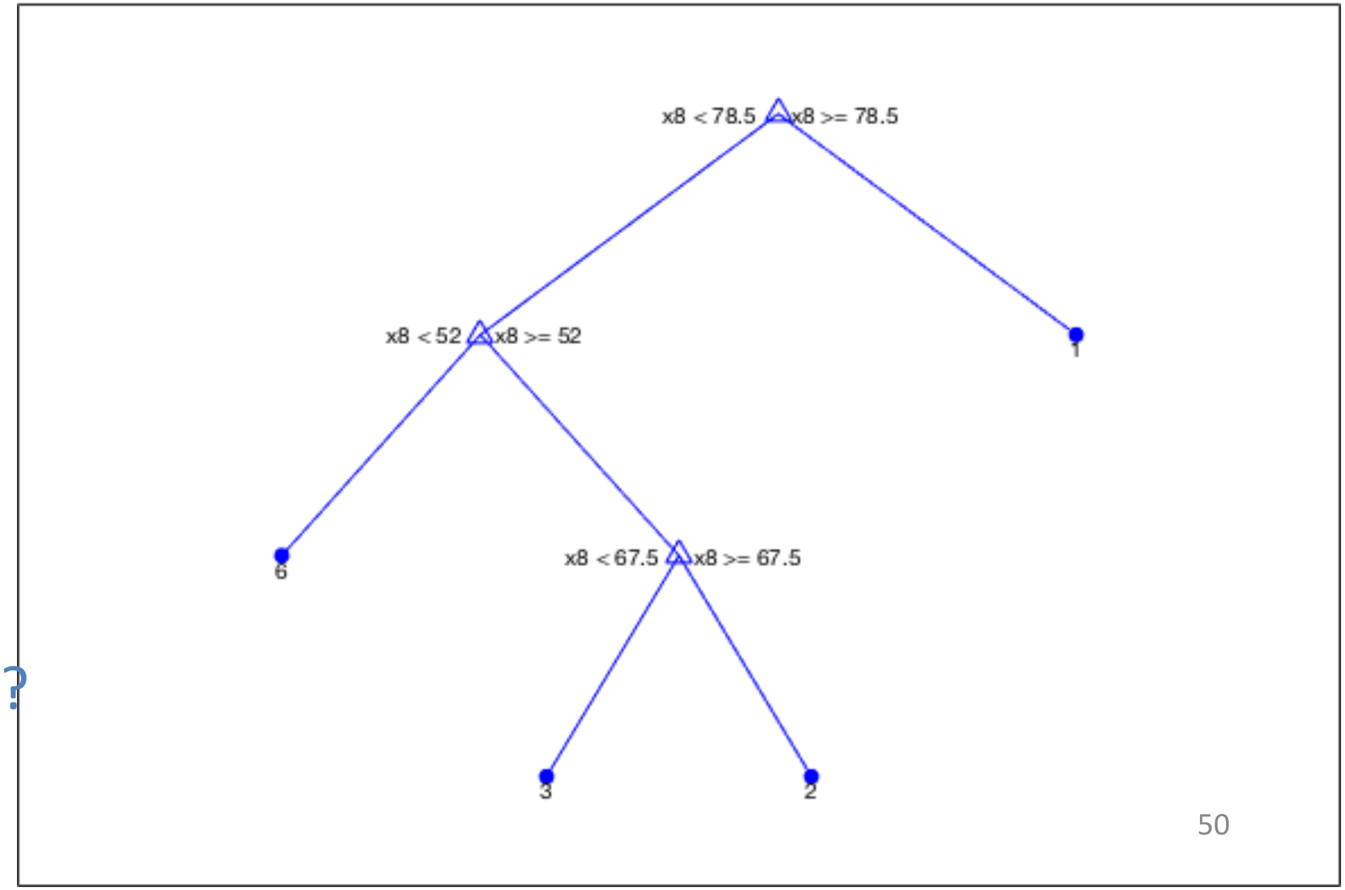
- Simple commands:

```
>> data = csvread( ‘cosc111data.csv’, 2 );
% read in only rows 2 onward

>> tree = fitctree( data(:, 1:17), data( :, 19) );
% syntax used: fitctree( X, Y )
% X is the set of attribute data
% Y is the actual class labels

>> view( tree )
% text output of tree branches

>> view( tree, ‘mode’, ‘graph’ )
% corresponding visualization of tree
```

Initial Tree Learned

```
Decision tree for classification
1  if x8<78.5 then node 2 elseif x8>=78.5 then node 3
2  if x8<52 then node 4 elseif x8>=52 then node 5
3  class = 1
4  class = 6
5  if x8<67.5 then
   node 6
   elseif x8>=67.5 then
   node 7
6  class = 3
7  class = 2
```

Debugging:
What could x8 be?



Actual Tree

Decision tree for classification

1 if x7<44.25 then node 2 elseif x7>=44.25 then node 3

2 class = 6

3 if x7<65.85 then node 4
elseif x7>=65.85 then
node 5

4 if x14<32213.5 then
node 6
elseif x14>=32213.5
then node 7

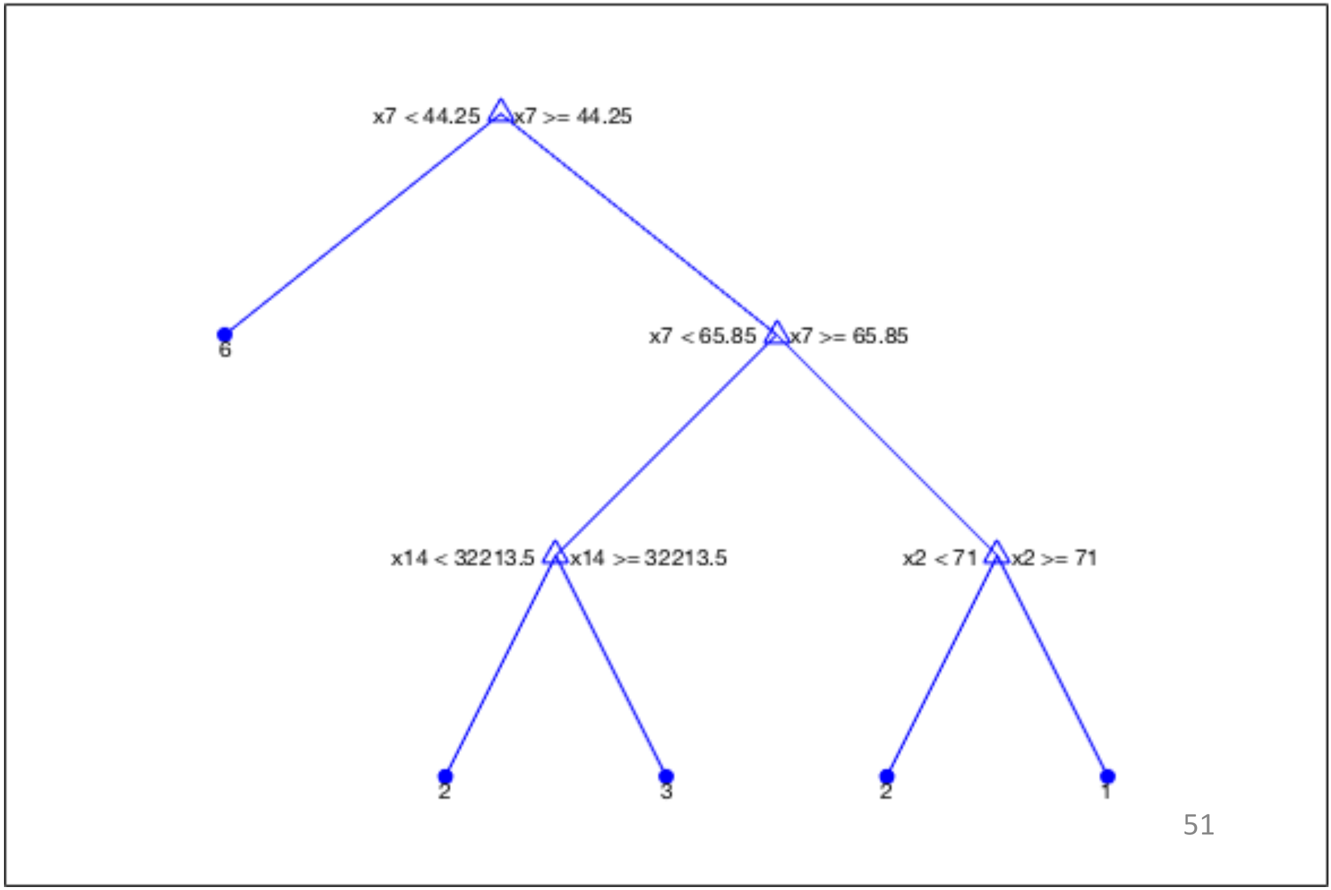
5 if x2<71 then node 8
elseif x2>=71 then
node 9

6 class = 2

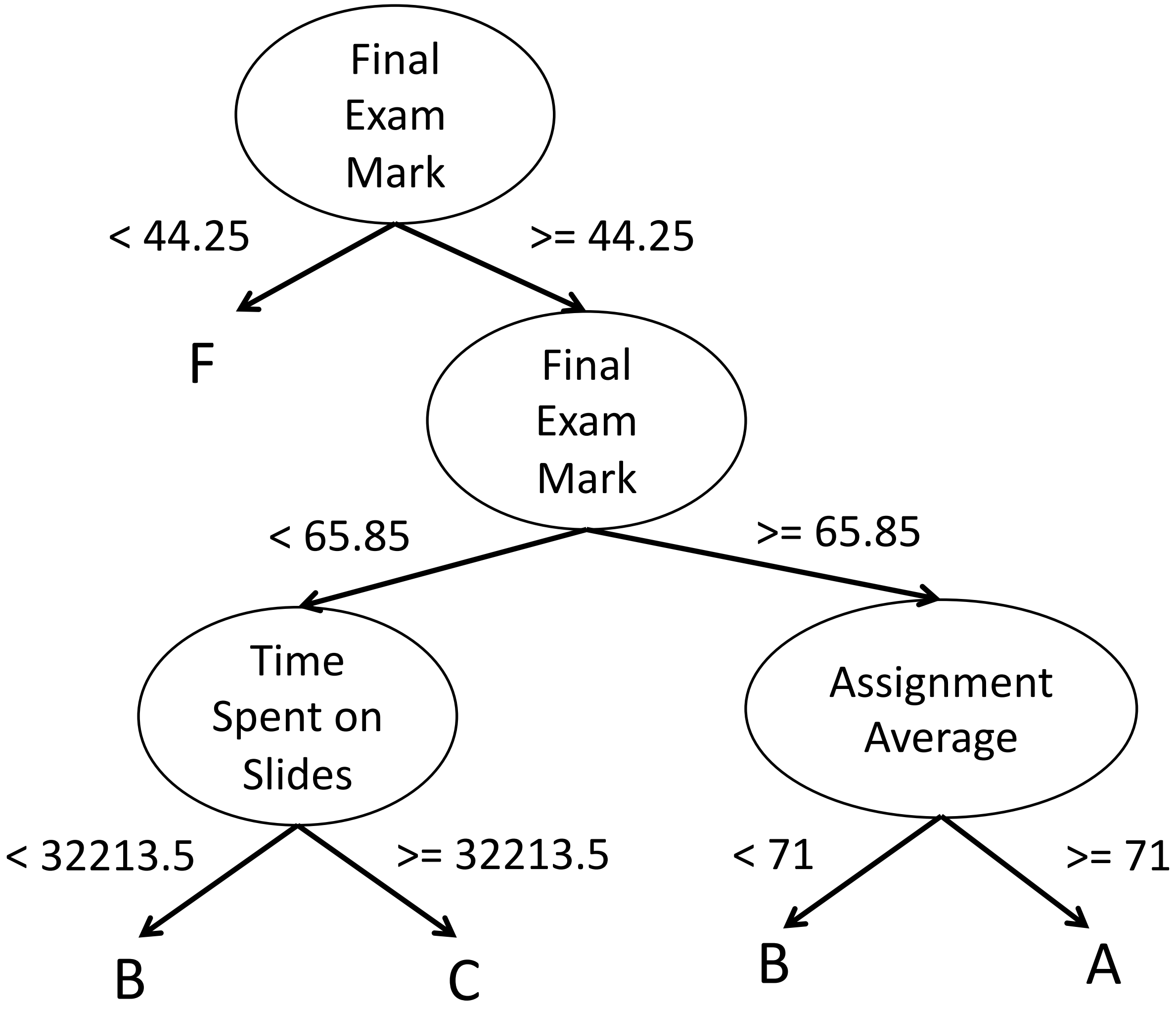
7 class = 3

8 class = 2

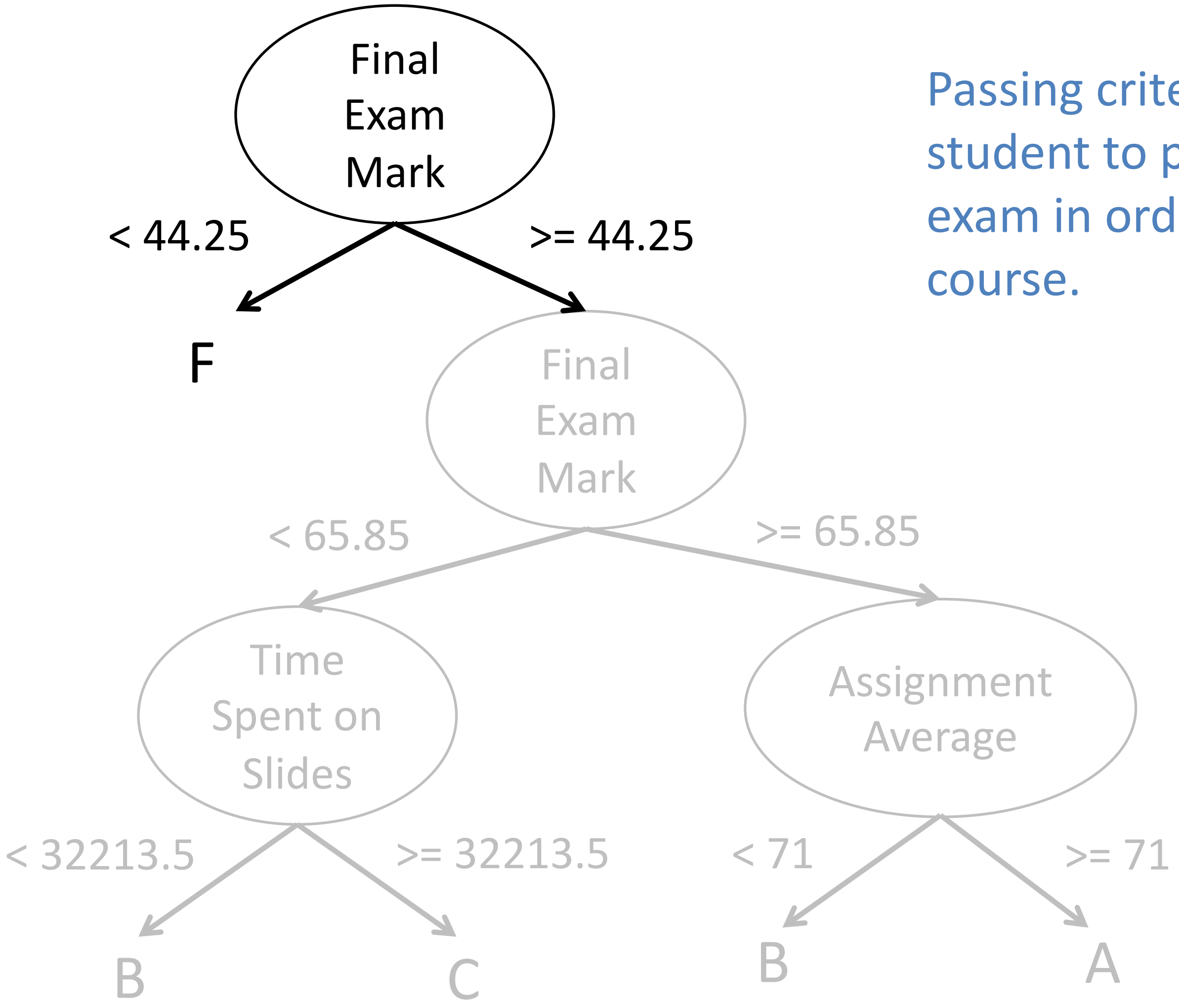
9 class = 1



Actual Tree – Better Labels

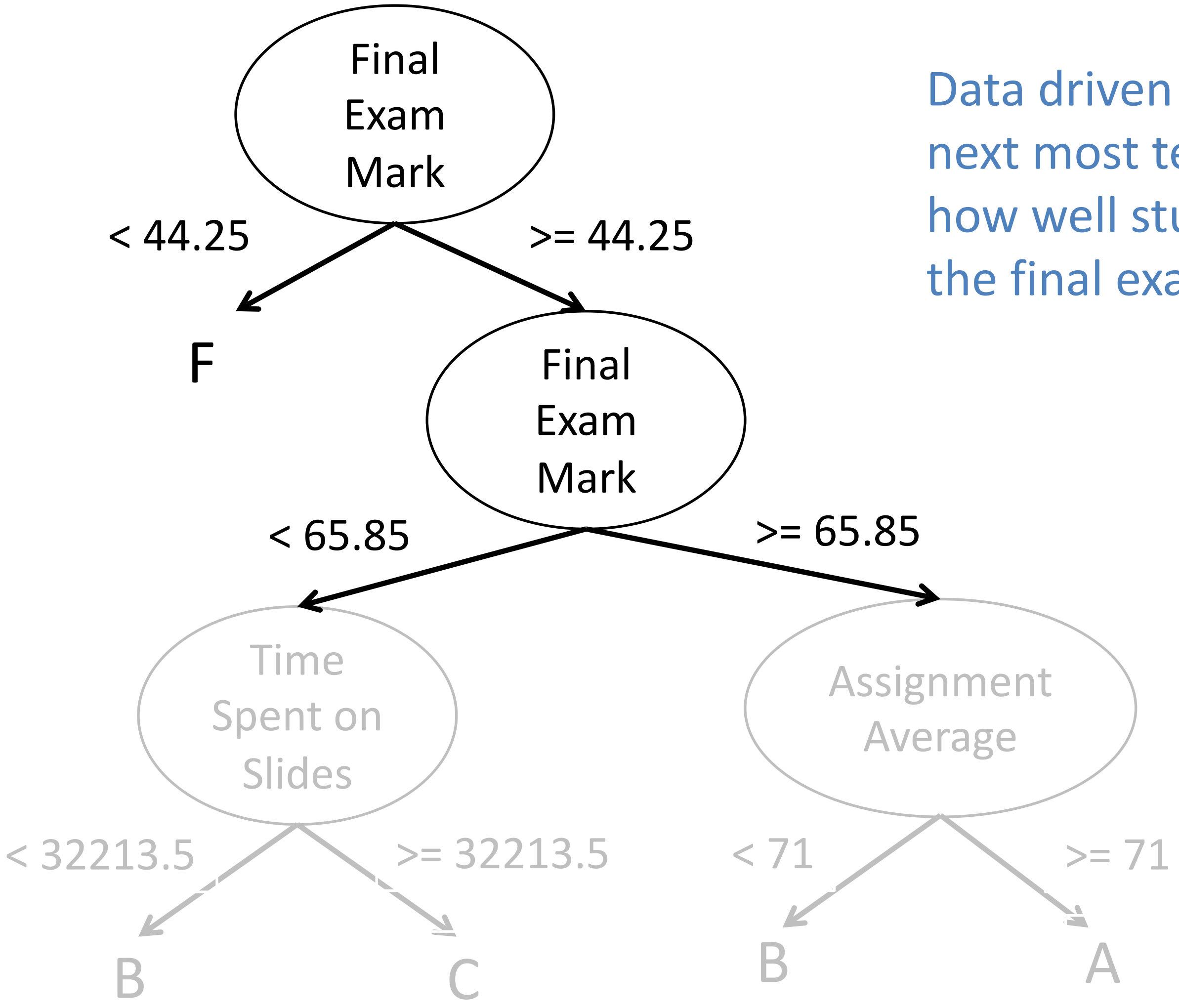


Explaining the Result



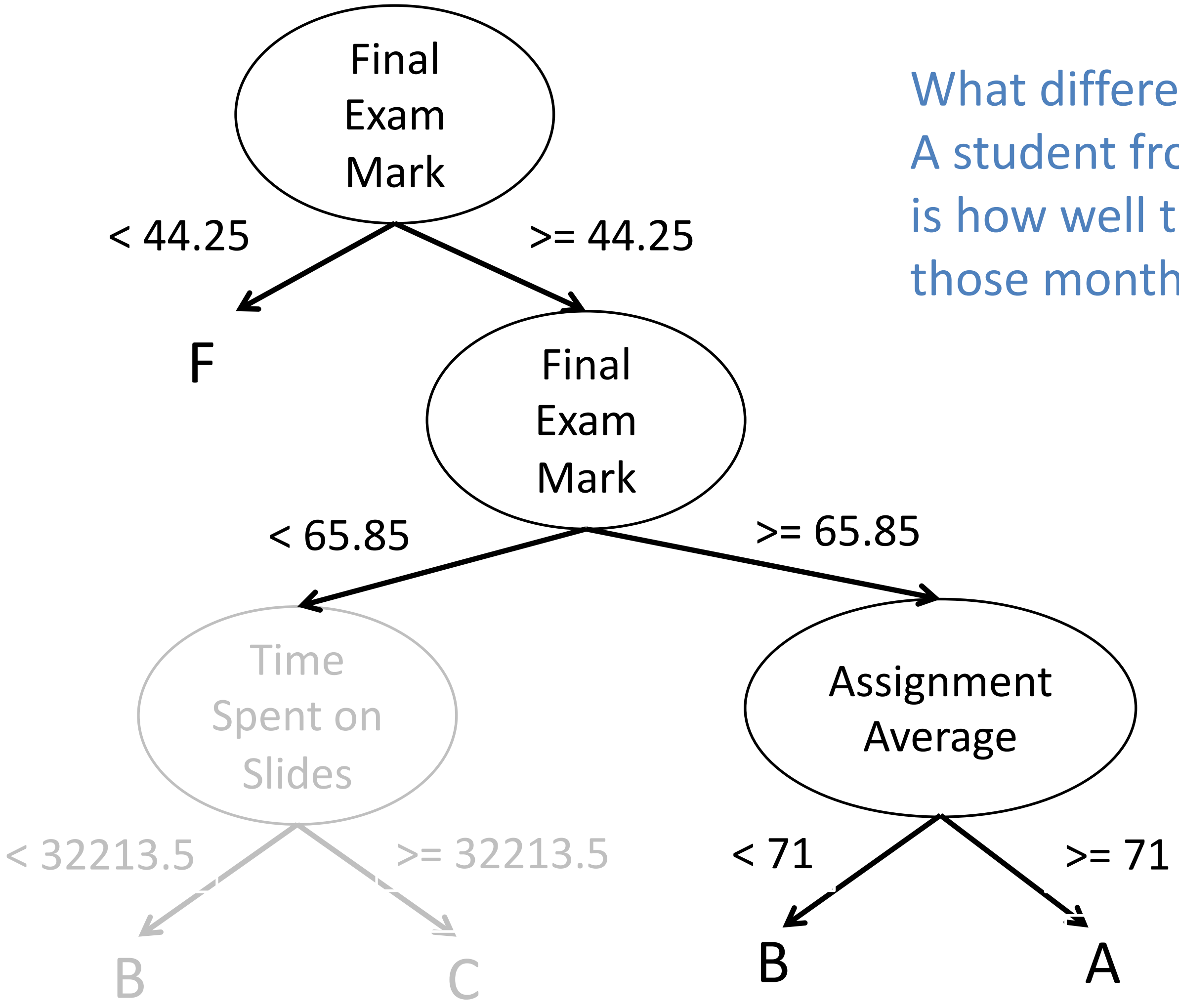
Passing criteria requires student to pass the final exam in order to pass the course.

Explaining the Result



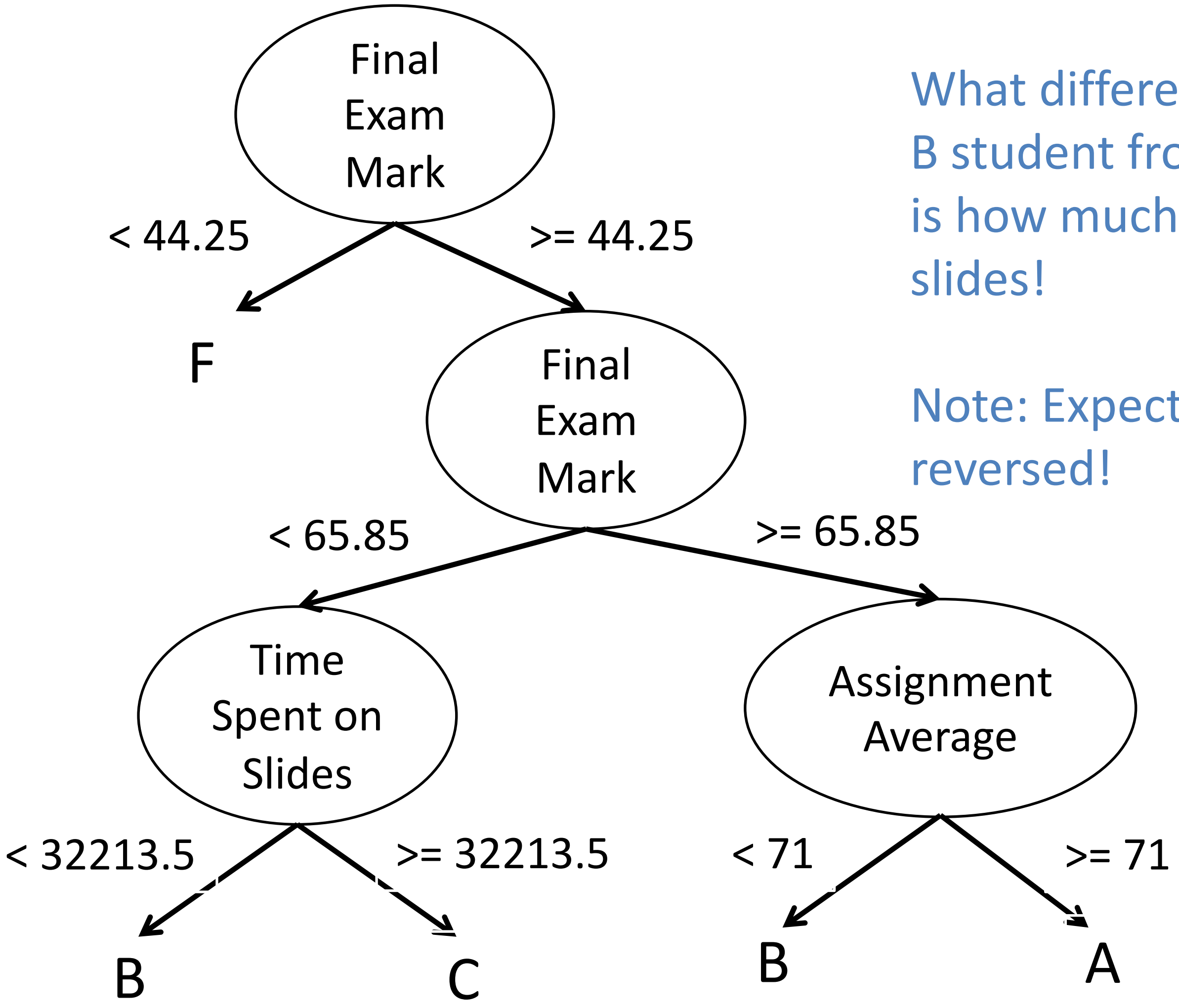
Data driven – finds that next most telling split is how well student does on the final exam!

Explaining the Result



What differentiates an A student from a B student is how well they do on those monthly assignments!

Explaining the Result



What differentiates a B student from a C student is how much they read slides!

Note: Expectations are reversed!

Classification in the Abstract

The line is a classifier!

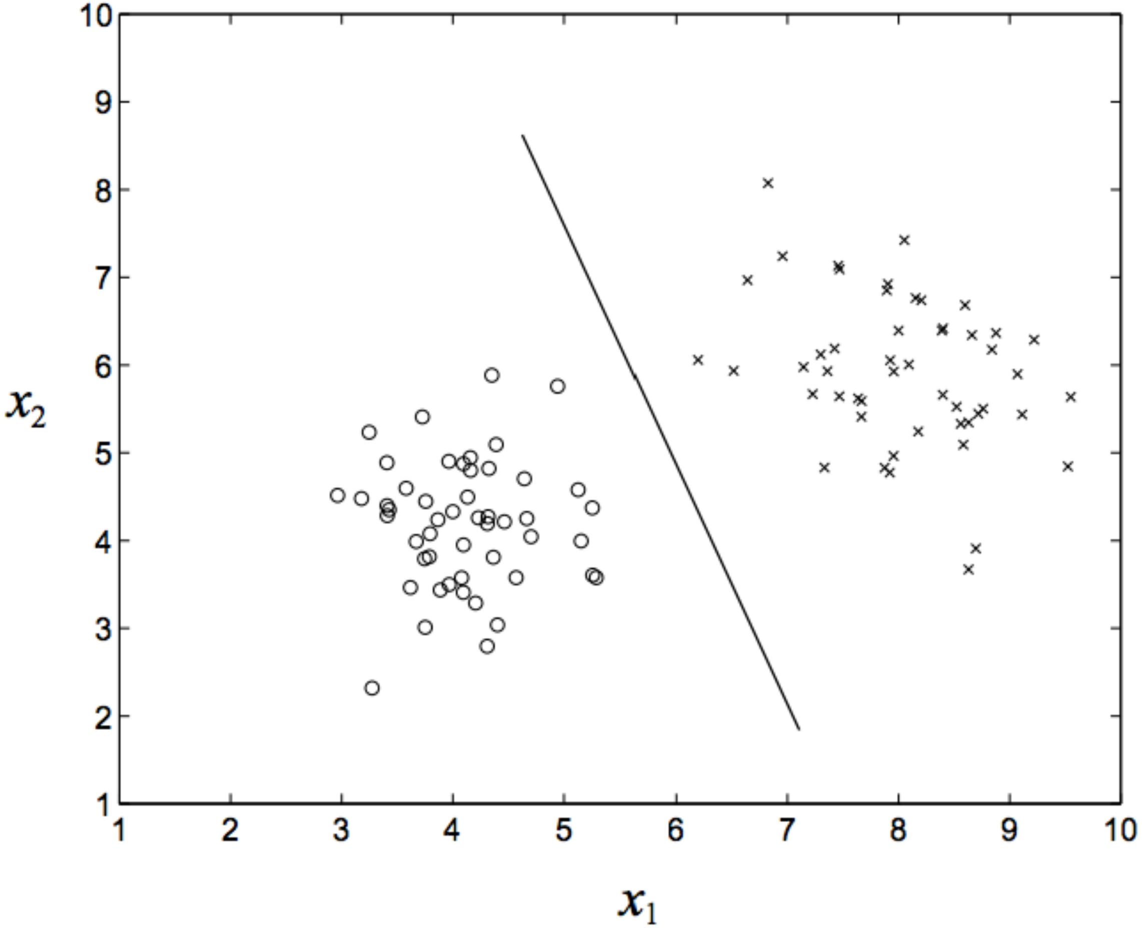
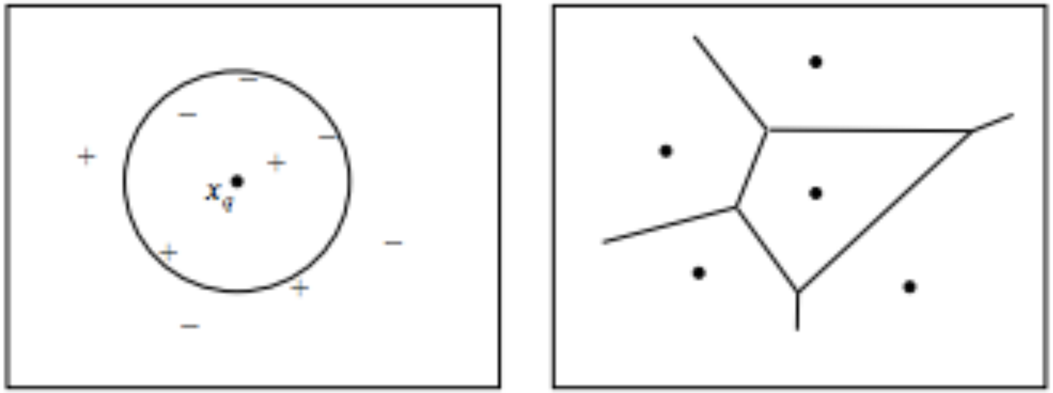


Image taken from http://www.cs.nyu.edu/~roweis/csc2515-2003/notes/lec2x.pdf

Advanced Classifiers



Split up the surface and take a different action for each decision surface

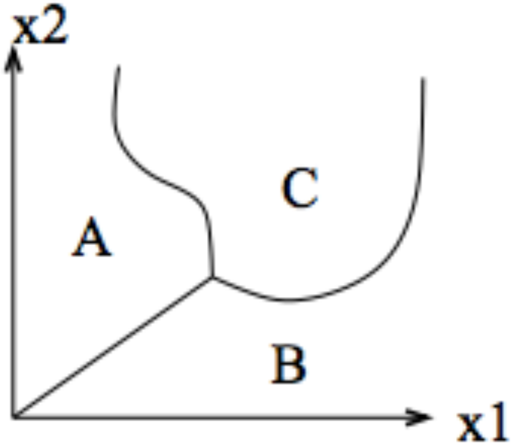


Image taken from http://www.cs.nyu.edu/~roweis/csc2515-2003/notes/lec2x.pdf

Tree Classifier

- Axes: x1,x2
- Decision points: t1, t2, t3, t4, t5
- Decision surfaces: A, B, C, D, E, F

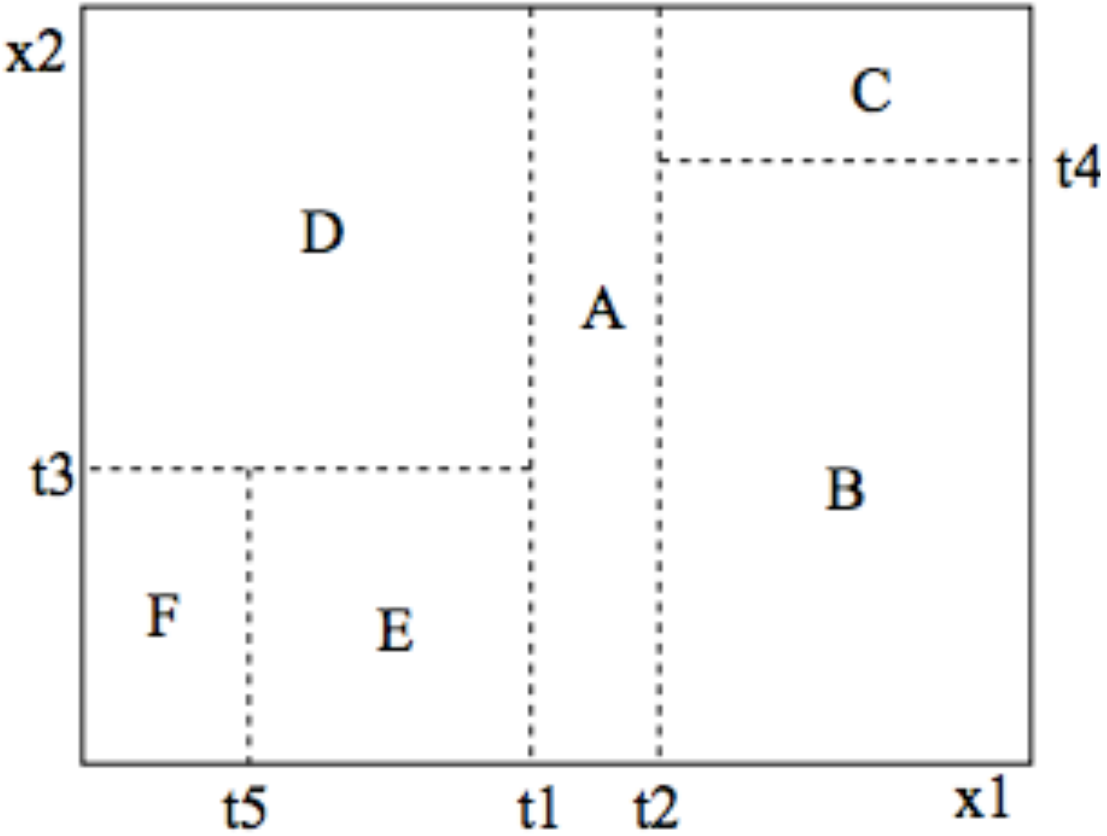
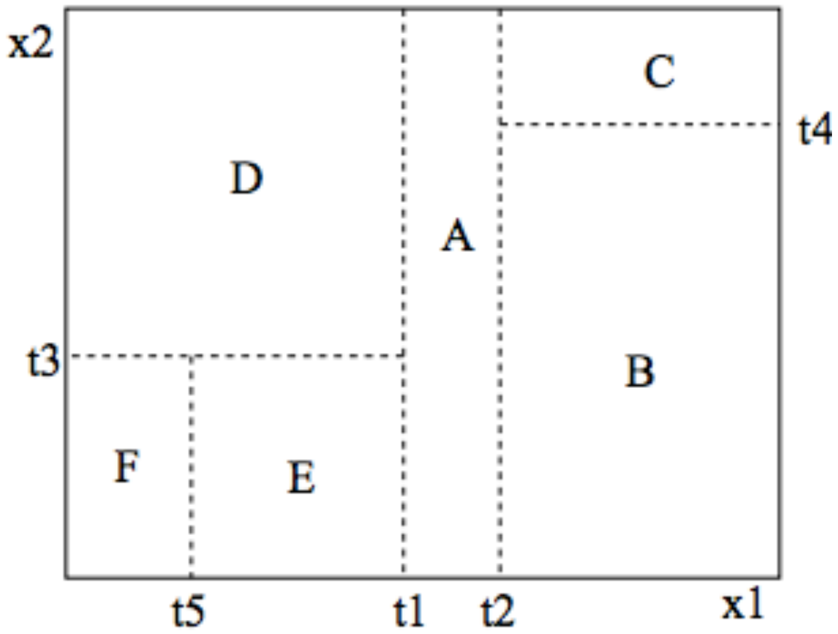


Image taken from http://www.cs.nyu.edu/~roweis/csc2515-2003/notes/lec2x.pdf

Tree Classifier



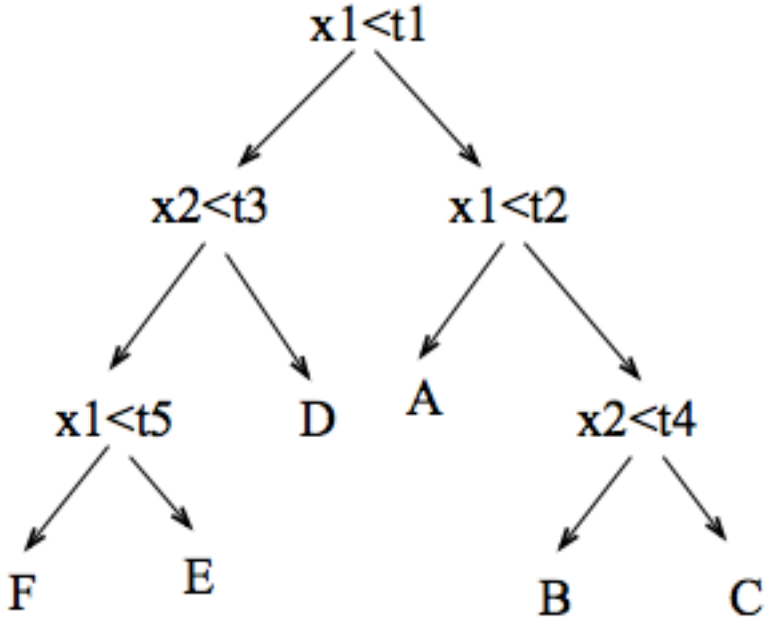


Image taken from http://www.cs.nyu.edu/~roweis/csc2515-2003/notes/lec2x.pdf

Visualization

Corresponding Tree Graph

Key Ideas

- Decision tree
  - Models decisions and possible consequences
  - Learn model from labeled dataset
  - Test/make predictions with unseen data
- Representation:
  - Nodes represent attribute
  - Branches indicate values to split on
  - Leaf nodes represent class label
- Algorithm:
  - Repeatedly pick an attribute to split on
  - Stop when node is homogeneous or no attributes left
  - Measure quality of split based on impurity and information gain