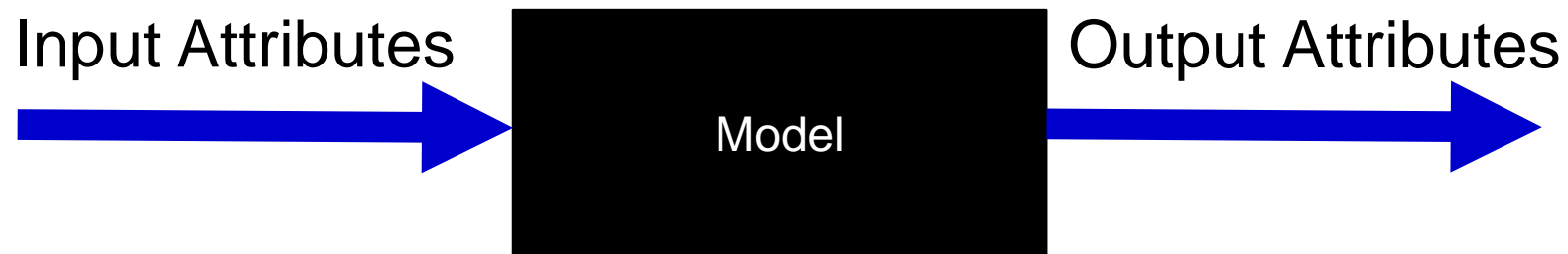# CS105 Lab 9: Data Mining

Brian Borucki – bborucki@bu.edu

Models

- Think of a model as a black box

- We feed the black box some input and get some output

Input Attributes → **Model** → Output Attributes

Models

- Say I'm modeling a person

- If I get certain behavior from the person on some input, I expect the same behavior from my model on the same input

Input Attributes $\longrightarrow$ | Model | $\longrightarrow$ Output Attributes
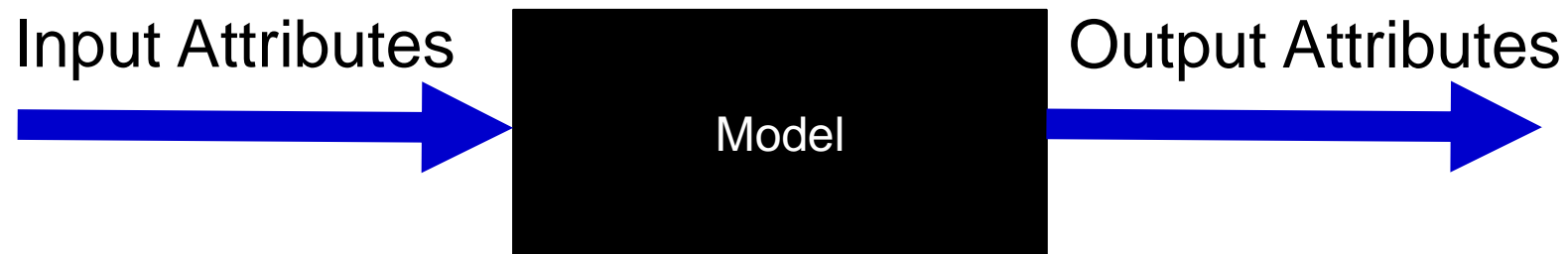
Types of Learning

- Classification – putting objects into bins, predicts behavior

    - "At risk" vs "Not at risk"

- Association – finds relationships that already exist

- Numeric Estimation – predicts numeric outcomes

    - How much should an insurance company charge a given person?

    - How will the stock market react to X?

Models

- We obtain models by training them on data
- We also test how good our models are using confusion matrices

Predicted

|	Coupe	Sedan	Van	Semi
Coupe	100	0	0	0
Sedan	0	100	0	0
Van	0	0	100	0
Semi	0	0	0	100

Actual

Models

- We obtain models by training them on data
- We also test how good our models are using confusion matrices

Predicted

	Coupe	Sedan	Van	Semi
Coupe	0	33	33	34
Sedan	33	0	34	33
Van	33	34	0	33
Semi	34	33	33	0

Actual

Models

- We obtain models by training them on data
- We also test how good our models are using confusion matrices

Predicted

	Coupe	Sedan	Van	Semi
Coupe	50	50	0	0
Sedan	50	50	0	0
Van	0	0	50	50
Semi	0	0	50	50
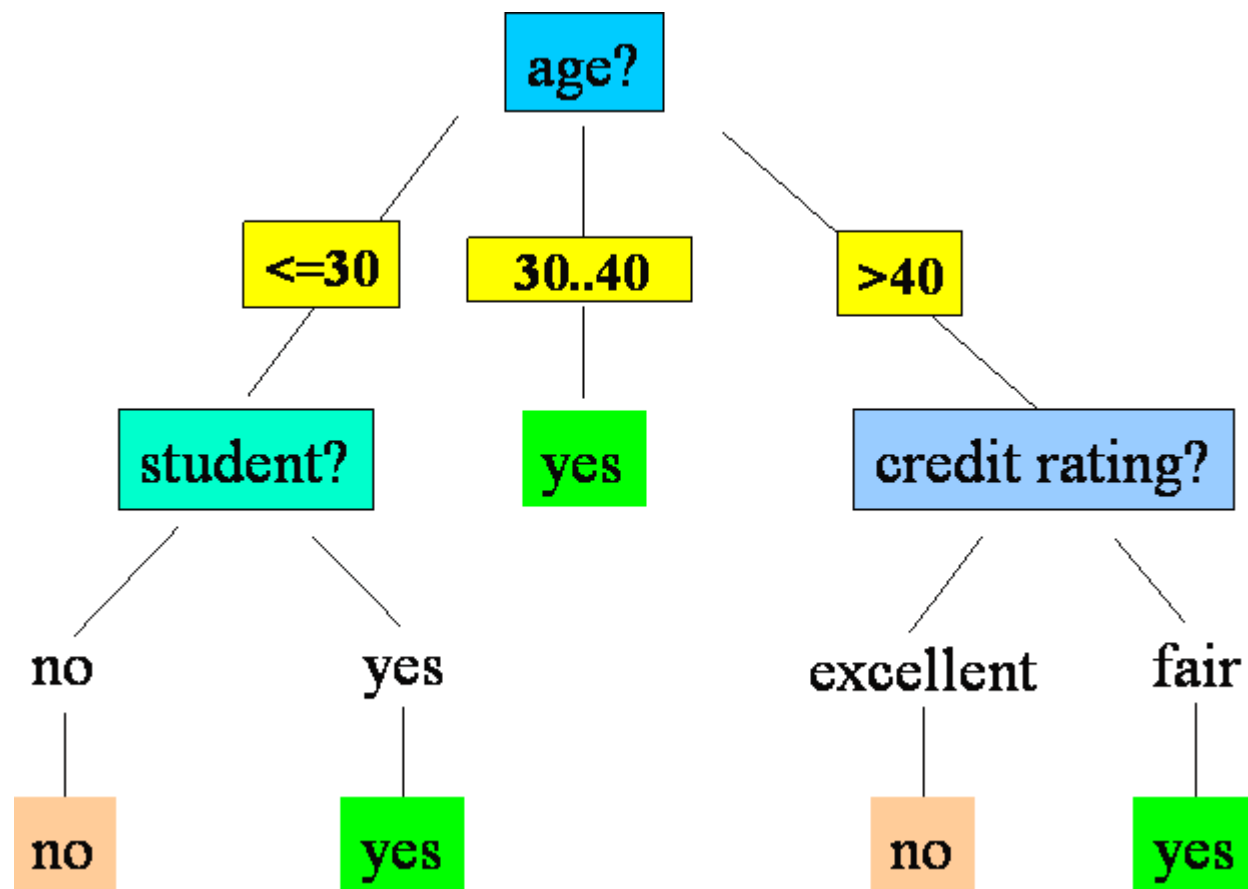
Actual

[age, income, student, credit, buys?]

[26 , high , no , fair , yes]

Creating a Confusion Matrix

Predicted

	Yes	No
Yes	0	0
No	0	0

Actual

Start by listing all possible classes along the top and side

Initialize the matrix with 0's in all places

For each tuple in your testing set:
    run in through the model
    increment the appropriate cell accordingly

Creating a Confusion Matrix

Predicted

	Yes	No
Actual Yes	0	0
No	0	0

The model predicted a "No"

Creating a Confusion Matrix

Predicted

	Yes	No
Yes	0	0
No	0	0

Actual

The model predicted a "No"

The actual was "Yes"

Creating a Confusion Matrix

Predicted

	Yes	No
Actual Yes	0	1
No	0	0

The model predicted a "No"

The actual was "Yes"

Increment the corresponding value

This Lab

- Any Questions?

- Spend time making sure you have a group
    - Professor Sullivan will talk more about the projects tomorrow in lecture

- Submit the completed ans.txt and my_scrub.py files

Project

- Raise your hand if you don't have a group yet (exactly 3 people to a group)

- Exchange contact information with your group members

- Chat briefly about interests, any ideas you may have for the project