



## Q: What is a classifier?

- A. *This option is intentionally left blank*
- B. A method to predict the future
- C. An algorithm that maps input data to a specific category
- D. A type of decision tree used for data mining
- E. A type of data storage for algorithms



# CPSC 100

## Computational Thinking

### Algorithm, Classifiers and Trees!

Instructor: Parsa Rajabi  
Department of Computer Science  
University of British Columbia



# Agenda

- Course Admin
- Recap
  - Classifiers
  - Decision Trees
- Entropy

# Learning Goals

# Learning Goals

After this lecture, you should be able to:

- Describe the **classification** steps.
- Explain the concept of a **rooted tree** and **decision tree**.
- Describe what the general decisions are in building a decision tree.
  - **Build a decision tree using entropy.**
- Describe **what considerations** are important in building a decision tree.

# Course Admin



# Course Admin

- **Lab #2**
  - Due on Friday, Jan 24 at 11:59pm
- **Post-Class (PC) Quiz #1**
  - Only 1 attempt, 60 minutes
  - Open book, open-AI\* (\*you must disclosure your usage)
  - Due on Sunday, Jan 26 at 11:59pm
- **Group Contracts**
  - Extended to Monday, Jan 27 at 11:59pm
- **PC Quiz #2**
  - To be released next week!

# Classifiers



# Classifier

- A **classifier** is an **algorithm** that maps the input data to a specific category
  - Classifiers are derived from patterns or correlations from data.

# Classifier: Training vs Test Data

- The data that classifiers learn the patterns has the “answer”
  - This data is called **training data**.
- Some of the training data is held back to check and see if the classifier works.
  - This is called **test data**.

Classifiers then apply these patterns to new data with no “answer”

- **Input:** Digital image
- **Output:** Cat/not a cat
- **Training data:**  
Labeled images of cats and  
images that are not cats



# Classification Task - Loan Applications

**Input:** Individual's loan application

- Address, age, gender, credit rating, etc.

**Output:** Acceptance/Rejection

**Training data:**

List of loan apps, decisions made, and for those who were approved, whether they repaid the loan or not

# Regular Trees

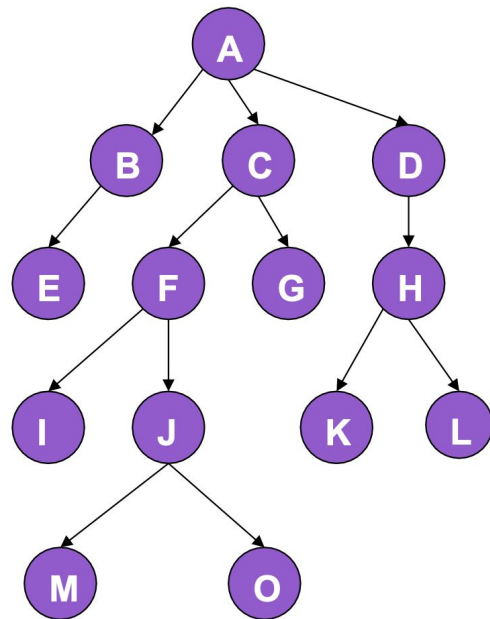


# Trees in Computer Science

- A Decision Tree is a way for a computer to make decisions based on a series of questions.

A **tree** is a **collection of nodes** such that

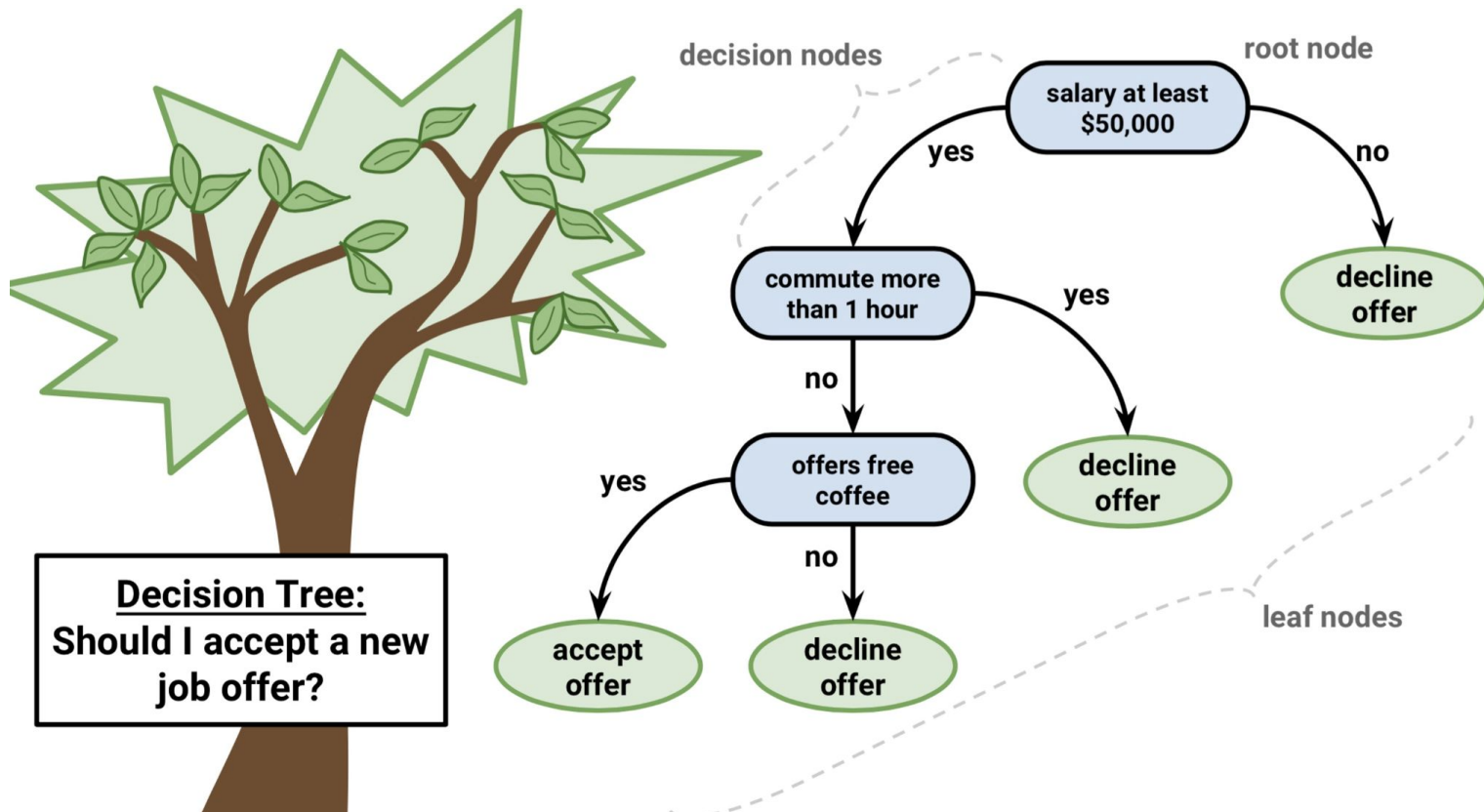
- One node is the designated **root**.
- A node can have zero or more children;
- a node with zero children is a leaf.
- All non-root nodes have a single parent.
- Edges denote parent-child relationships.
- Nodes and/or edges may be labeled by data.
  - Each node on this tree is labeled by a letter





# Decision trees

Trees whose node labels are **attributes**, edge labels are **conditions**



# Building Decision Trees

- Should you get an ice cream?
- You might start out with the following data

*Attributes*

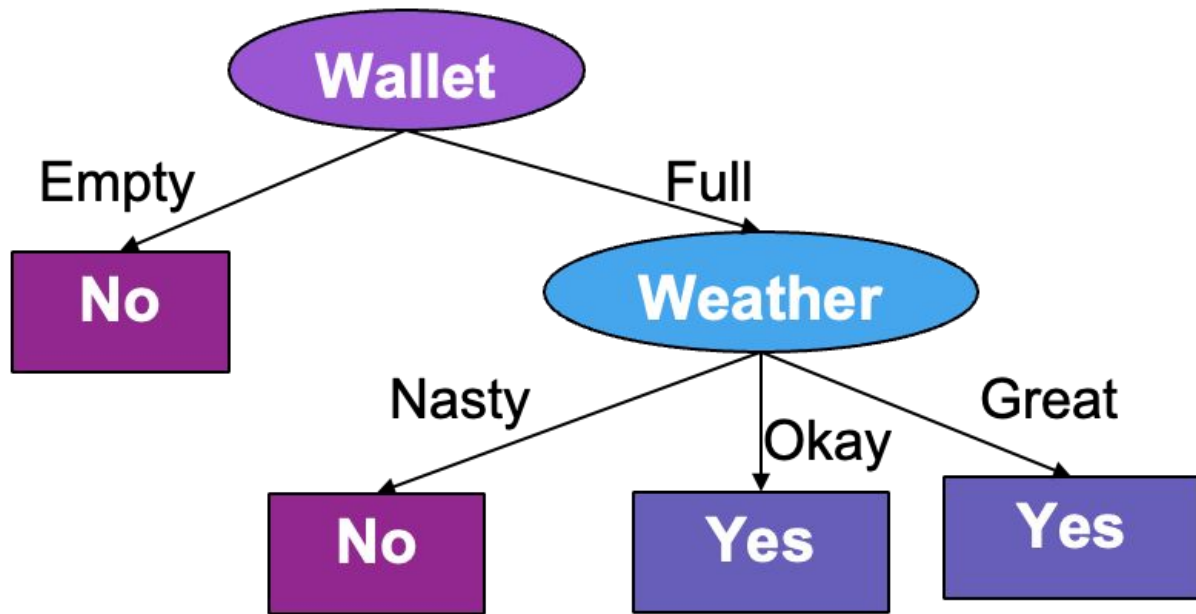
*Conditions*

Weather	Wallet	Ice Cream?
Great	Empty	No
Nasty	Empty	No
Great	Full	Yes
Okay	Full	Yes
Nasty	Full	No



# Should you get an ice cream?

Weather	Wallet	Ice Cream?
Great	Empty	No
Nasty	Empty	No
Great	Full	Yes
Okay	Full	Yes
Nasty	Full	No







**Q: In classification, how is the accuracy of a classifier evaluated?**



- A. By comparing training data with random data
- B. By matching the classifier's results with decisions from test data
- C. By ensuring the classifier can handle large datasets
- D. By improving the efficiency of the algorithm

# Soccer League:

## Do we cancel the game?

# Soccer League Data

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No



# Soccer League: Cancel Game?

- Build a **decision tree** to help officials decide
- Assume that decisions are the same given the same information
- The leaf nodes should be whether or not to play
- The non-leaf nodes should be **attributes** (e.g., Outlook, Windy)
- The edges should be **conditions** (e.g., sunny, hot, normal)

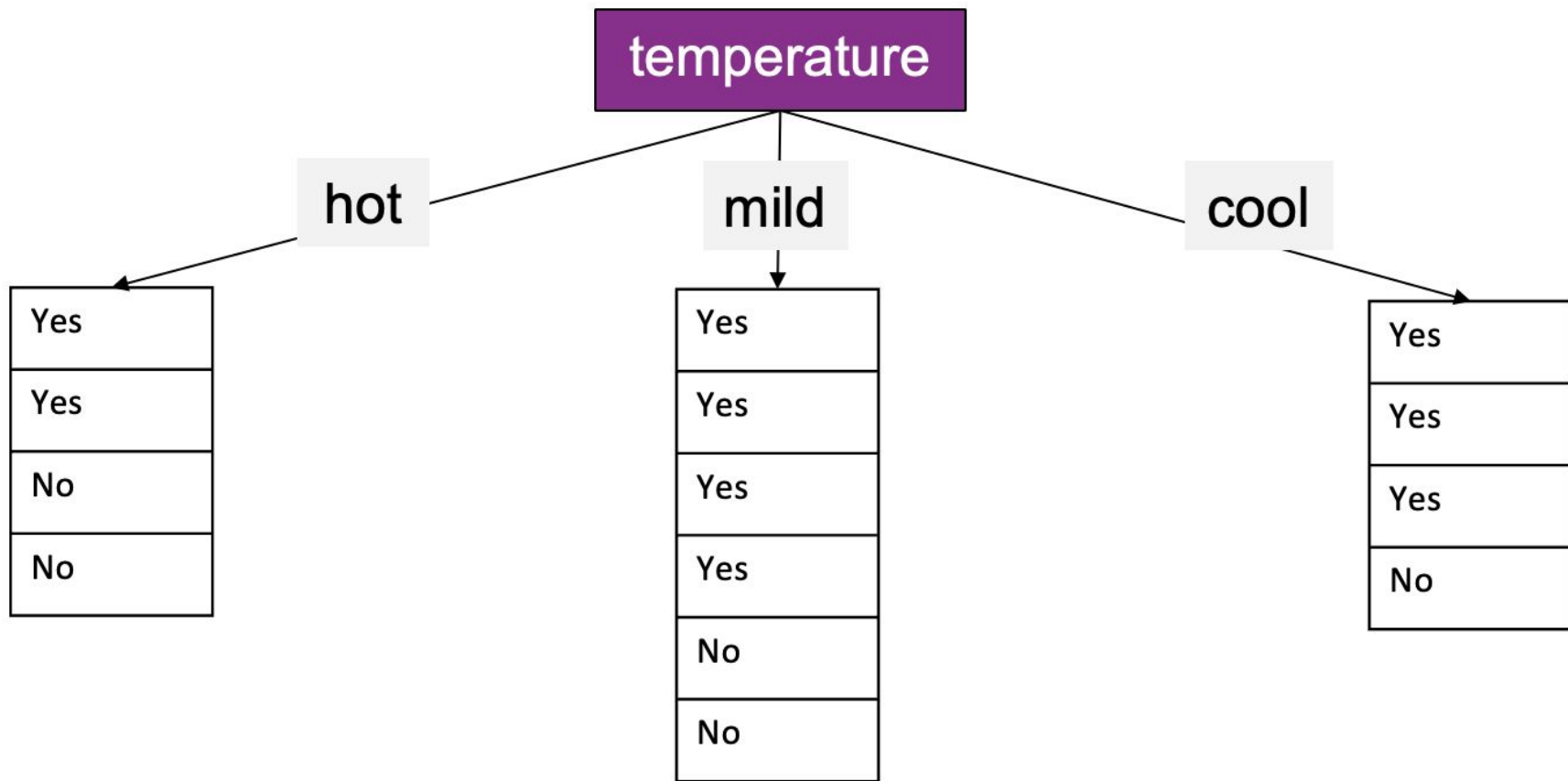
Want to have as few mixed “Yes” and “No” answers together in groups as possible.

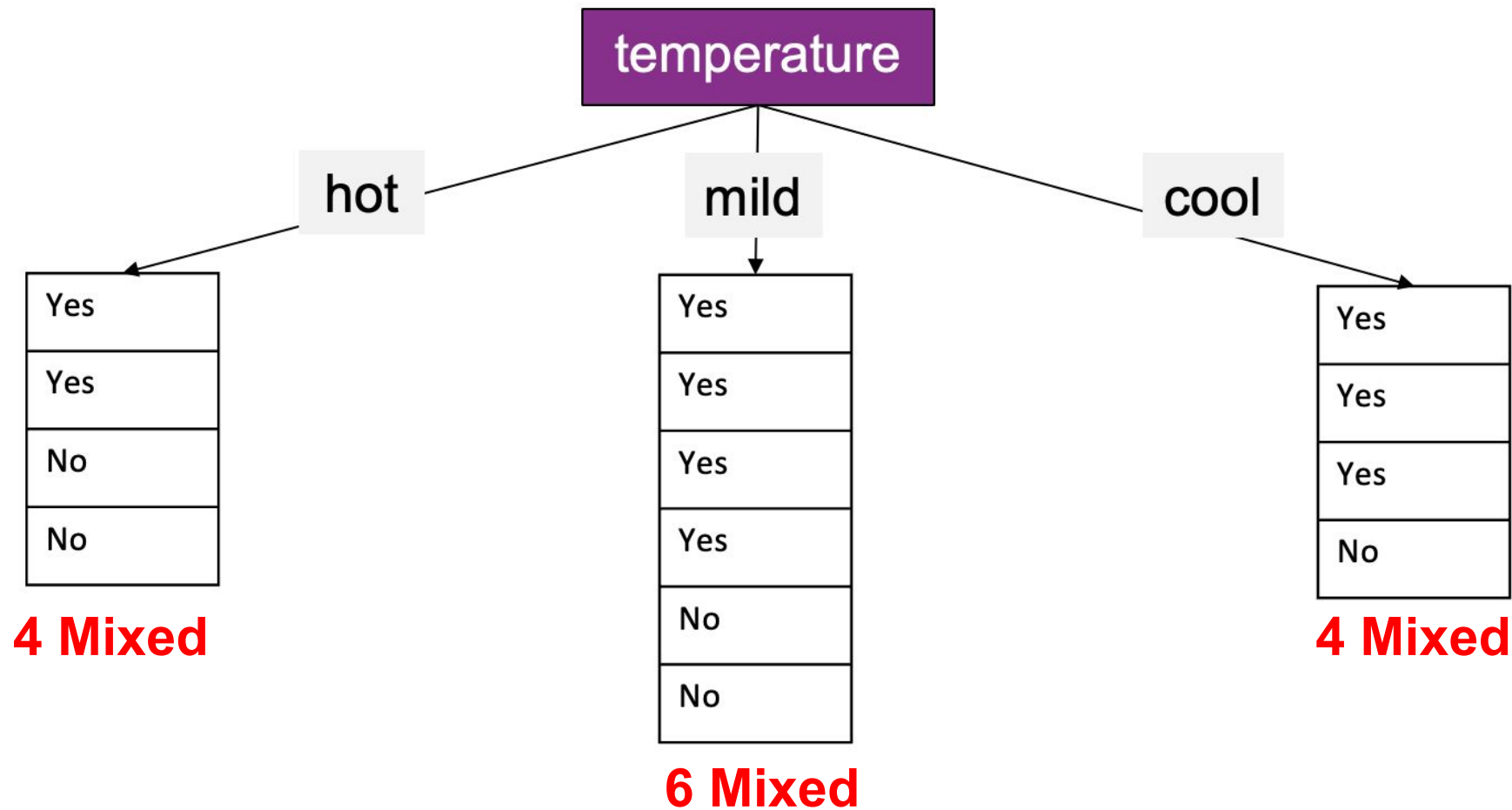
At the start we have 14 mixed Yes’s/No’s

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

**What happens if  
we split data on  
Temperature?**

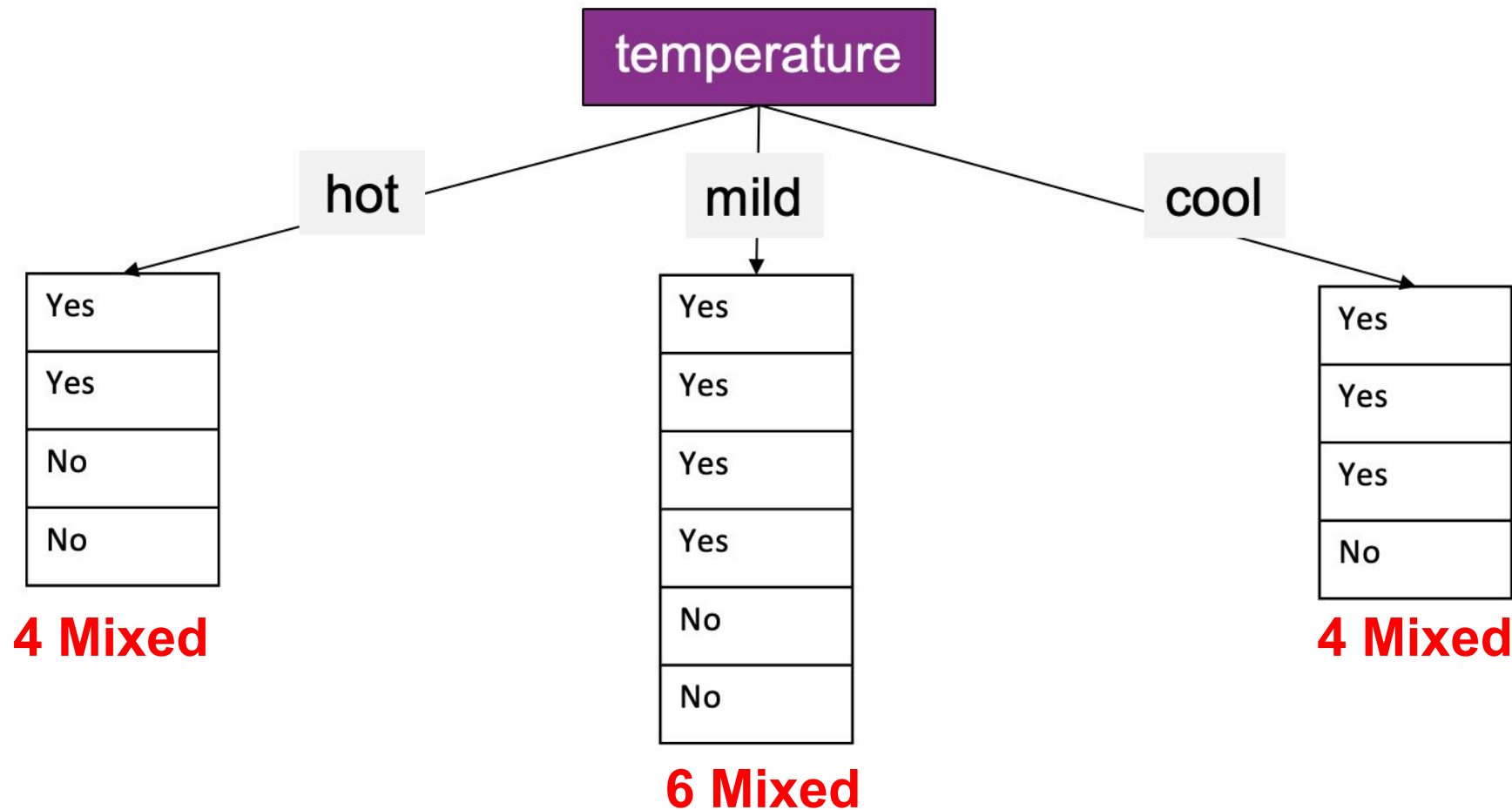




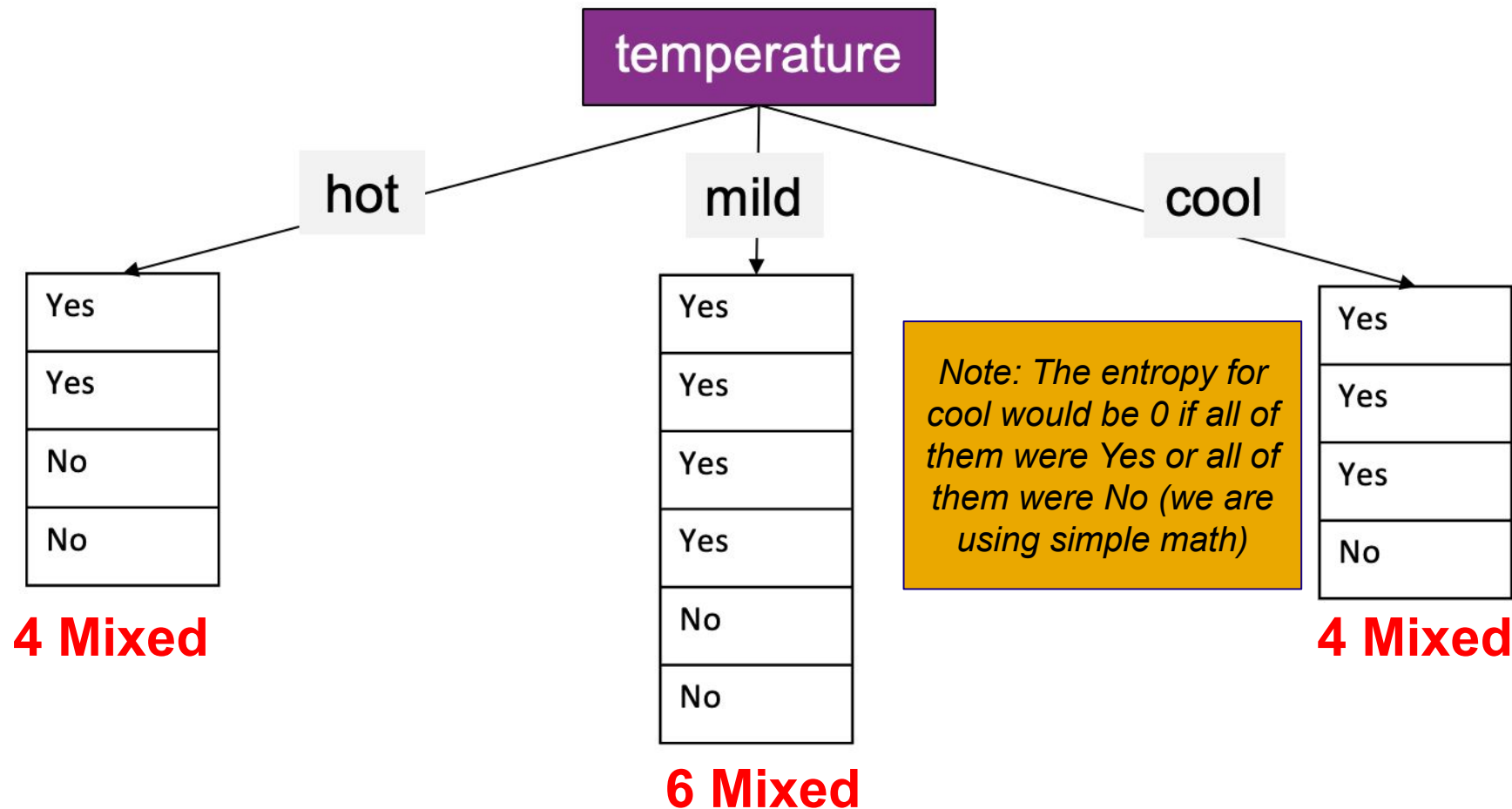


What is the  
uncertainty  
(entropy) in our  
data?

**Overall entropy = 4 + 4 + 6 = 14**



**Overall entropy = 4 + 4 + 6 = 14**



# In-class Activity

*[Groups of 2-3]*

What's the  
entropy if you  
split on  
Outlook?

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

**Q: What's the entropy if you split on Outlook?**



iClicker

A. 0

B. 5

C. 10

D. 14

E. None of the above

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No





**What's the  
entropy if you  
split on  
Windy?**

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

# What's the entropy if you split on Humidity?

Outlook	Temperature	Humidity	Windy	Play?
sunny	hot	high	false	No
sunny	hot	high	true	No
overcast	hot	high	false	Yes
rain	mild	high	false	Yes
rain	cool	normal	false	Yes
rain	cool	normal	true	No
overcast	cool	normal	true	Yes
sunny	mild	high	false	No
sunny	cool	normal	false	Yes
rain	mild	normal	false	Yes
sunny	mild	normal	true	Yes
overcast	mild	high	true	Yes
overcast	hot	normal	false	Yes
rain	mild	high	true	No

# Recap

# Wrap up



# Wrap Up

- **Lab #2**
  - Due on Friday, Jan 24 at 11:59pm
- **Post-Class (PC) Quiz #1**
  - Only 1 attempt, 60 minutes
  - Open book, open-AI\* (\*you must disclosure your usage)
  - Due on Sunday, Jan 26 at 11:59pm
- **Group Contracts**
  - Extended to Monday, Jan 27 at 11:59pm
- **PC Quiz #2**
  - To be released next week!