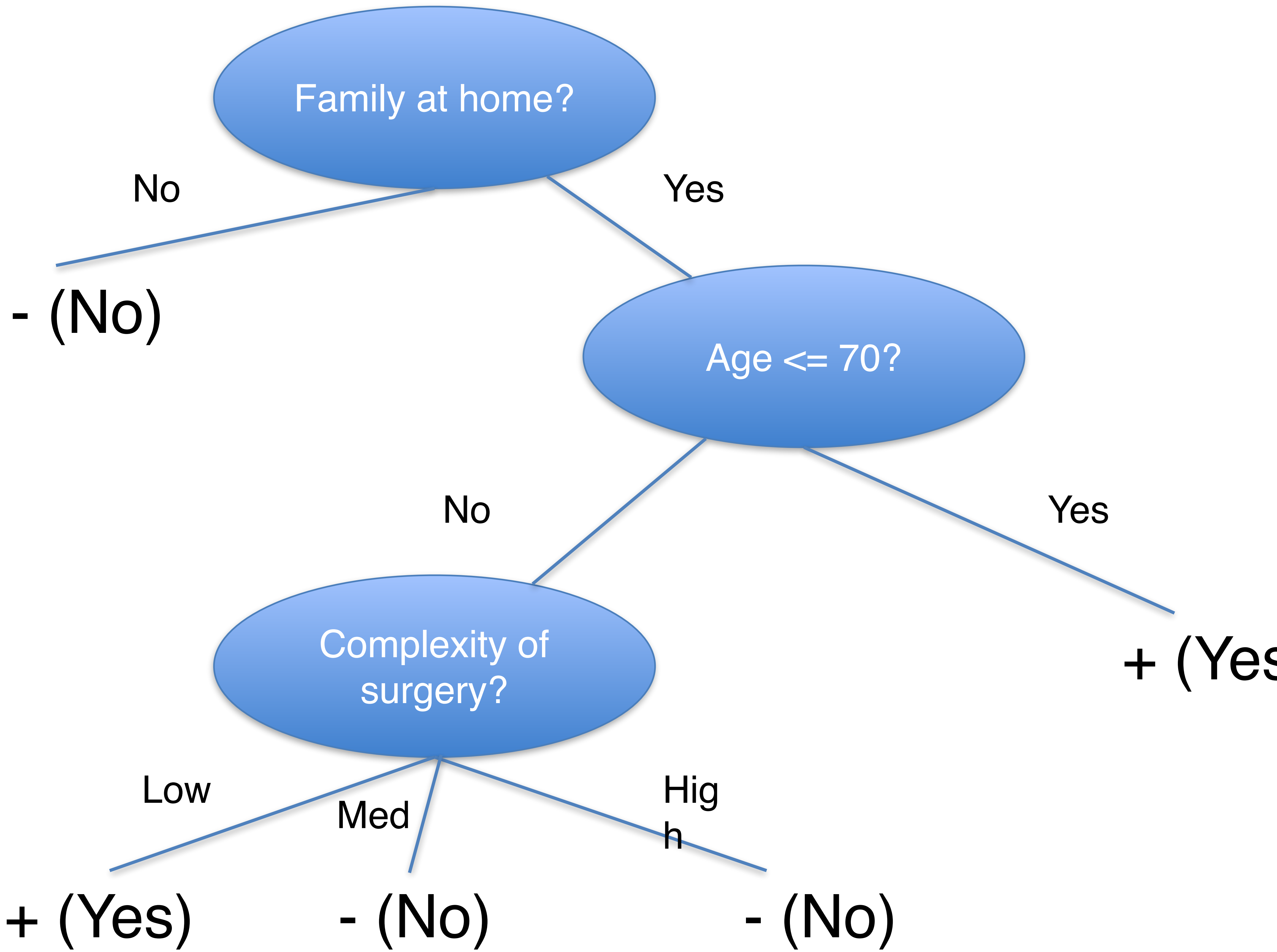
Lecture 16: Decision Tree Learning 2

Core 109S IDWT?, Spring 2017
Michael Hay

http://en.akinator.com/personnages/jeu

Recap

Decision tree for “Send patient home post-op?”



Learning a decision tree

Input: a collection of *labeled* examples

Output: a decision tree

Key problem: which attribute should be chosen?

If all examples have same value for target attribute*

The tree is a leaf that stores value of the target attribute

Else

Pick an attribute for the decision node

Construct one branch for each possible value of that attribute

Split examples: each branch gets subset of examples that agree with attribute value associated with that branch

For each branch, repeat this process on subset of examples assigned to that branch

* we may revise this condition later

Entropy-related measures

- entropy(X, D) the entropy of attribute X in dataset D
- segmentEntropy(X, Y, D) the remaining entropy of Y after we segment the data in D on attribute X
- InfoGain(X, Y, D) =
  entropy(Y,D) - segmentEntropy(X, Y, D)
- InfoGain measures how much information about Y is gained when we segment data based on X

Question

Instructions: ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

The first attribute selected for a decision node is the attribute with the highest information gain.

Claim: The second attribute selected for a decision node will be the attribute with the second highest information gain.

A. True

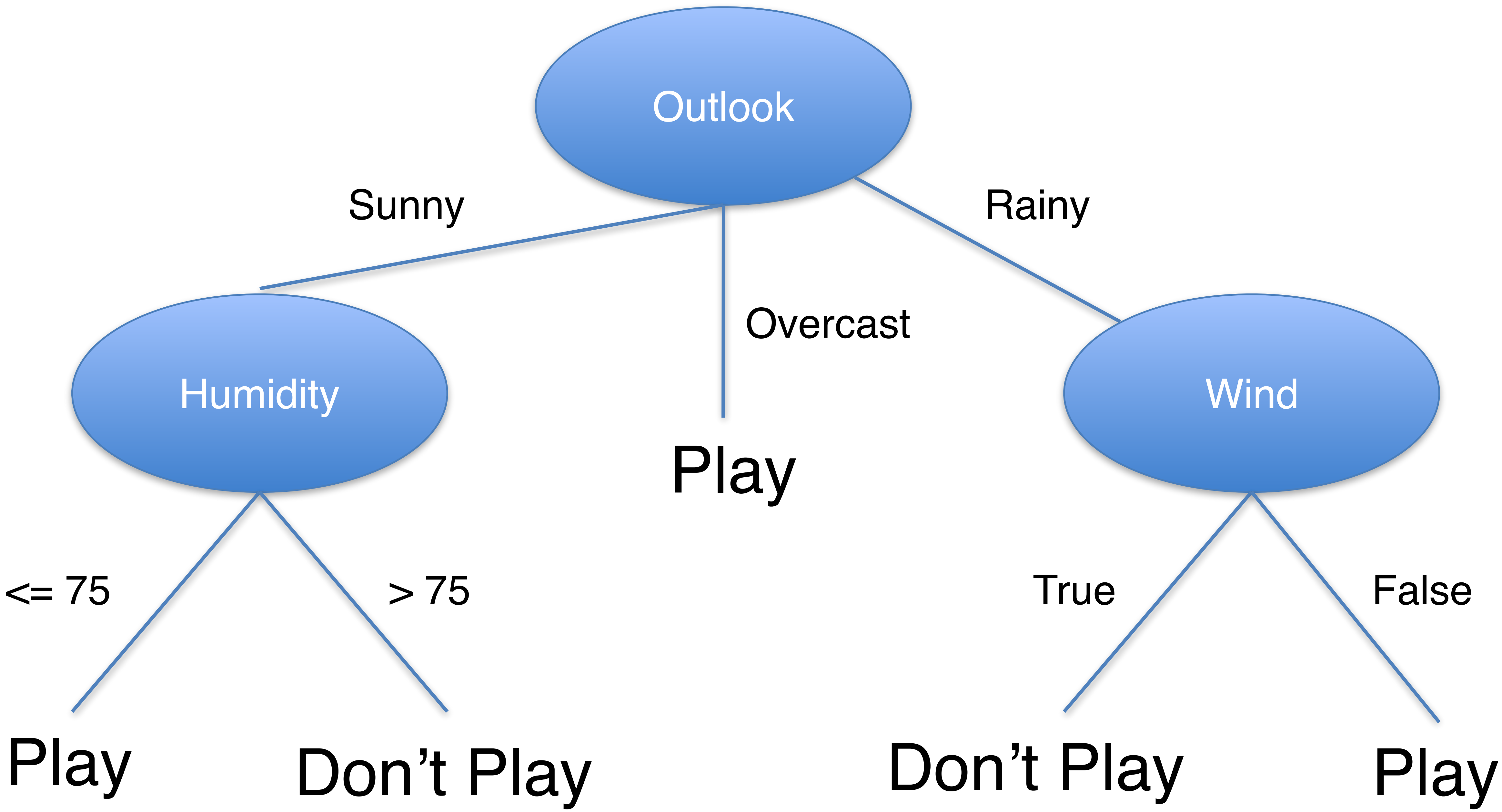
B. False

C. Not enough information given in problem

Today

- InfoGain for numeric attributes
- Comparison of trees vs. perceptrons

“Is it a good day to play golf?”



Attributes with Numeric Values: the Golf Tree

	Outlook	Temp	Humidity	Wind	Class
Example1	Sunny	85	85	False	Don't Play
Example2	Sunny	80	90	True	Don't Play
Example3	Overcast	83	88	False	Play
Example4	Rainy	70	96	False	Play
Example5	Rainy	68	80	False	Play
Example6	Rainy	65	70	True	Don't Play
Example7	Overcast	64	65	True	Play
Example8	Sunny	72	95	False	Don't Play
Example9	Sunny	69	70	False	Play
Example10	Rainy	75	80	False	Play
Example11	Sunny	75	70	True	Play
Example12	Overcast	72	90	True	Play
Example13	Overcast	81	75	False	Play
Example14	Rainy	71	96	True	Don't Play

Attributes with Numeric Values

- Split the numeric range into two groups:

  values <= threshold

  values > threshold
- How to select the threshold:
  - Sort the examples by the values of the attribute.
  - Search the examples, noting ***transition points***: places where adjacent examples belong to different classes.
  - The average value at transition points represent ***potential splits***.
  - Evaluate each ***split*** by applying the information gain formula.
  - Choose the best ***split***.
- Compare the gain for the best split against information gain for the remaining attributes.

Attributes with Numeric Values: the Golf Tree

Considering only the examples with Outlook=Sunny

	Humidity	Class
Example9	70	Play
Example11	70	Play
Example1	85	Don't Play
Example2	90	Don't Play
Example8	95	Don't Play

Only one transition point here: 70 to 85, potential split: 77.5, info gain?

Question

Instructions: ~1 minute to think/ answer on your own; then discuss with neighbors; then I will call on one of you

	Temp	Class
Example7	64	Play
Example6	65	Don't Play
Example5	68	Play
Example9	69	Play
Example4	70	Play
Example14	71	Don't Play
Example8	72	Don't Play
Example12	75	Play
Example10	75	Play
Example11	75	Play
Example2	80	Don't Play
Example13	81	Play
Example3	83	Play
Example1	85	Play

What are the *potential splits* for Temp?

- Search the examples, noting ***transition points***: places where adjacent examples belong to different classes.
- The average value at transition points represent ***potential splits***.

Today

- InfoGain for numeric attributes
- Comparison with perceptrons

Example: perceptron

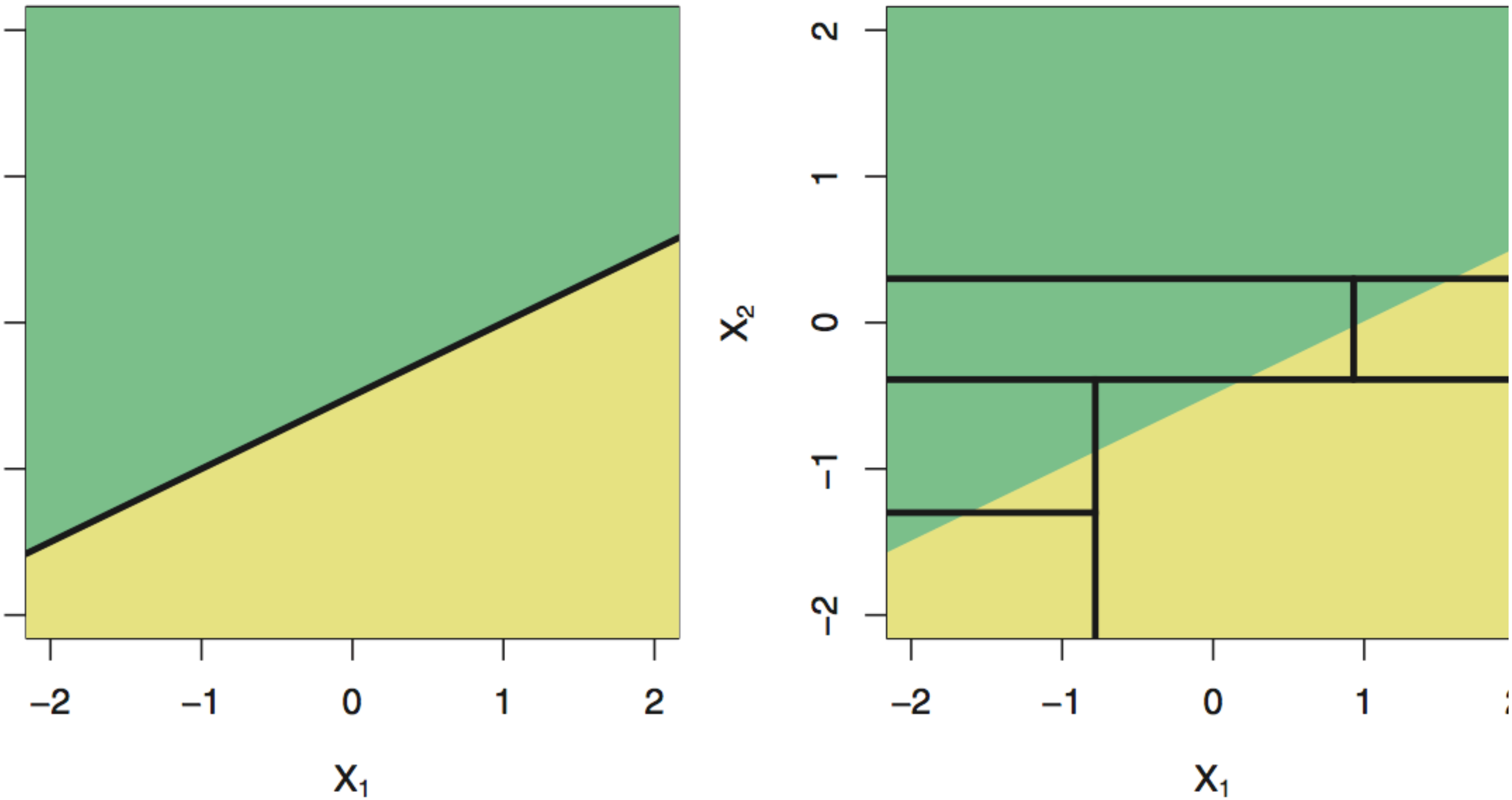
- Given input x = (x_1, …, x_d) *(the d attributes of customer app.)*
- Assign a **weight** w_i for each attribute value x_i. and choose some **threshold** value

$$\text{Approve credit if} \quad \sum_{i=1}^{d} w_i x_i \geq \text{threshold}$$

$$\text{Deny credit if} \quad \sum_{i=1}^{d} w_i x_i < \text{threshold}$$

Credit: Yaser Abu Mostafa https://work.caltech.edu/

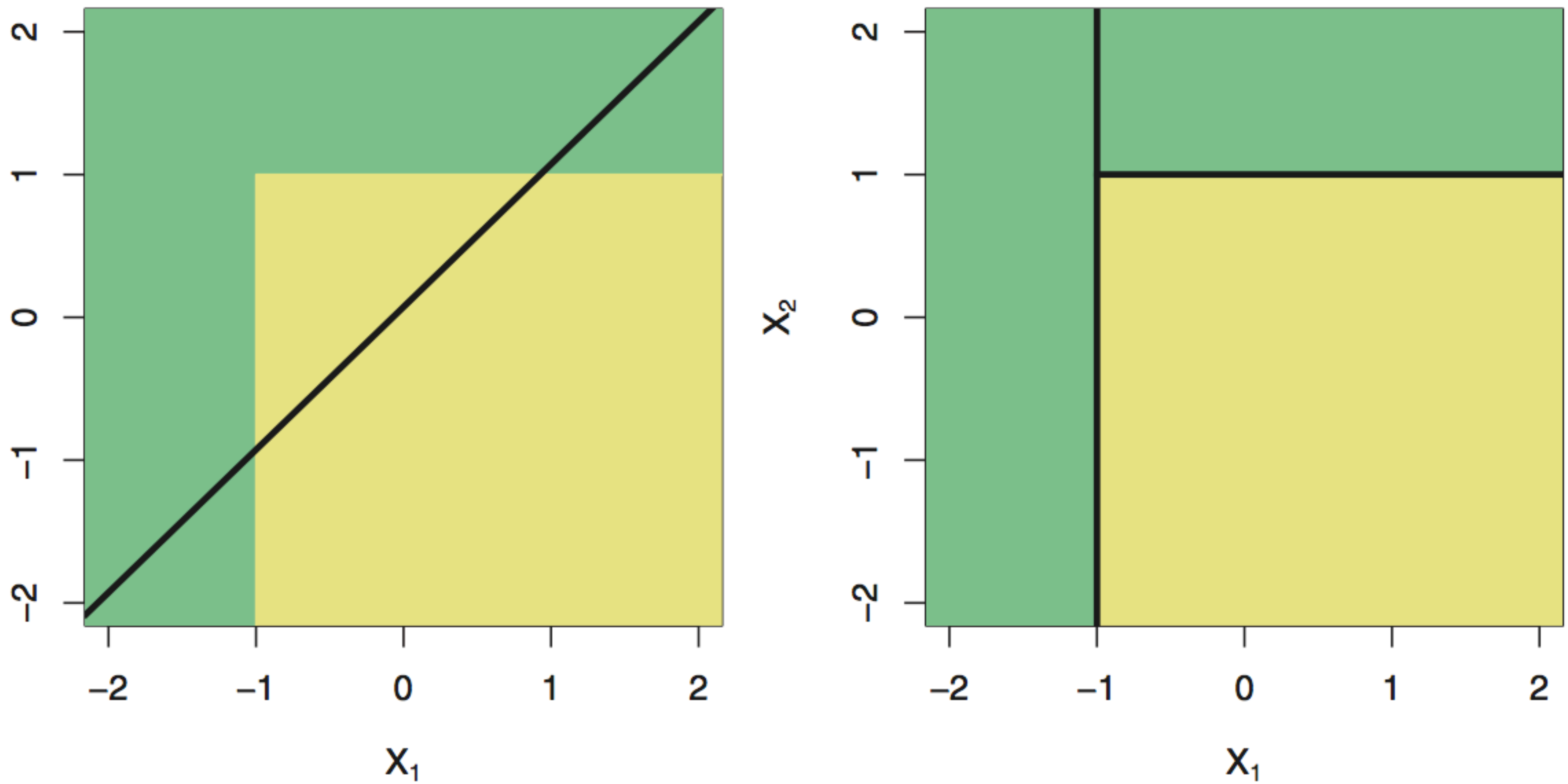
Linear vs. trees



True boundary is linear; tree can only approximate with axis-parallel splits

Credit:James et al. http://www-bcf.usc.edu/~gareth/ISL/

Linear vs. trees



True boundary is non-linear: linear model (perceptron) gives poor approximation

Credit:James et al. http://www-bcf.usc.edu/~gareth/ISL/

Comparison

Perceptron	Decision Tree
Parameteric	Non-parametric
Numerical features	All feature types
Finds separating line	Partitions space into regions (finer with more data)
"Robust": similar line on two different samples of data	Non-robust: shape of tree can change a lot on different samples

Summary of trees

- Key idea behind decision trees is *segmentation:* we are splitting training dataset into subset based on value of a selected attribute
- We can use *information gain* to select the "best" attribute to split on
- For numeric attributes, we can consider "transition points"
- Note: for numeric attributes, it may make sense to split on the same attribute multiple times within the same tree.

Discuss HW 2