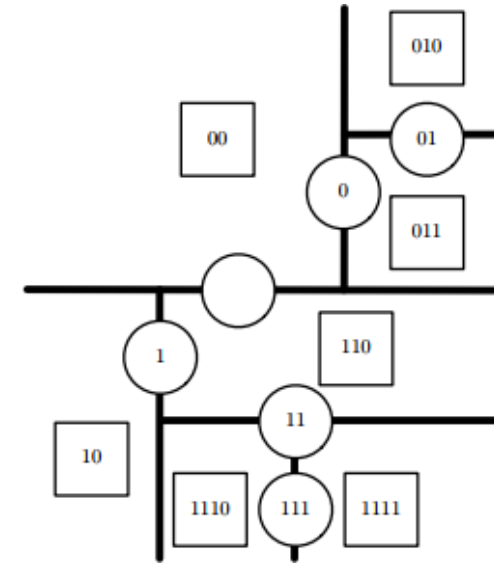
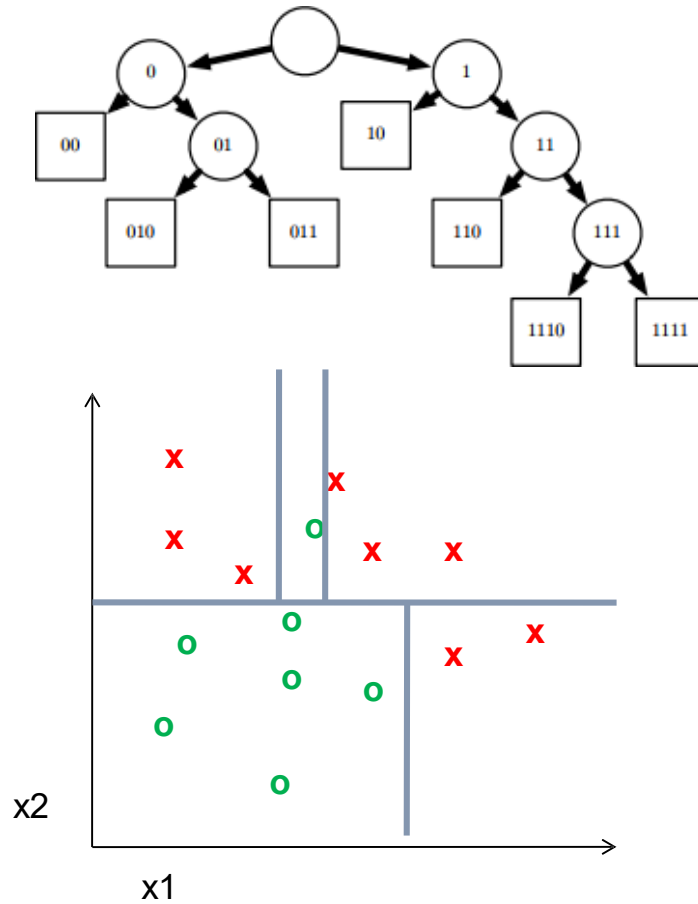




The Wisdom of the Crowd: On Ensemble Methodology

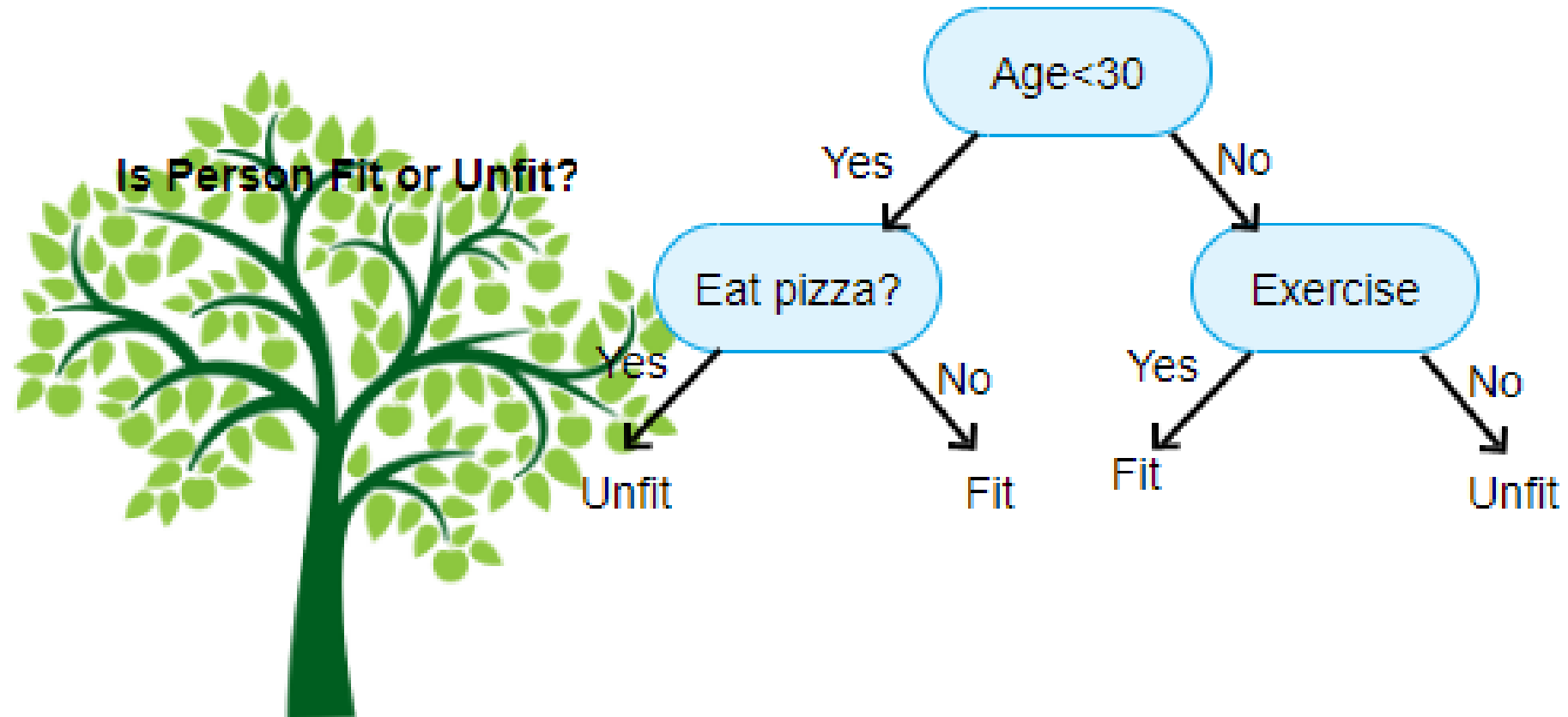
Alexander G. Ororbia II
Introduction to Machine Learning
CSCI-335
4/10/2026

Decision Trees – Aggregated Piecewise Functions



Divides (input) space into regions (nodes on lines categorize samples, leaf nodes on regions correspond to examples received). Piecewise-constant function (cannot learn function with more local maxima than number of training samples).

Decision Trees – Interrogating Your Data!



On Continuous Variables & Gain Ratio

(Yay, More White Board!)

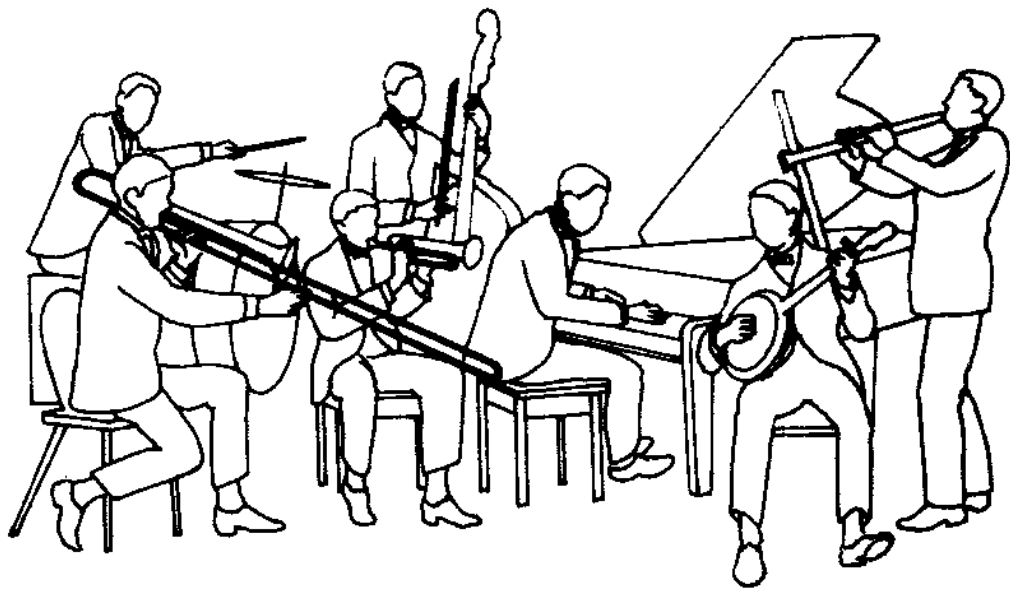
Class-Labeled Training Tuples from the *AllElectronics* Customer Database

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

**Ok, but is one tree
enough?**

Ok, but is one tree
enough?

No!



An ensemble is kinda like a garage band...



Some aspiring machine intelligence scientists can be musicians too, you know!



Good ol' Led Zep!

The Wingmen

Ensembling: Bagging and Boosting

- Ensembles can be very powerful
 - Put together all classifiers (weak + strong learners)
 - **Intuition** = different learners will pick up on different (statistical) regularities in feature space
 - Combine different perspectives to build a more intelligent complex system (a "committee machine")
 - Ensemble = more powerful than component classifiers
- **Bagging & boosting** = create data subsets through sampling with replacement
 - Build a classifier on each subset
 - Bagging, introduced by Breiman (1996), stands for "**bootstrap aggregating"**
 - *Bootstrap* => sampling with replacement

Methods of Combining Models

- Classification/regression performance can be improved by combining multiple models
 1. Committee: train L different models
 - Make predictions by average of the predictions
 2. Boosting: Variant of Committee
 - Train multiple models in sequence
 - Error function used to train a model depends on performance of previous models
 3. Decision Tree
 - Select one of L models to make the prediction
 - Choice of model is a function of input variables

Enlarging the Hypothesis Space

Combinations of Simple Algorithms: Powerful

e.g. using linear separators, can represent non-linearly separable attribute regions

- *Enlarges the hypothesis space* (i.e. space of representable functions)

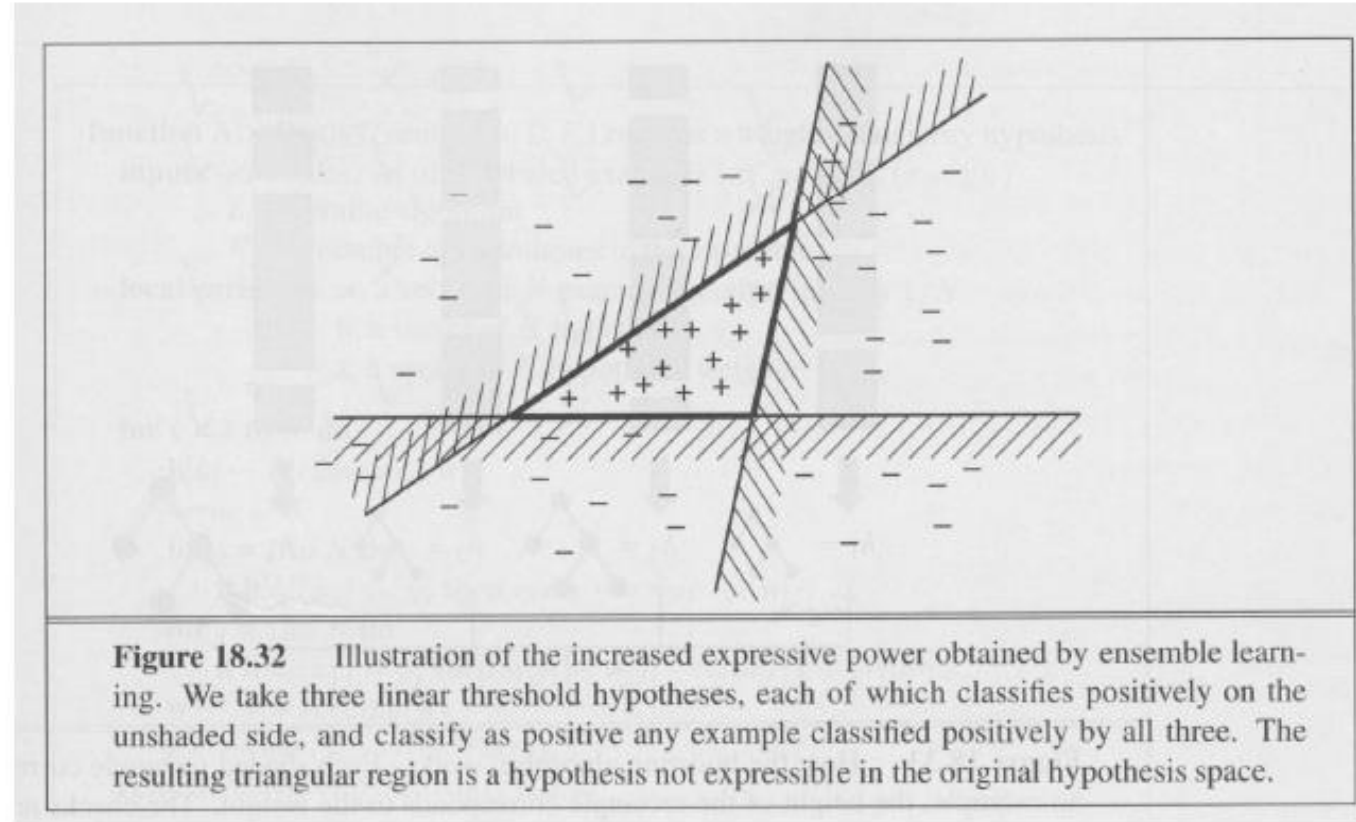


Sort of like a
“deforested”
learning algorithm?

Decision stumps (single decision tree nodes partitioning attributes) used frequently to learn complex functions

If base algorithm supports simple and efficient learning (e.g. decision stump), ensemble permits complex functions to be learned without high algorithmic or computational complexity

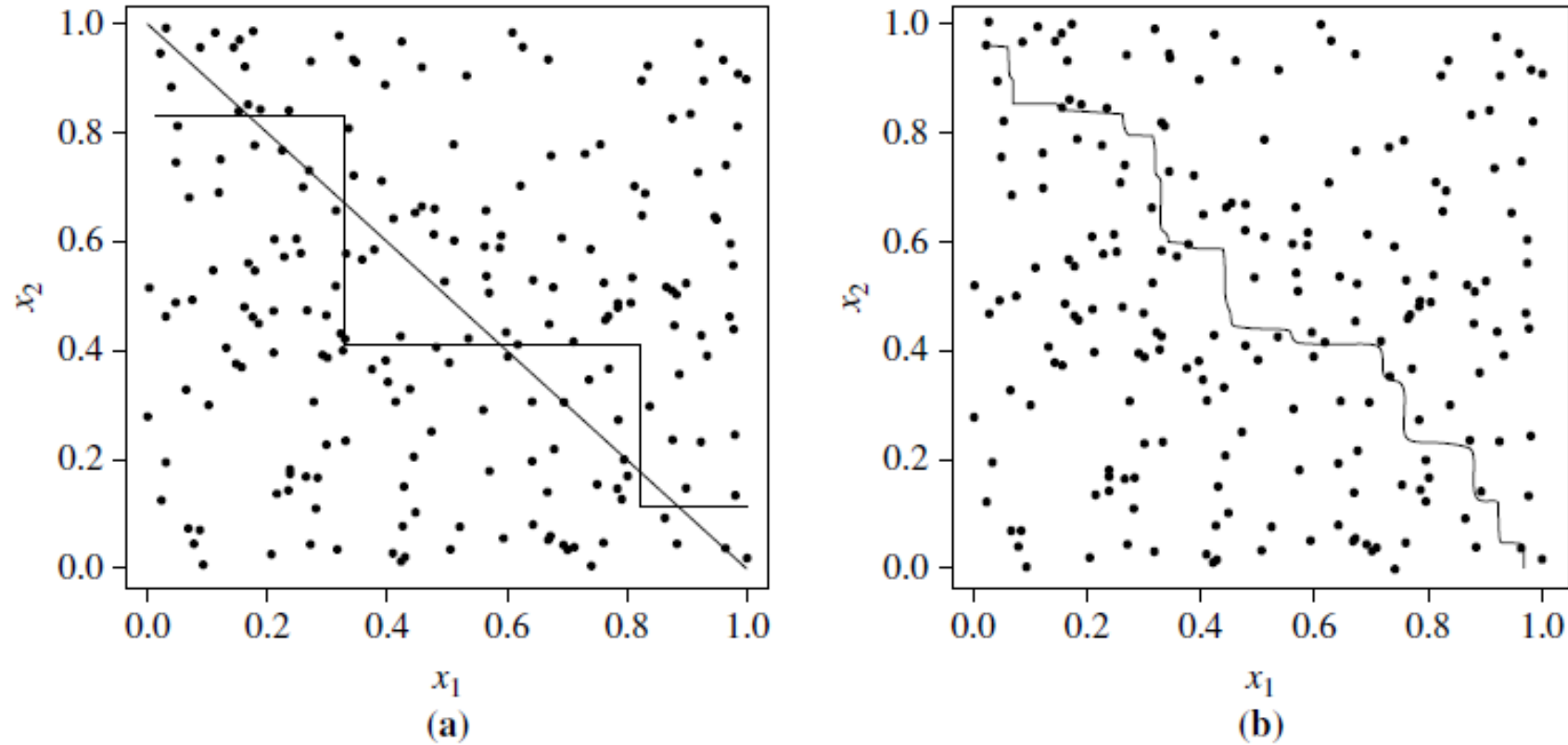
Example of an Ensemble



An *ensemble* consisting of three linear separations of 2D data for a two-class classification problem

Ensemble Methodology (White Board)

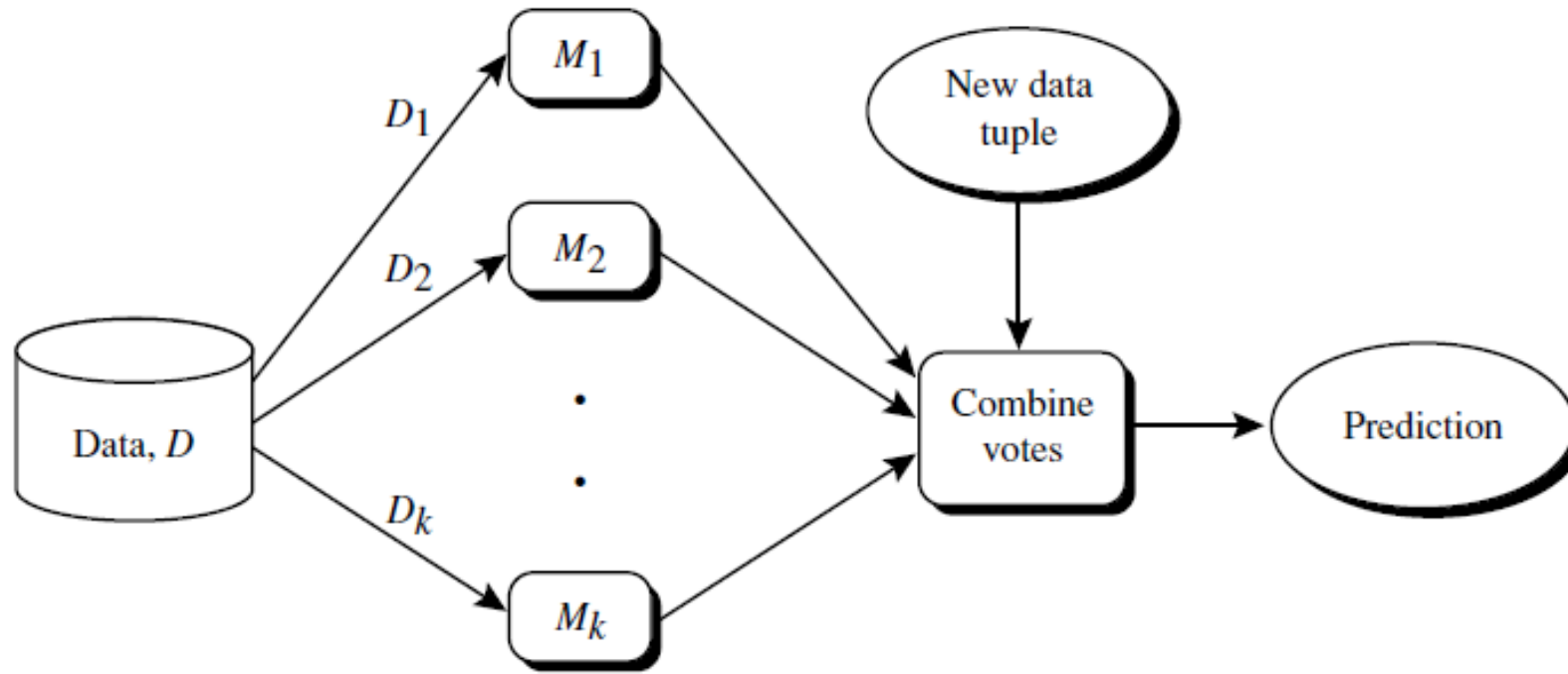
Decision Boundary of an Ensemble



Decision boundary by (a) a single decision tree and (b) an ensemble of decision trees for a linearly separable problem (i.e., where the actual decision boundary is a straight line). The decision tree struggles with approximating a linear boundary. The decision boundary of the ensemble is closer to the true boundary. *Source:* From Seni and Elder [SE10]. © 2010 Morgan & Claypool Publishers; used with permission.

Ensemble Methodology, Bagging (White Board)

A *Bagging* Ensemble System



Increasing classifier accuracy: Ensemble methods generate a set of classification models, M_1, M_2, \dots, M_k . Given a new data tuple to classify, each classifier “votes” for the class label of that tuple. The ensemble combines the votes to return a class prediction.

Questions?
QUESTIONS?

