



Lecture 33

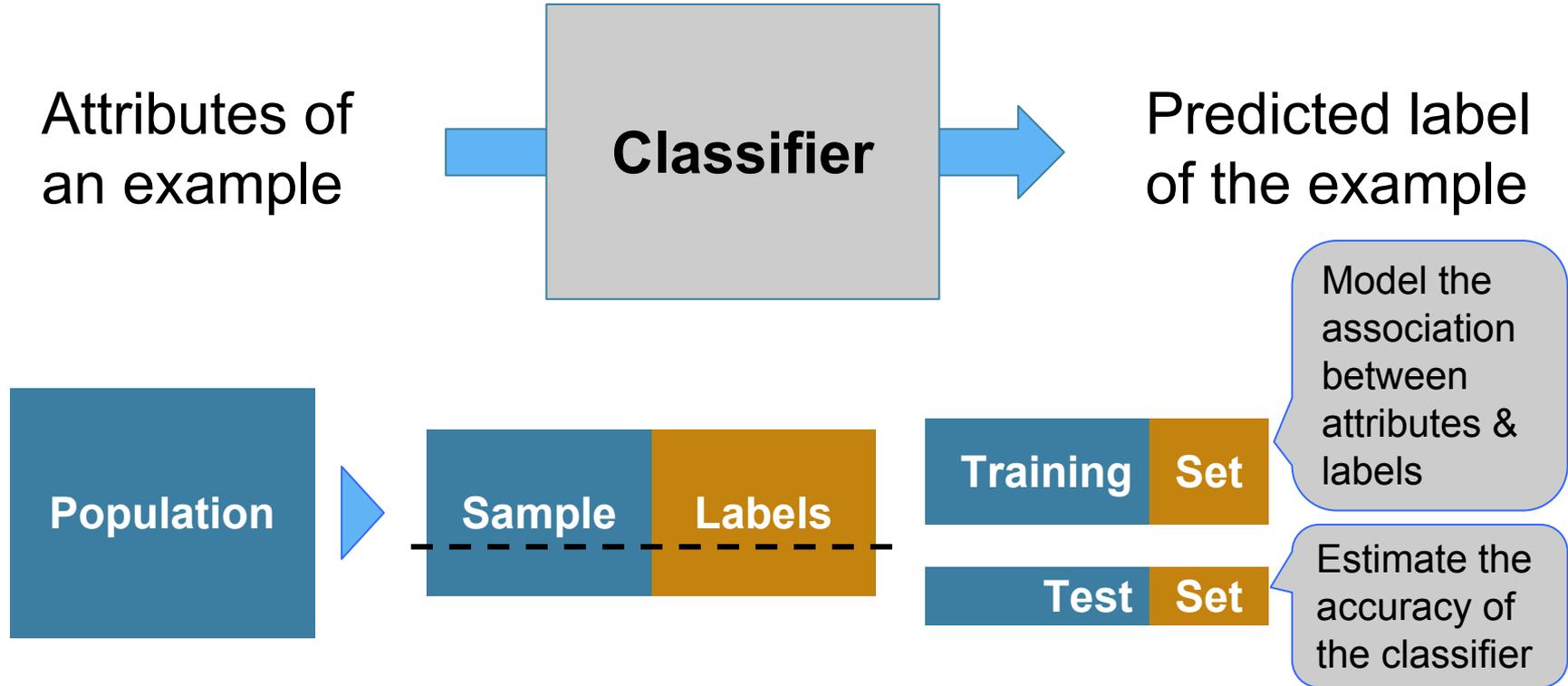
Machine Learning

Slides created by Vinitra Swamy (vinitra@berkeley.edu)
Contributions from John DeNero (denero@berkeley.edu) and Ani Adhikari (adhikari@berkeley.edu)

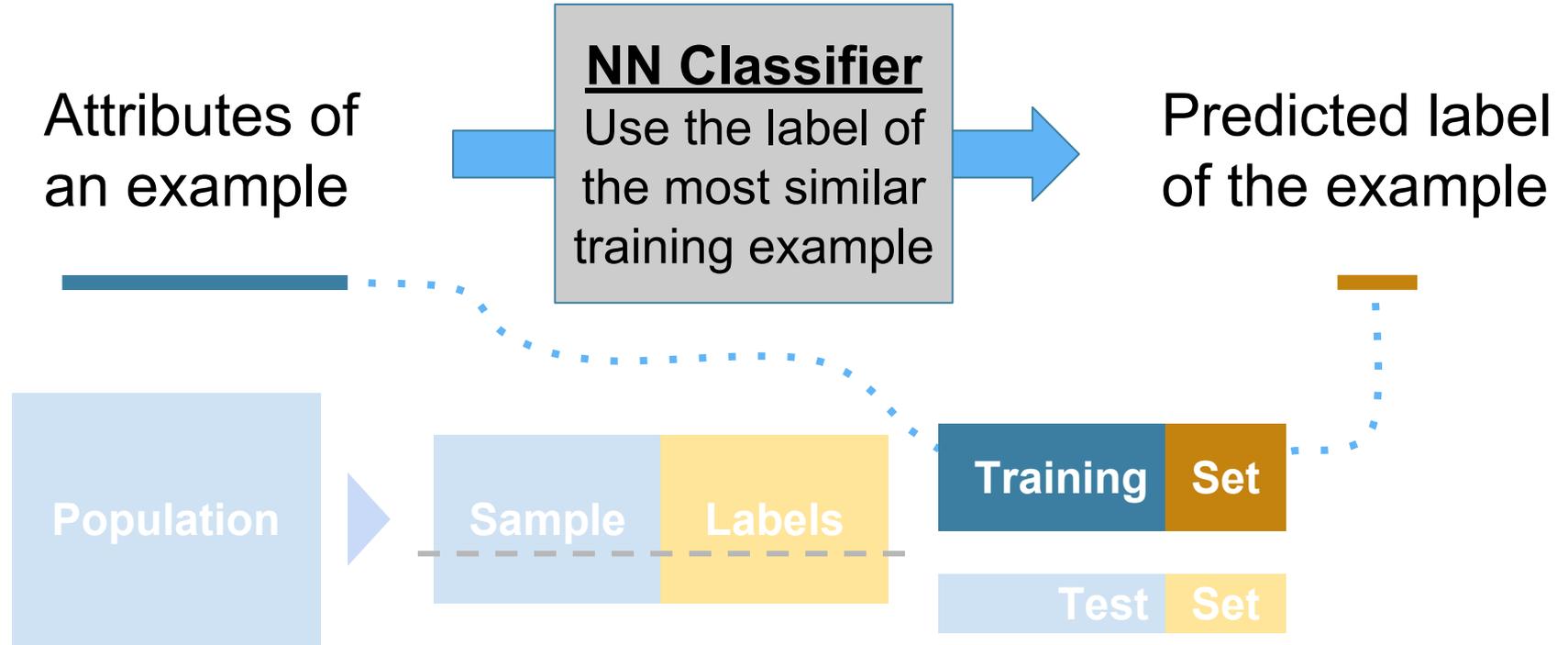
Announcements

Review: Classifiers

Training a Classifier



Nearest Neighbor Classifier



Finding the k Nearest Neighbors

To find the k nearest neighbors of an example:

- Find the distance between the example and each example in the training set
 - Augment the training data table with a column containing all the distances
 - Sort the augmented table in increasing order of the distances
 - Take the top k rows of the sorted table (Demo)
-

The Classifier

To classify a point:

- Find its k nearest neighbors
- Take a majority vote of the k nearest neighbors to see which of the two classes appears more often
- Assign the point the class that wins the majority vote

(Demo)

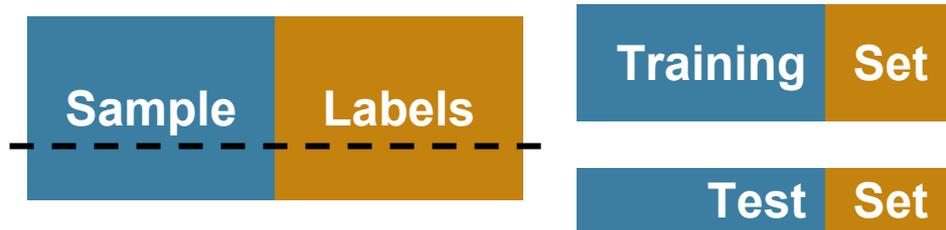
Evaluation

Accuracy of a Classifier

The accuracy of a classifier on a labeled data set is the proportion of examples that are labeled correctly

Need to compare classifier predictions to true labels

If the labeled data set is sampled at random from a population, then we can infer accuracy on that population



(Demo)

Machine Learning

What is ML?

A **machine learning** algorithm enables a computer to

- identify patterns in observed data
- build models that explain the world
- and predict things without having explicit pre-programmed rules and models.

All you'll need to know from this lecture -- the difference between supervised and unsupervised ML

Supervised Machine Learning

Input: Labeled data

Output: Prediction for unlabeled example

High computational complexity

Unsupervised Machine Learning

Input: Unlabeled data

Objective: Recognize underlying patterns in data

Low computational complexity

Semi-Supervised Machine Learning

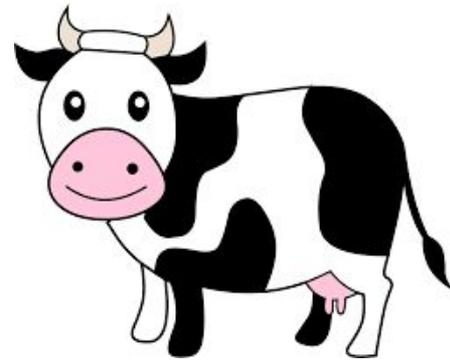
Input: Some labeled data, but majority unlabeled



Dog
4 legged
animal

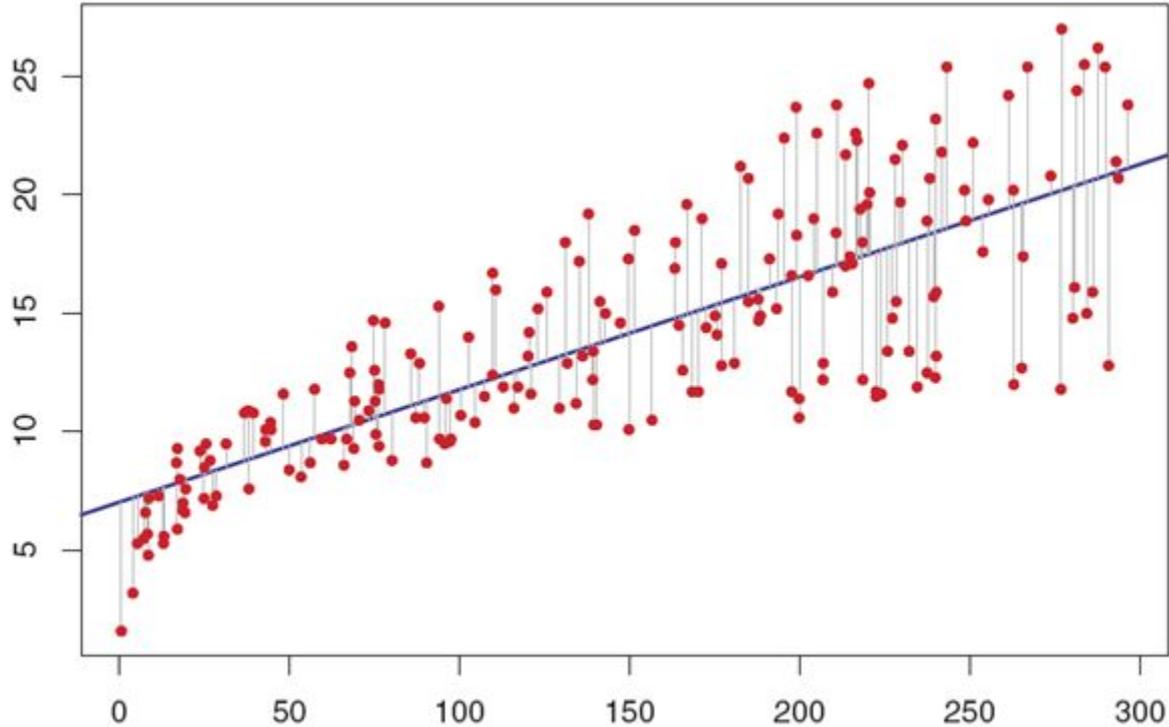


Dog



Dog

What we've learned: Regression



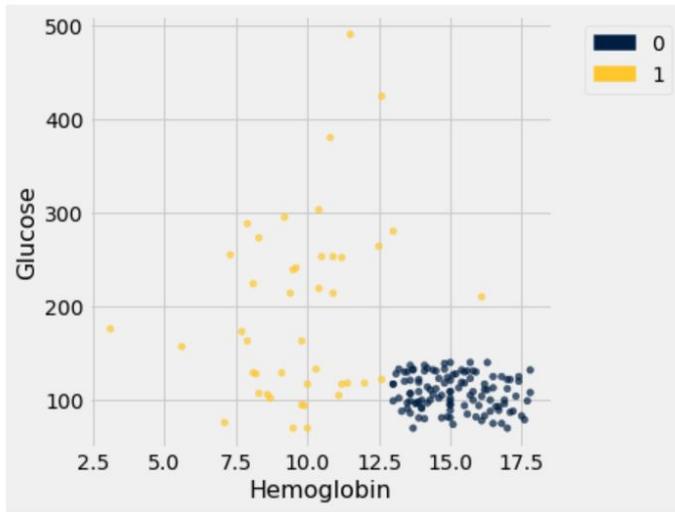
Is Linear
Regression
supervised?

Yes!

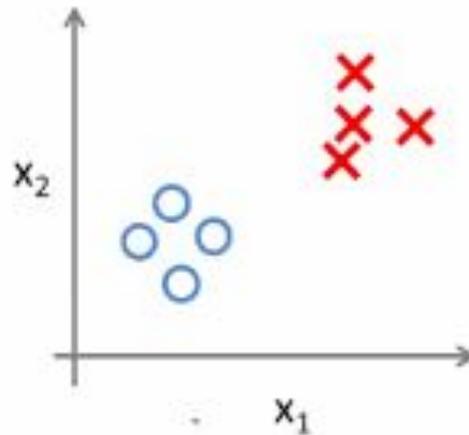
What we've learned: Classification

Is Classification supervised? Yes!

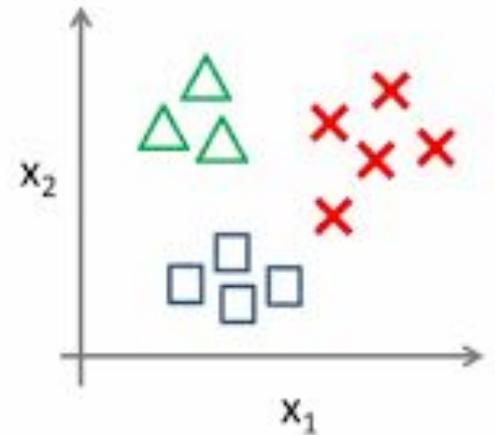
```
ckd.scatter('Hemoglobin', 'Glucose', colors='Class')
```



Binary classification:



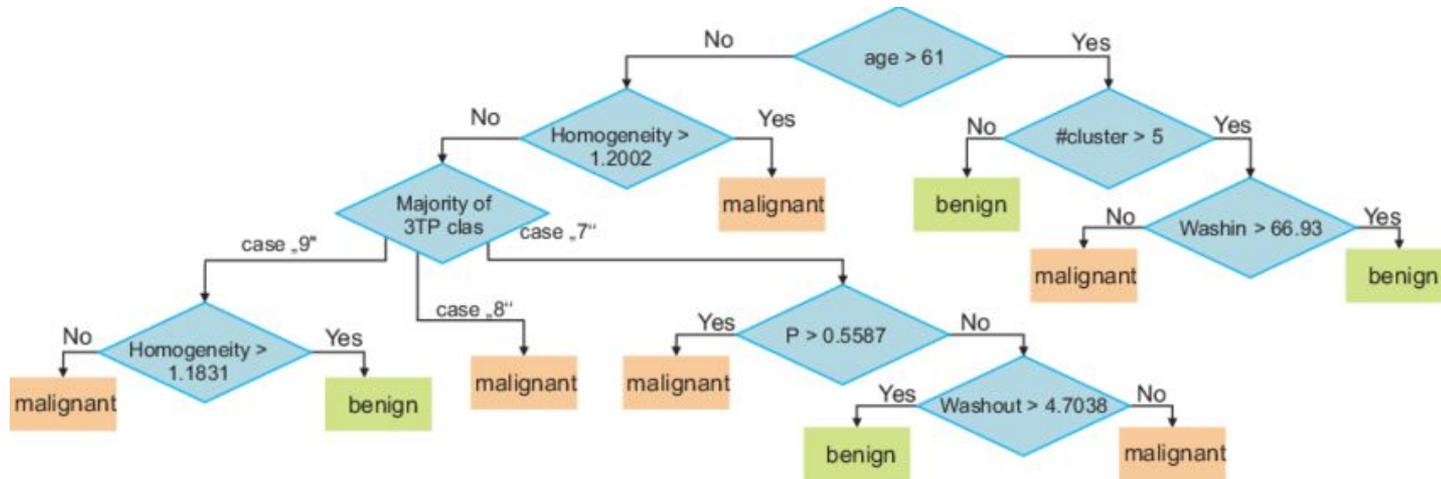
Multi-class classification:



**Interesting Material (that will
not be tested!)**

Other Interesting Techniques

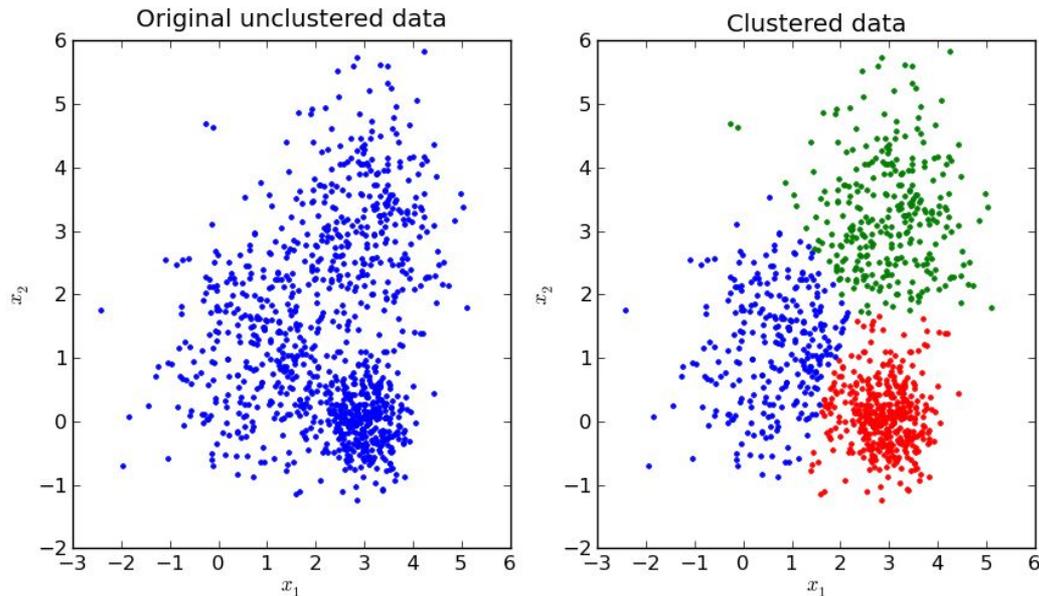
Decision Trees -- supervised? Yes!



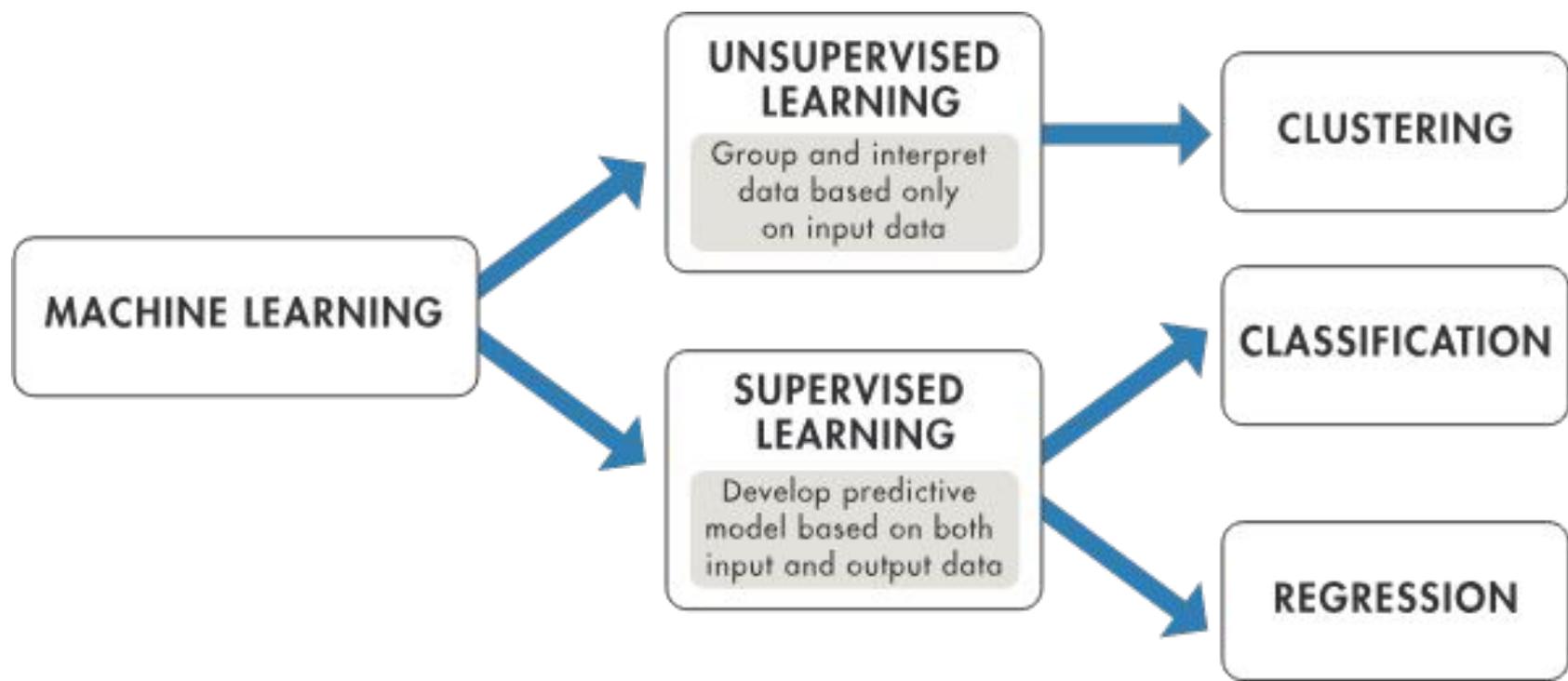
If you like this, take DATA 100, STAT 154, CS 189

Other Interesting Techniques

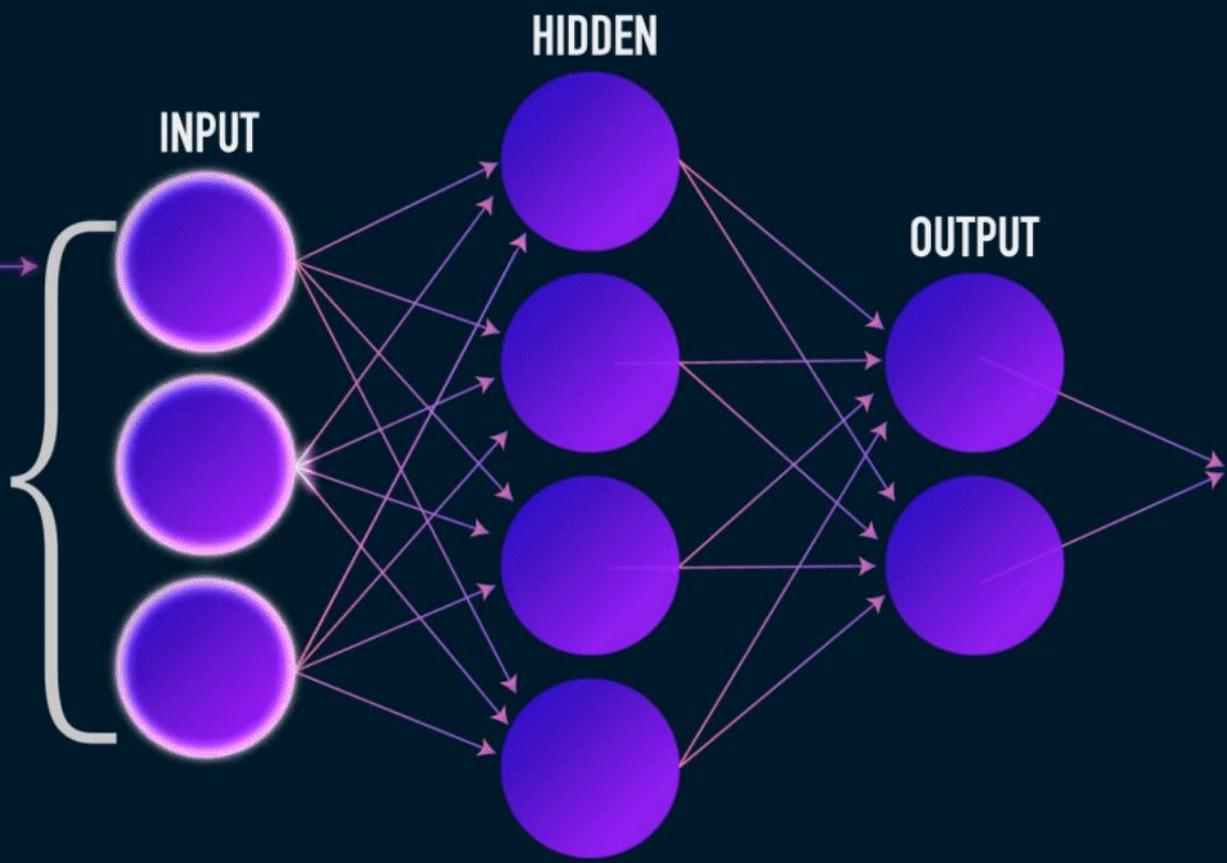
Clustering -- supervised? No!

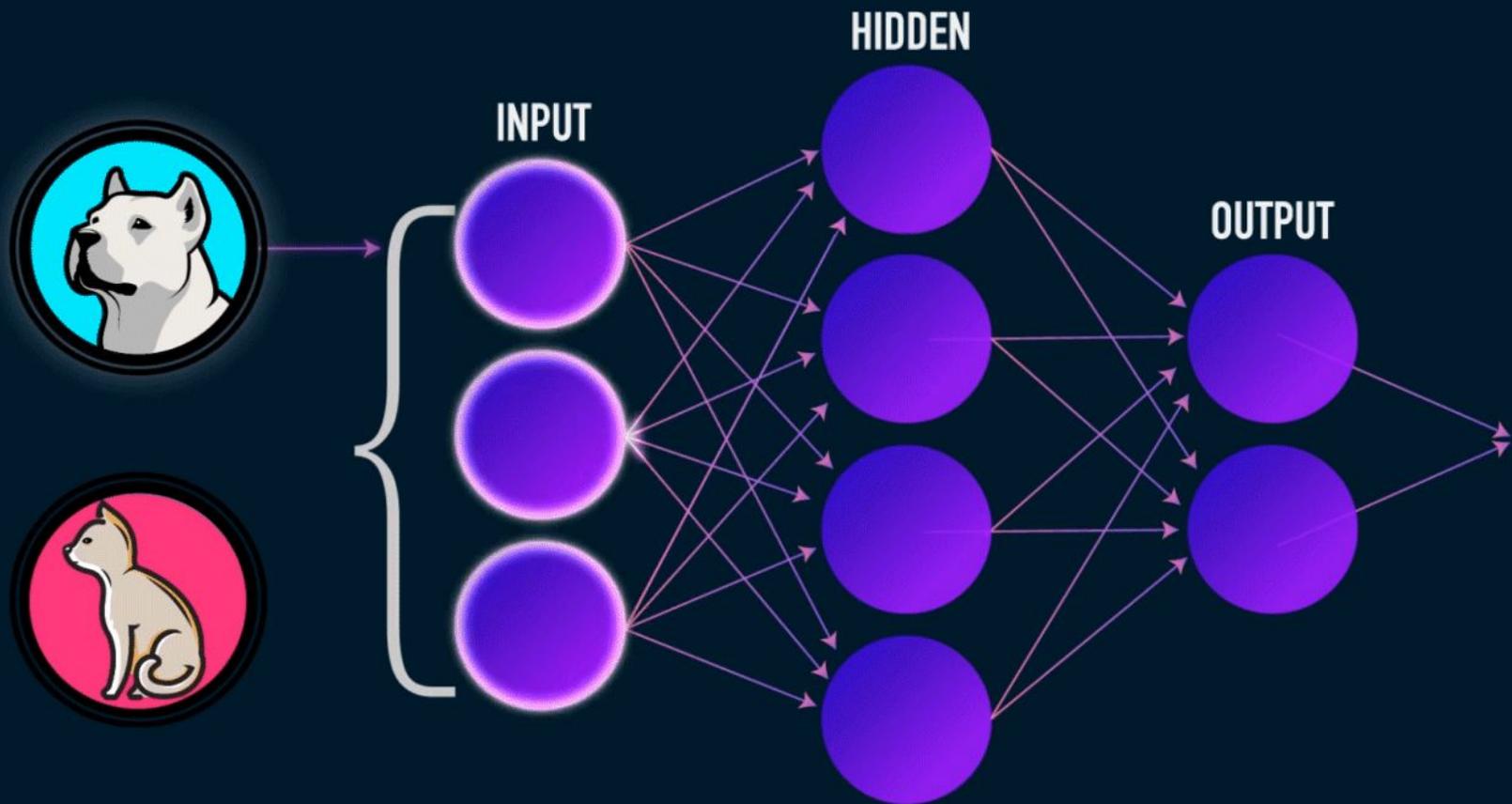


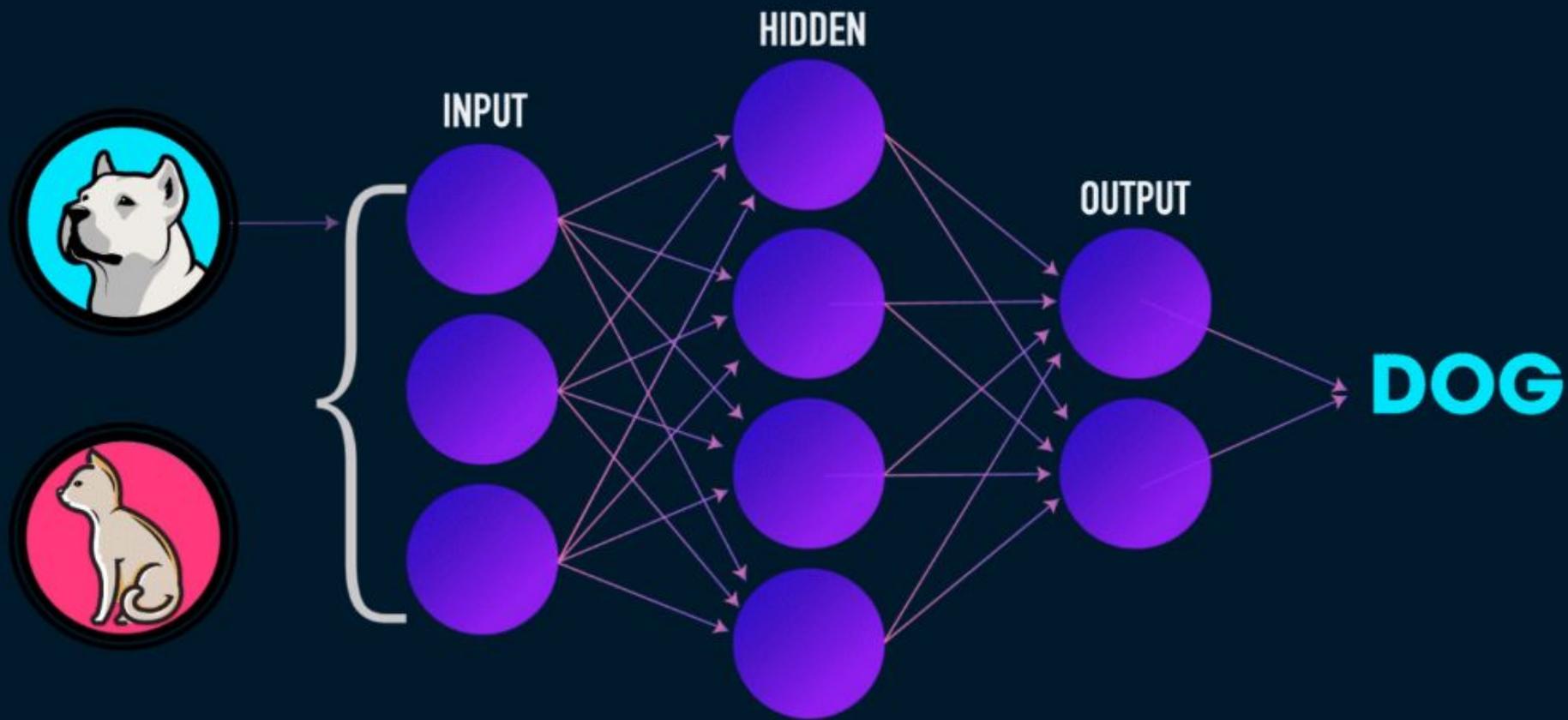
If you like this, take DATA 100, STAT 154, CS 189

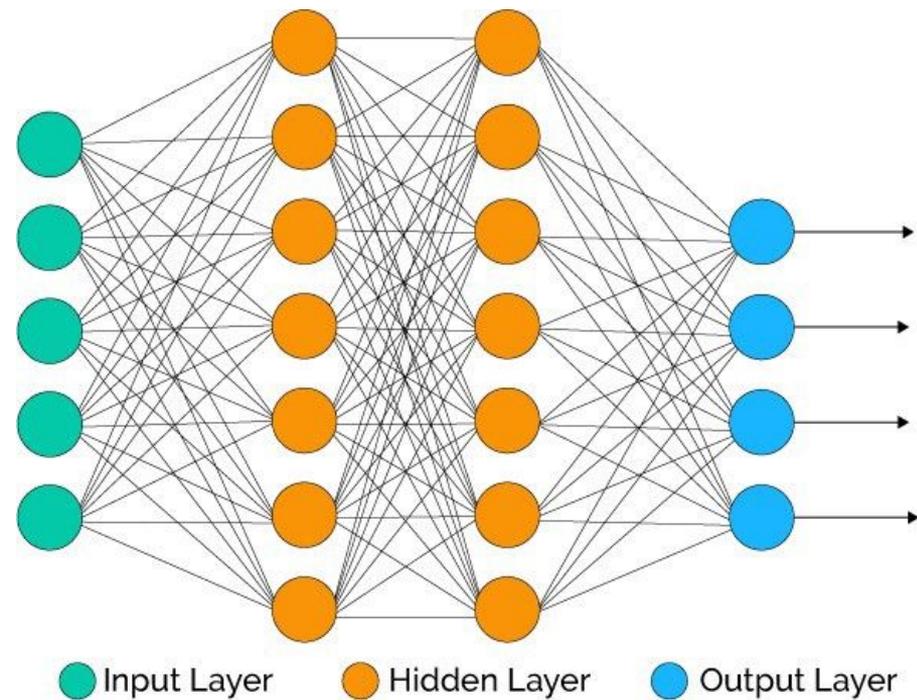


Neural Networks





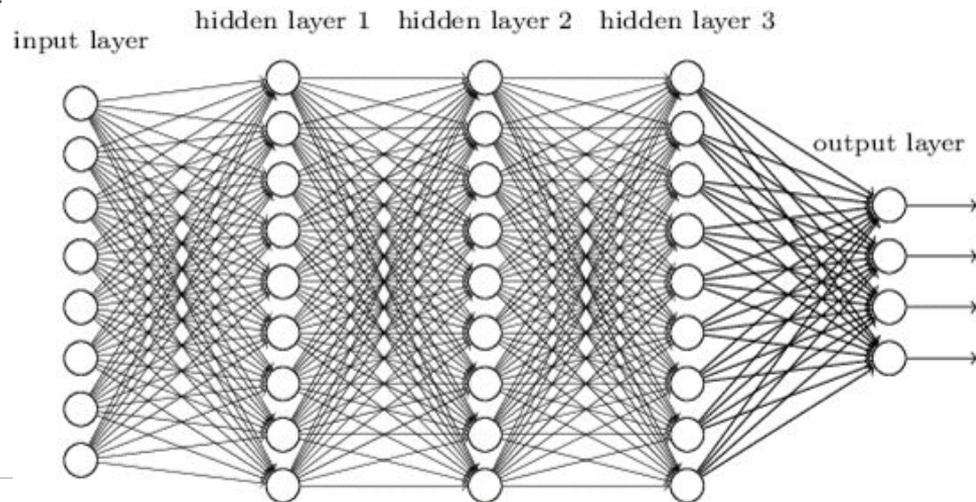




- Interpretability
- Security

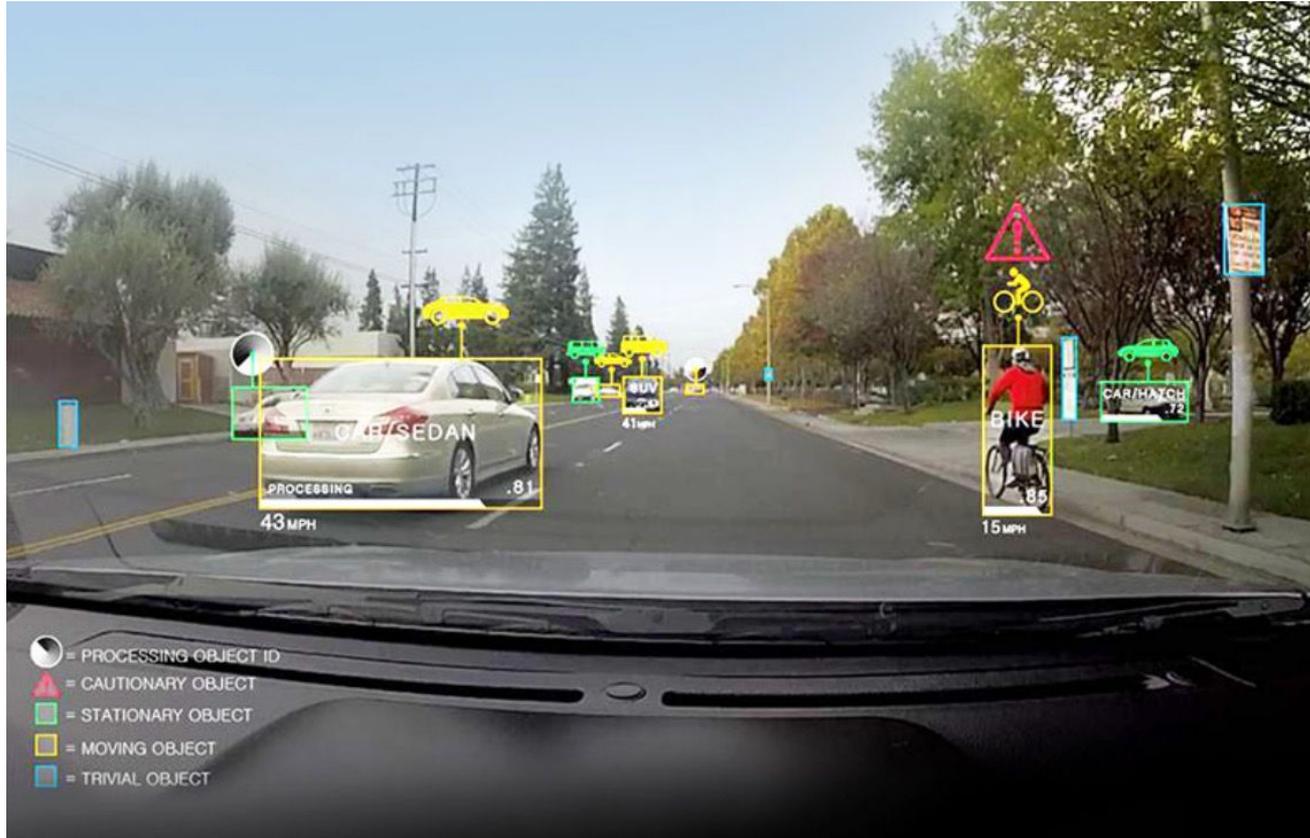
Deep Learning

Supervised? Yes!



Data Science problems for the next 10 years

Self-Driving Cars



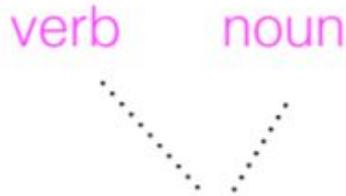
State of the Art:
Deep Learning +
Computer Vision



Natural Language Processing

How can a computer read a book?

- Machine Translation
- Question Answering
- Ambiguity



“One morning I shot
an elephant in my pajamas”

Can a computer play
Jeopardy?

Yes! IBM Watson can.



If you like this, take INFO 159 (NLP), L&S 88 (Literature + Data Connector)

Conversational Agents



If you like this, take CS 188 (AI) and look into Human-Computer Interaction

And many more...

- Education
- Social Science
- Humanities
- Economics
- Environmental Science
- (We'll never finish listing them all)

You are data scientists now -- go out and change the world!
