TYPES OF ML ALGORITHMS

Lara J. Martin (she/they)
TA: Aydin Ayanzadeh (he)

11/30/2023

CMSC 671

By the end of class today, you will be able to:

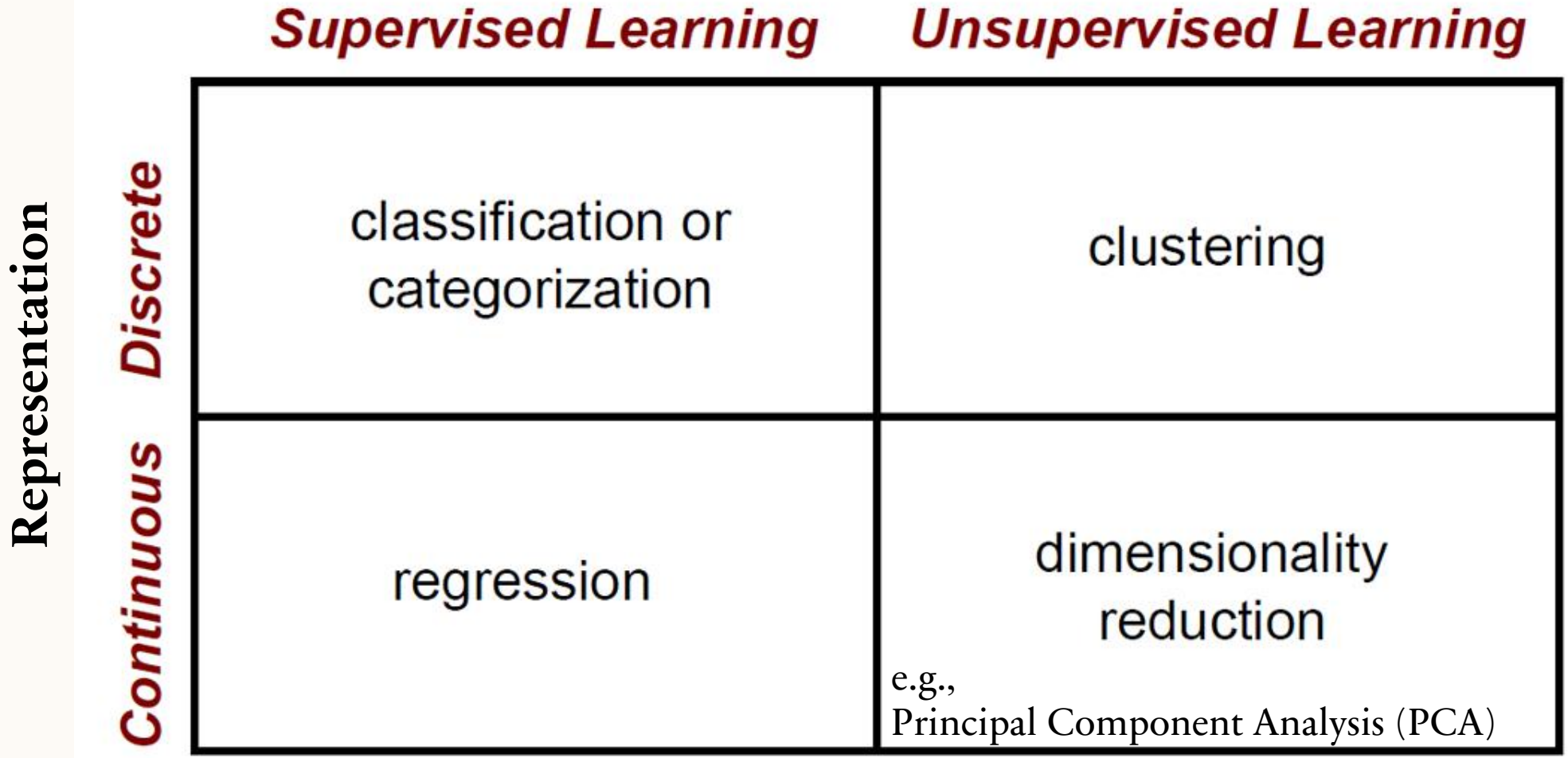
- Distinguish between types of machine learning algorithms (in more detail)
- Determine the classes and features of a given algorithm
- Determine whether a given algorithm is supervised or unsupervised

Modified from slides by Dr. Cynthia Matuszek

COURSE UPDATES

- Milestone 3 is due December 7th (a week from today)
  - Draft of paper, official conference template
  - Get preliminary results

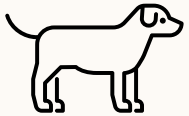
REVIEW: TYPES OF ALGORITHMS



REVIEW:
DIVIDING UP DATA FOR TRAINING

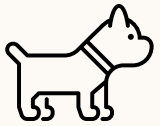

cat

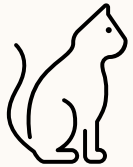

mouse


dog

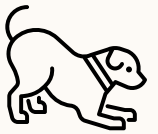

hamster

[position of lines]
[loops and dots]
[any other information relevant to our problem]


dog


cat


mouse


dog

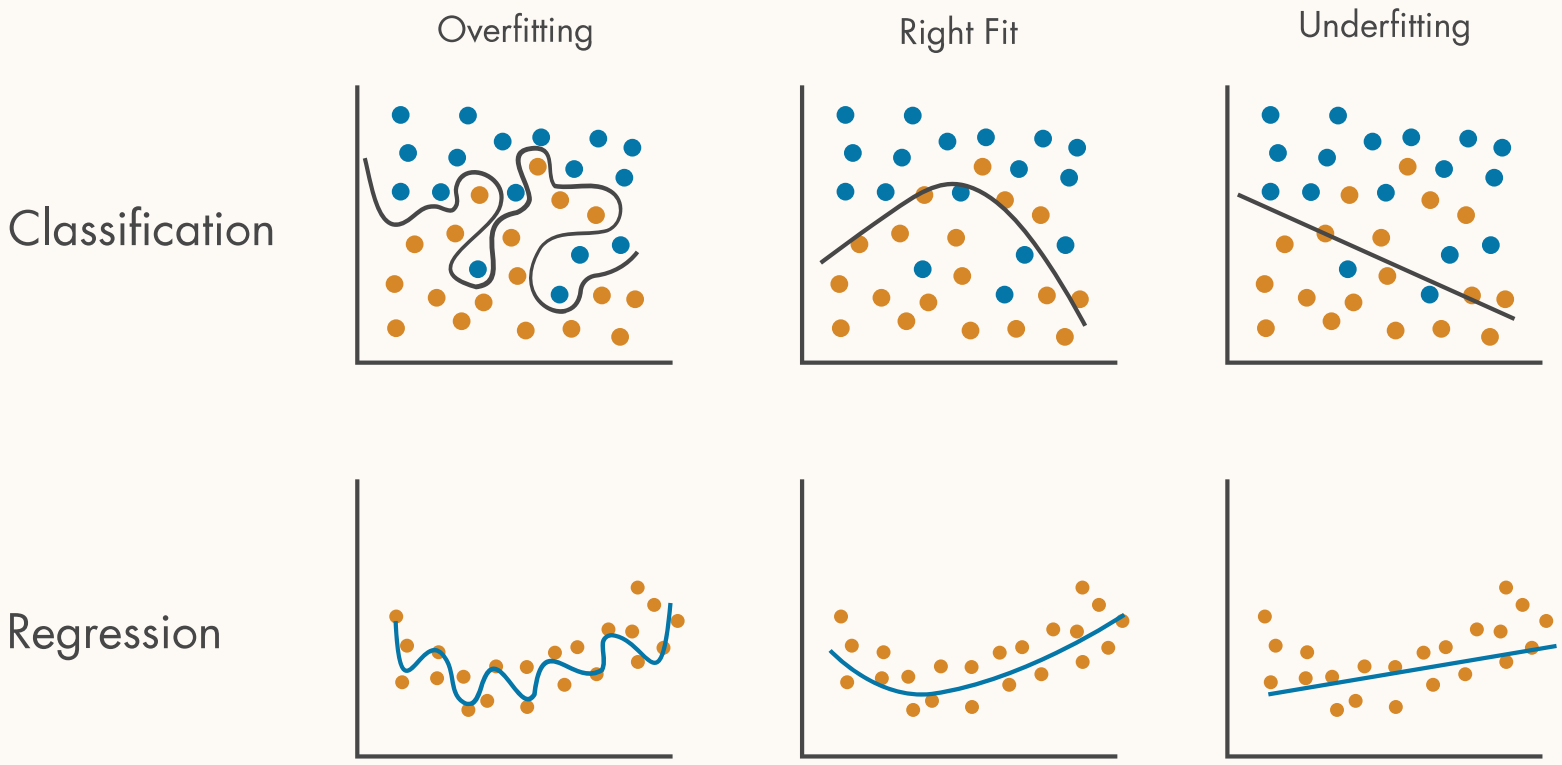
Training
~80%

Validation/Development
~10%

Testing
~10%

Why would we do this? To prevent overfitting!

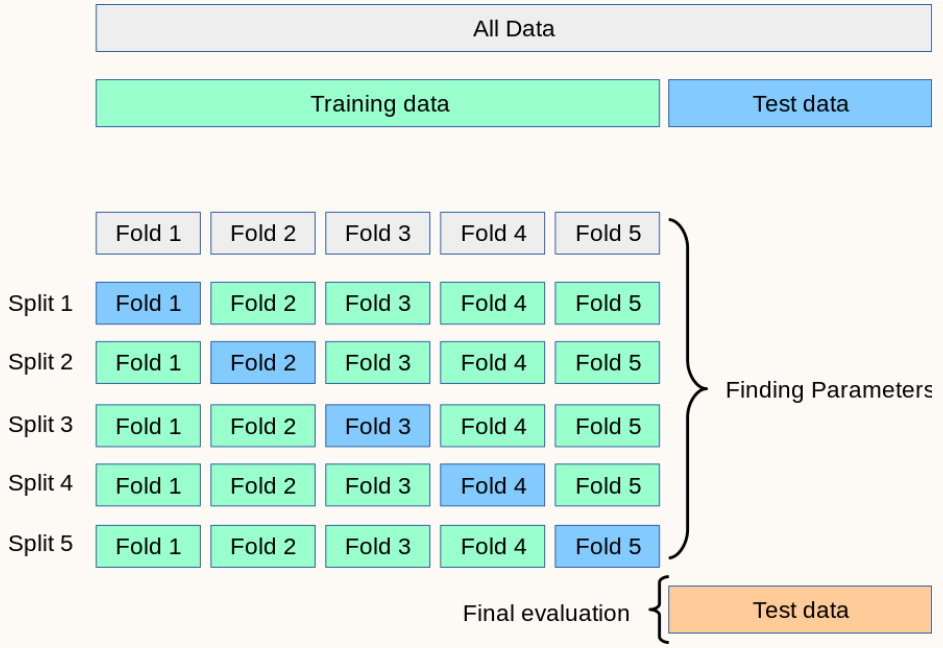
OVERFITTING



https://www.mathworks.com/discovery/overfitting.html

K-FOLD CROSS-VALIDATION

- Another way of avoiding overfitting
- Not usually helpful when you need a model for generation → better for data analysis



https://scikit-learn.org/stable/modules/cross_validation.html

THE MACHINE LEARNING FRAMEWORK

$$y = f(\mathbf{x})$$

output — prediction function — feature(s) of input

- **Training:**
  given a *training set* of labeled examples $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$,
  estimate the prediction function f
  by minimizing the prediction error on the training set
- **Testing:**
  apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

Slide credit: Svetlana Lazebnik

STEPS

Training

Training Images

Training Labels

Validation Labels

Image Features

Training

Validation

Learned model

Testing

Test Image

Image Features

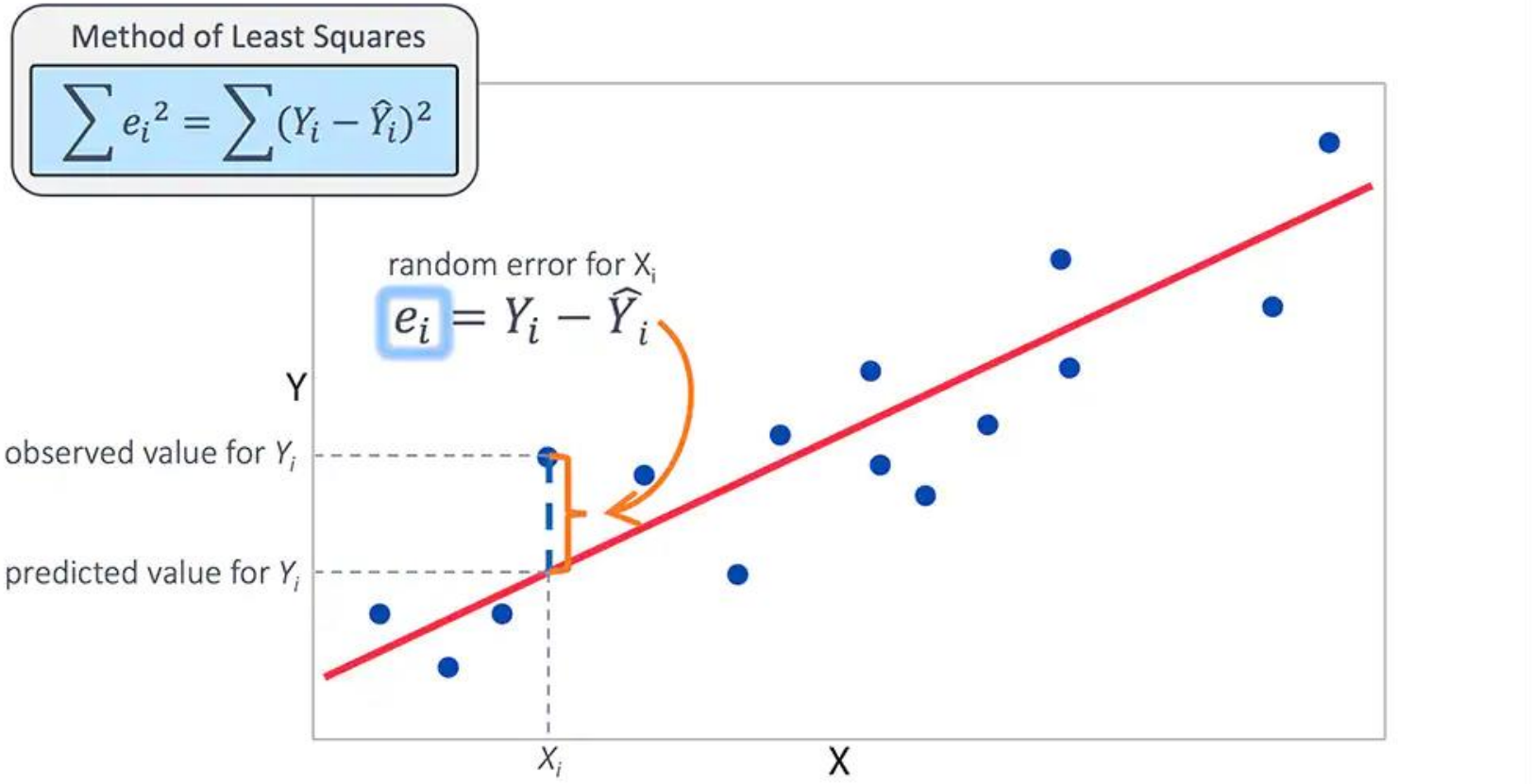
Learned model

Prediction

SOME POPULAR MODELS

- **Linear Models**
- **Support Vector Machines (SVMs)**
- **K-Means**
- **K-Nearest Neighbor (KNN)**
- **Decision Trees**
- **Logistic Regression**
- **Hidden Markov Models (HMM)**
- Naïve Bayes
- Reinforcement Learning
- Neural Networks

LINEAR MODELS



$$y = mx + b$$

slope

height

class

feature

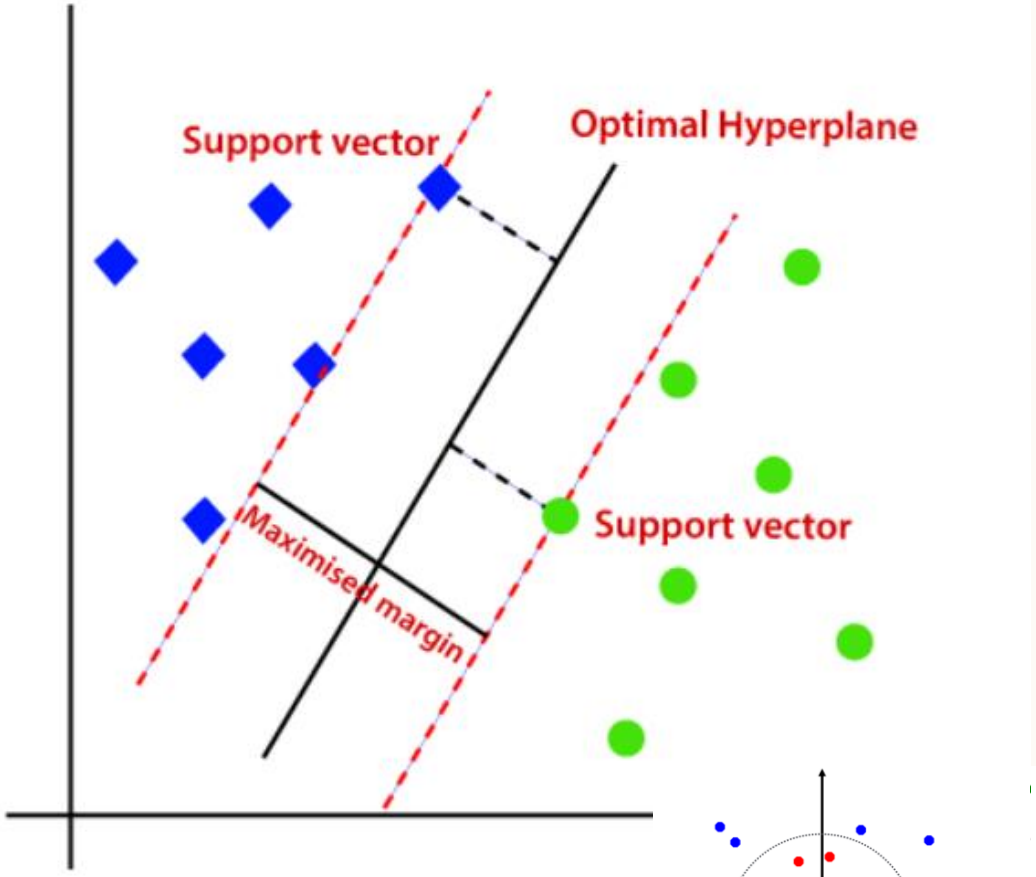
What would be our classes and features?

Is this supervised or unsupervised?

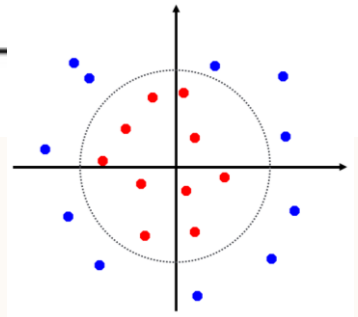
Supervised

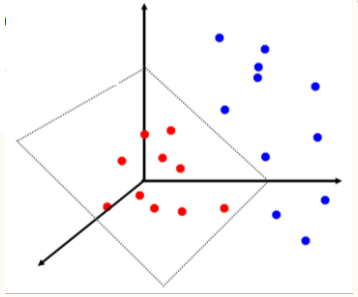
https://www.jmp.com/en_no/statistics-knowledge-portal/what-is-multiple-regression/fitting-multiple-regression-model.html

SUPPORT VECTOR MACHINE (SVM)



- Find the separator with the max margin
- Only support vectors matter → other training examples are ignorable
- Can be non-linear, depending on the "kernel" (Kernel SVM)
  - Choosing the right kernel can be hard





Images from Chris Callison-Burch & https://medium.com/@viveksalunkhe80/support-vector-machine-svm-88f360ff5f38

K-MEANS CLUSTERING

1. Specify how many clusters (k)
2. Cluster the points **x** such that each one has a minimal distance to its mean/centroid of the cluster μ_i

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} ||\mathbf{x} - \mu_i||^2$$

What would be our classes and features?

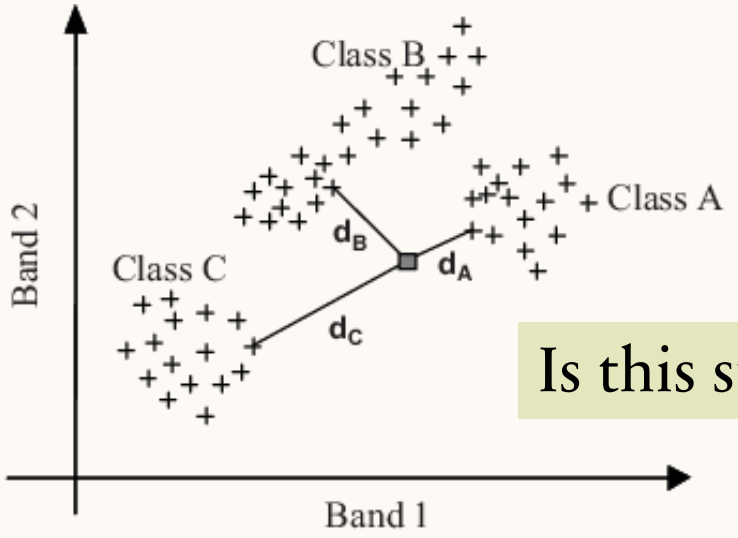
Is this supervised or unsupervised?

Unsupervised

https://datascience.fm/k-means-clustering/

K-NEAREST NEIGHBOR (KNN)

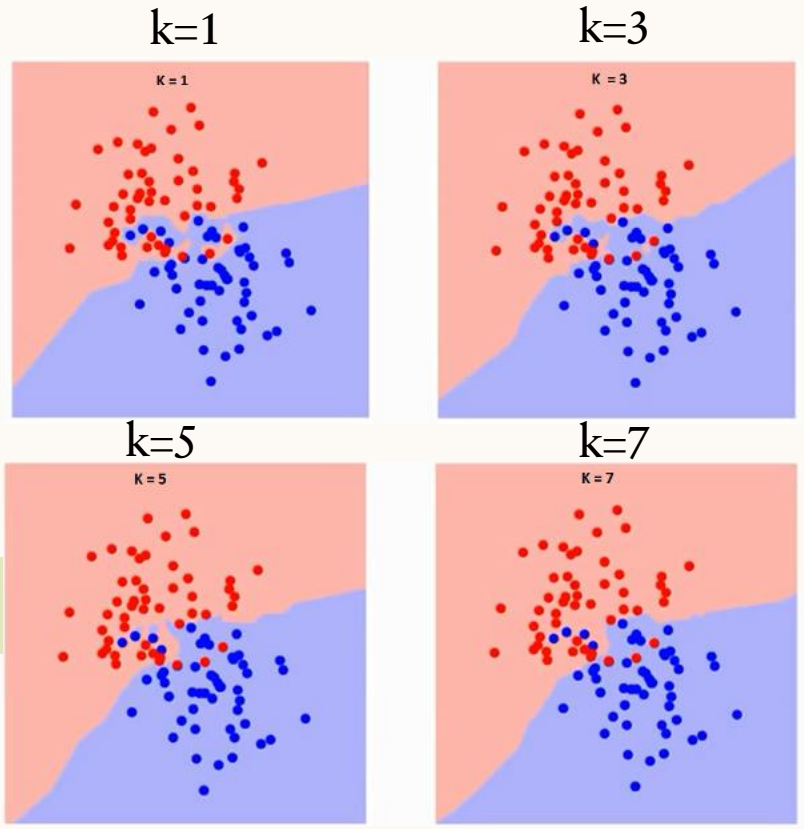
1. Find the distance between the current point to every other point
2. Pick the top k closest points
3. Return the predicted class



What would be our classes and features?

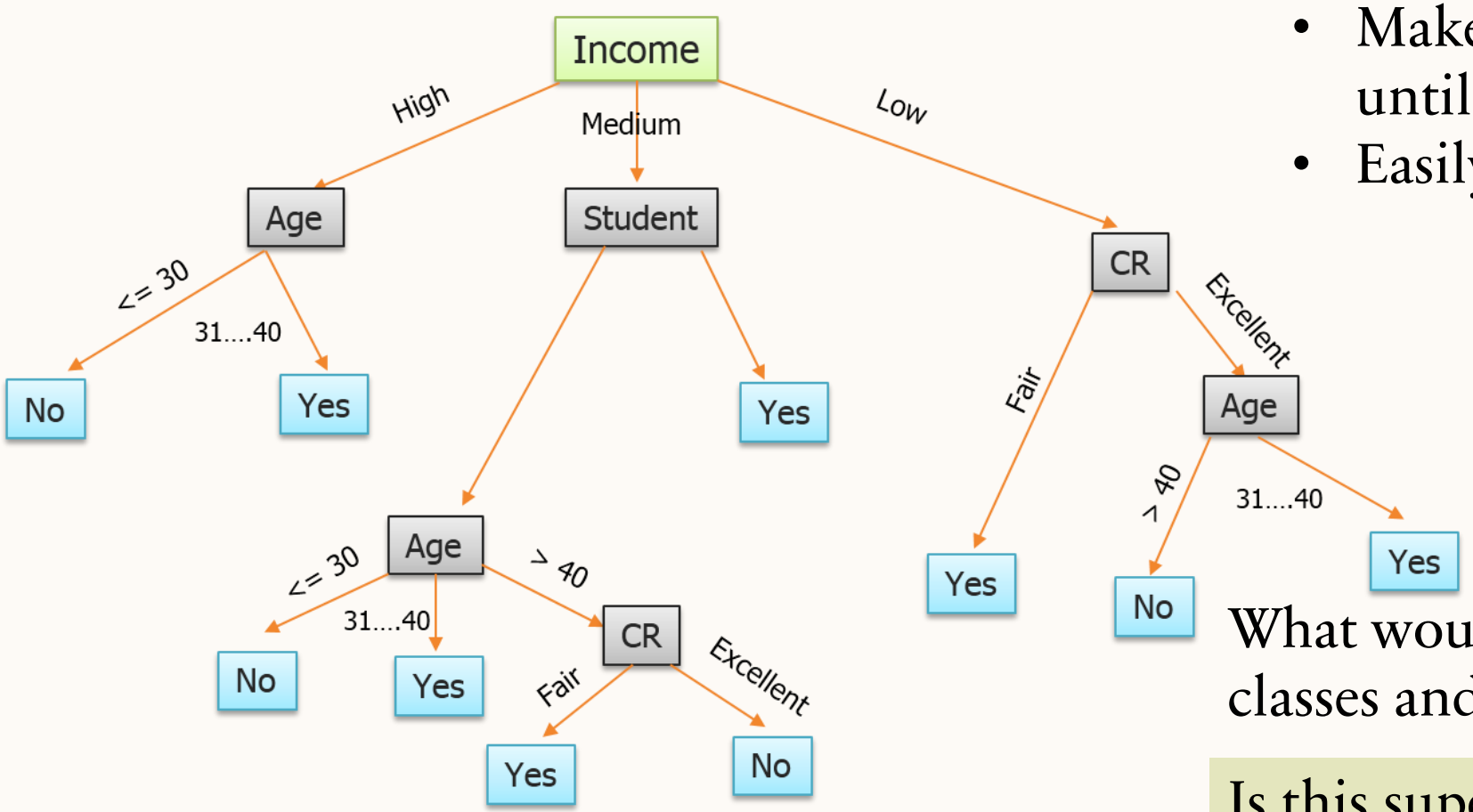
Is this supervised or unsupervised?

Supervised



https://www.researchgate.net/figure/Nearest-Neighbor-Classifier_fig1_221398952
https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

DECISION TREES



- Make a decision at each node until you reach the leaves
- Easily interpretable

What would be our classes and features?

Is this supervised or unsupervised?

Supervised

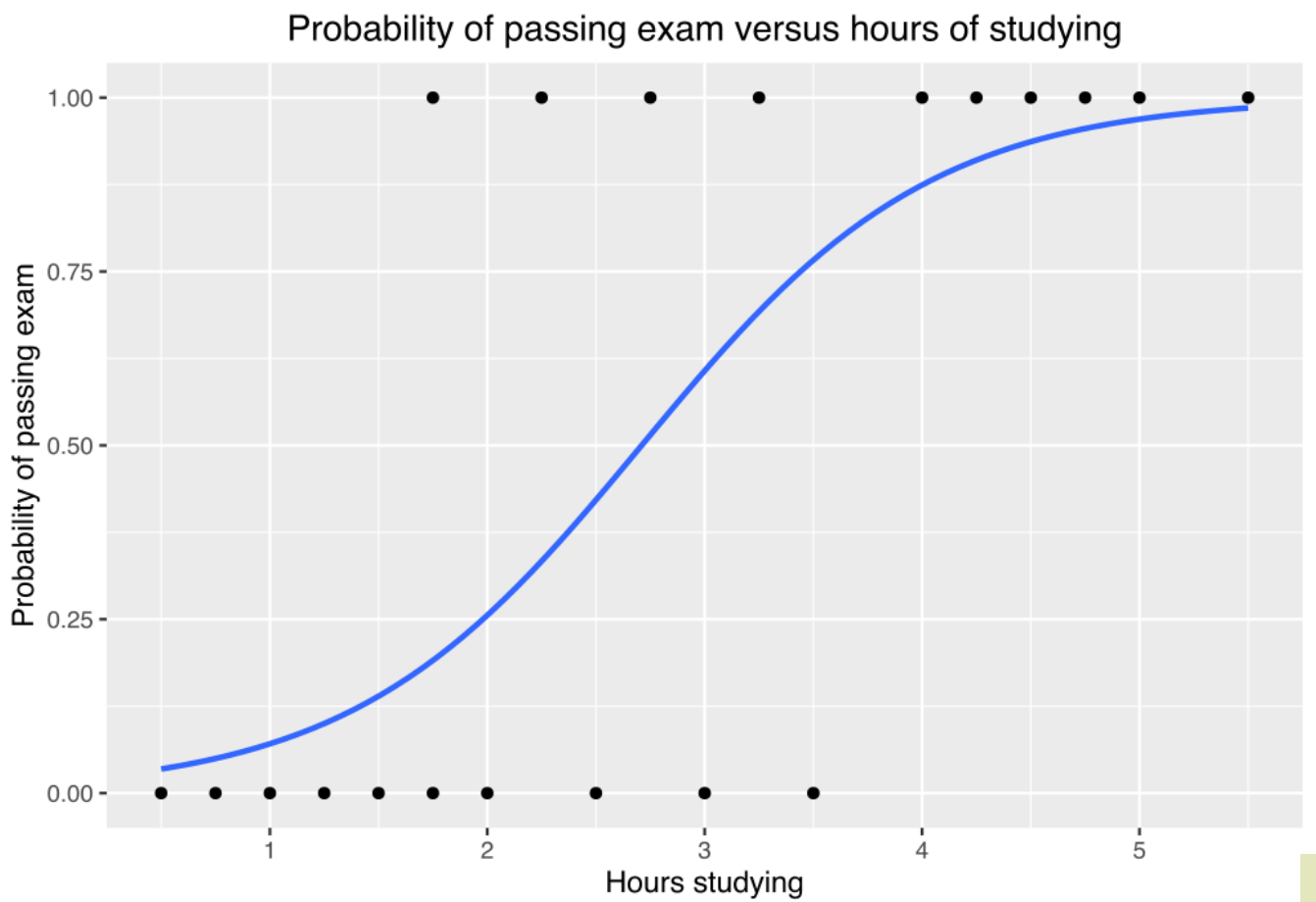
MAKE A DECISION TREE

Datum	Features										Outcome (Label)
	alternatives	bar	Friday	hungry	people	$	rain	reservation	type	wait time	Wait?
X_1	Yes	No	No	Yes	Some	$$$	No	Yes	French	0-10	Yes
X_2	Yes	No	No	Yes	Full	$	No	No	Thai	30-60	No
X_3	No	Yes	No	No	Some	$	No	No	Burger	0-10	Yes
X_4	Yes	No	Yes	Yes	Full	$	Yes	No	Thai	10-30	Yes
X_5	Yes	No	Yes	No	Full	$$$	No	Yes	French	>60	No
X_6	No	Yes	No	Yes	Some	$$	Yes	Yes	Italian	0-10	Yes
X_7	No	Yes	No	No	None	$	Yes	No	Burger	0-10	No
X_8	No	No	No	Yes	Some	$$	Yes	Yes	Thai	0-10	Yes
X_9	No	Yes	Yes	No	Full	$	Yes	No	Burger	>60	No
X_{10}	Yes	Yes	Yes	Yes	Full	$$$	No	Yes	Italian	0-30	No
X_{11}	No	No	No	No	None	$	No	No	Thai	0-10	No
X_{12}	Yes	Yes	Yes	Yes	Full	$	No	No	Burger	30-60	Yes

Hints:

- Try to combine as many data points as possible into a single leaf
- You can use a feature more than once but it's better to make as few decisions as possible

LOGISTIC REGRESSION



$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}$$

μ is the midpoint of the curve

s is the scale parameter (squish)

- Parameters are often estimated using MLE

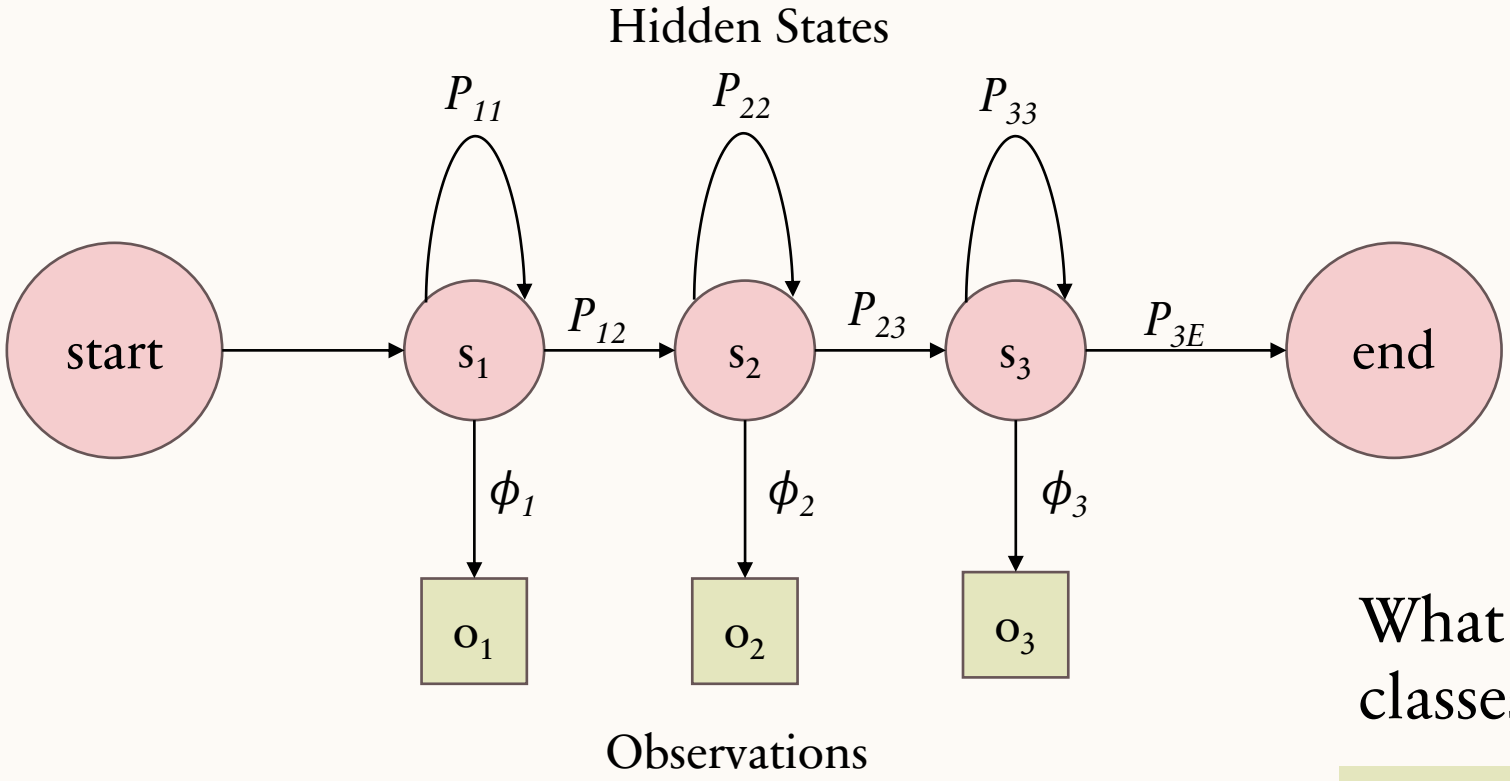
What would be our classes and features?

Is this supervised or unsupervised?

Supervised

https://en.wikipedia.org/wiki/File:Exam_pass_logistic_curve.svg

HIDDEN MARKOV MODELS (HMM)



- Markovian – probabilities to transition from one state to another
- Good for sequences

What would be our classes and features?

Is this supervised or unsupervised?

Semi-supervised

Diagram based off of https://www.sciencedirect.com/topics/mathematics/hidden-markov-model