INTRO TO MACHINE LEARNING

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

11/28/2023

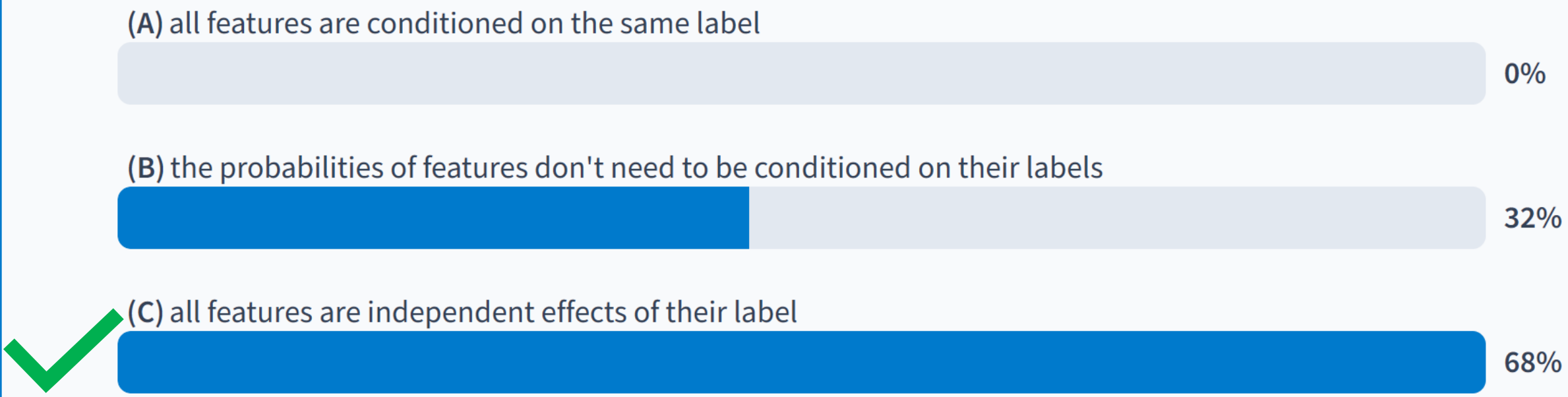
CMSC 671

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

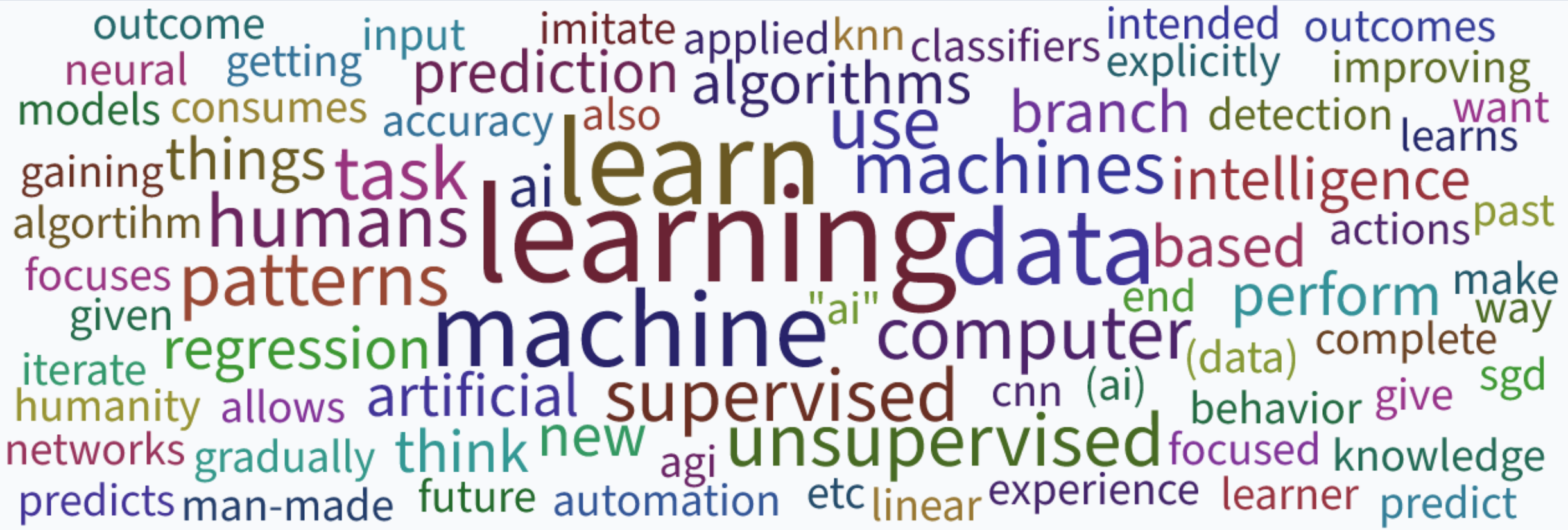
- Distinguish between types of machine learning algorithms
- Identify when you would want to use particular algorithm types

Modified from slides by Dr. Chris Callison-Burch, Dr. Cassandra Kent, & Dr. Cynthia Matuszek

Naive Bayes is naive because it assumes...



What do you think of when you hear "machine learning"?



COURSE UPDATES

- Course Evaluations are available until December 12 at 11:59pm
- Milestone 2 is due tonight (11:59pm)
  - Take feedback into consideration!!
- Milestone 3 is due December 7th (next Thursday)
  - Draft of paper, official conference template
  - Get preliminary results

EXTRA CREDIT REPORT

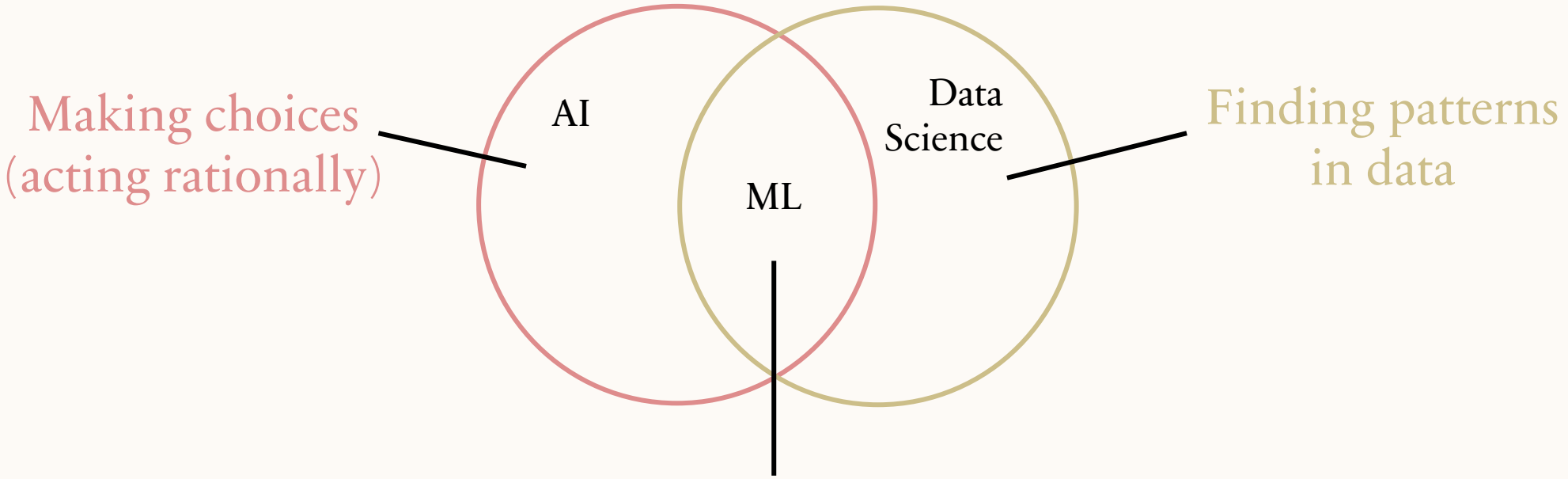
- Goal: use research to back up your stance on something from the news about AI

- 0.5-2 points on final grade
  - 0.5 if you give a solid effort but don't satisfy other requirements
  - 2 points if you satisfy all the requirements

https://laramartin.net/Principles-of-AI//homeworks/extra-credit-report.html

MACHINE LEARNING

WHAT IS MACHINE LEARNING?



Collecting data -> Finding patterns -> Selecting actions

The goal: learning **models**

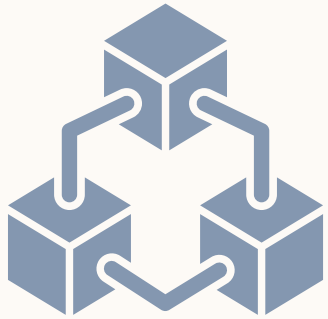
WHAT IS A MODEL?

- Model: simplification of some phenomenon
  - Learning patterns from evidence

- Typically model something we don't know or have trouble specifying
  - How the environment can change
  - Actual states given only (partial) observations
  - How my sensors work
  - How my actions work
  - …

HOW DO WE LEARN MODELS?

Take past experiences
(lots of data; corpus)



Find patterns
(the ML algorithm)

Use on new experiences
(save & test the model)

TYPES OF MODELS

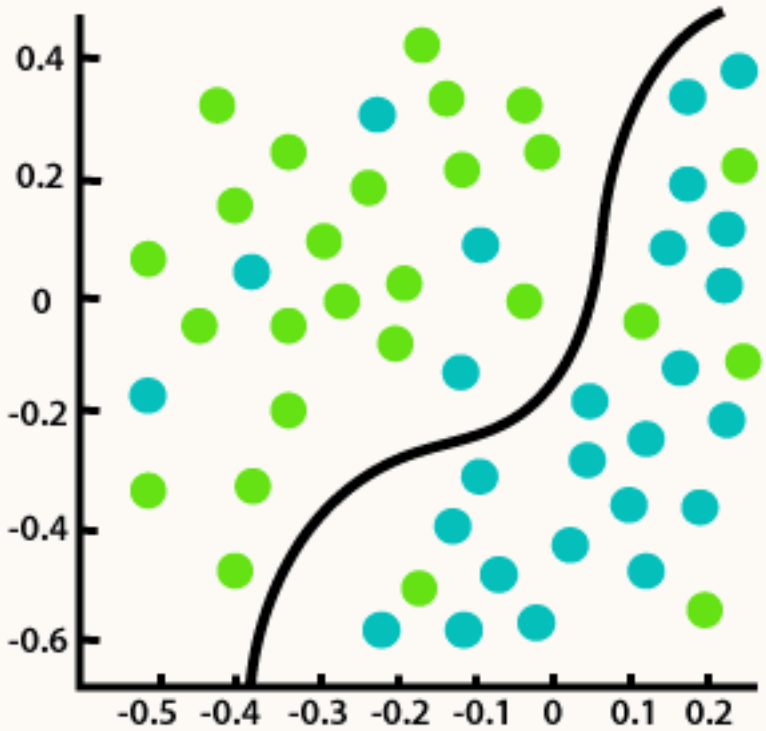
Classification

- Model outputs comes from a finite set of values
- Discrete result
- *Examples*:
  - What type of animal is this a picture of?
  - Predicting the weather (sunny, cloudy, or rainy?)
  - Ranking: Is this result *better* than this result?

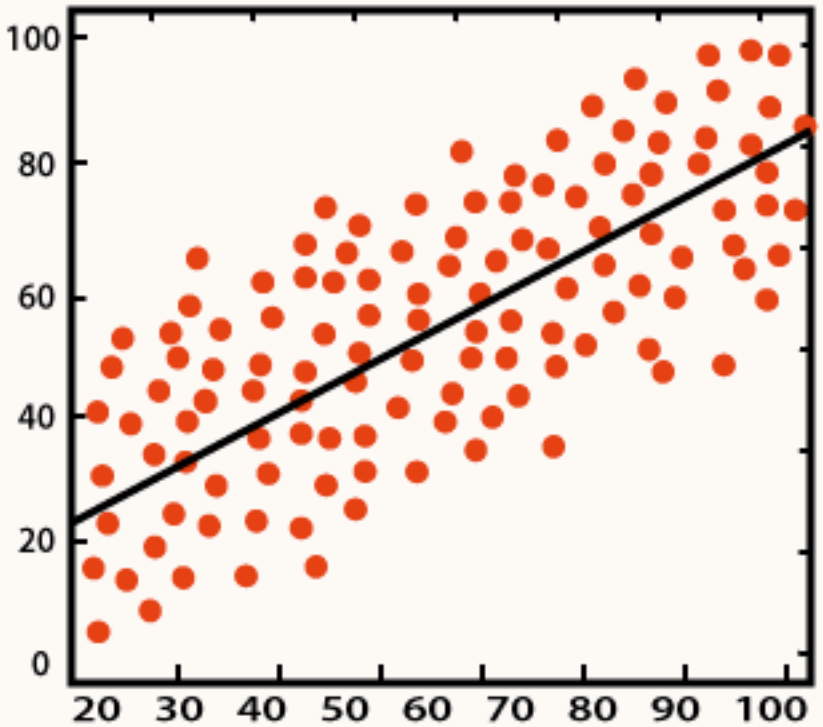
Regression

- Model outputs are continuous values
- Continuous result
- *Examples*:
  - How far will I move if I drive my motors at this speed for 1 second?
  - Predicting the weather (temperature)
  - Ranking: *how good* is this result?

TYPES OF MODELS



Classification



Regression

https://medium.com/unpackai/classification-regression-in-machine-learning-7cf3b13b0b09

WHAT ARE SOME OTHER EXAMPLES OF THESE?

Classification

- Spam vs not spam
- Is a food poisonous
- Picture identification/labeling

Regression

- House prices
- Finding relationship between two features
- Weather forcasting

TYPES OF LEARNING

Supervised learning

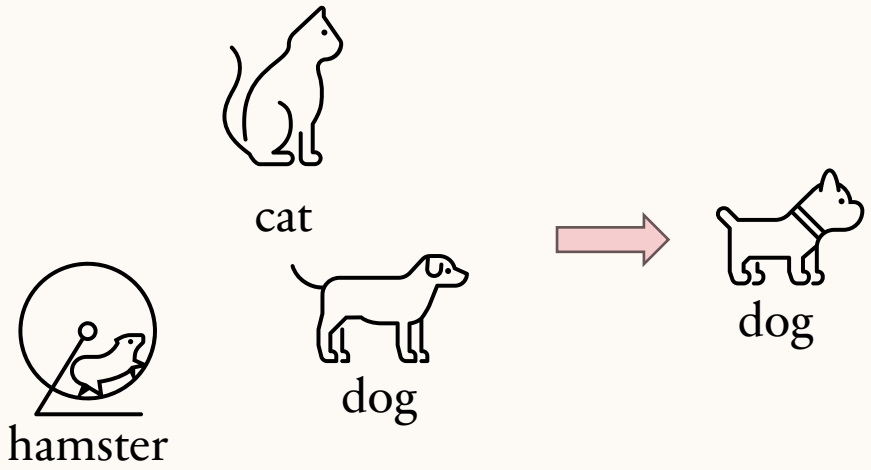
- Data has feedback (labels)
- Data consists of input-output pairs
- Learn mapping from input to output
- Examples:
  - Dataset classification
  - How likely is it that this person will get into a car accident?

Unsupervised learning

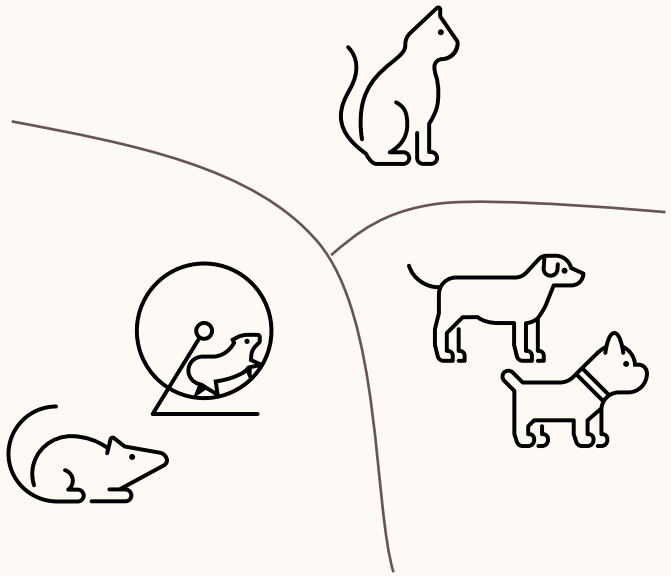
- No explicit feedback in data
- Learn patterns directly from data
- *Examples*:
  - Clustering
  - Do these people fall under multiple groups?

TYPES OF LEARNING

Supervised learning



Unsupervised learning



WHAT ARE SOME OTHER EXAMPLES OF THESE?

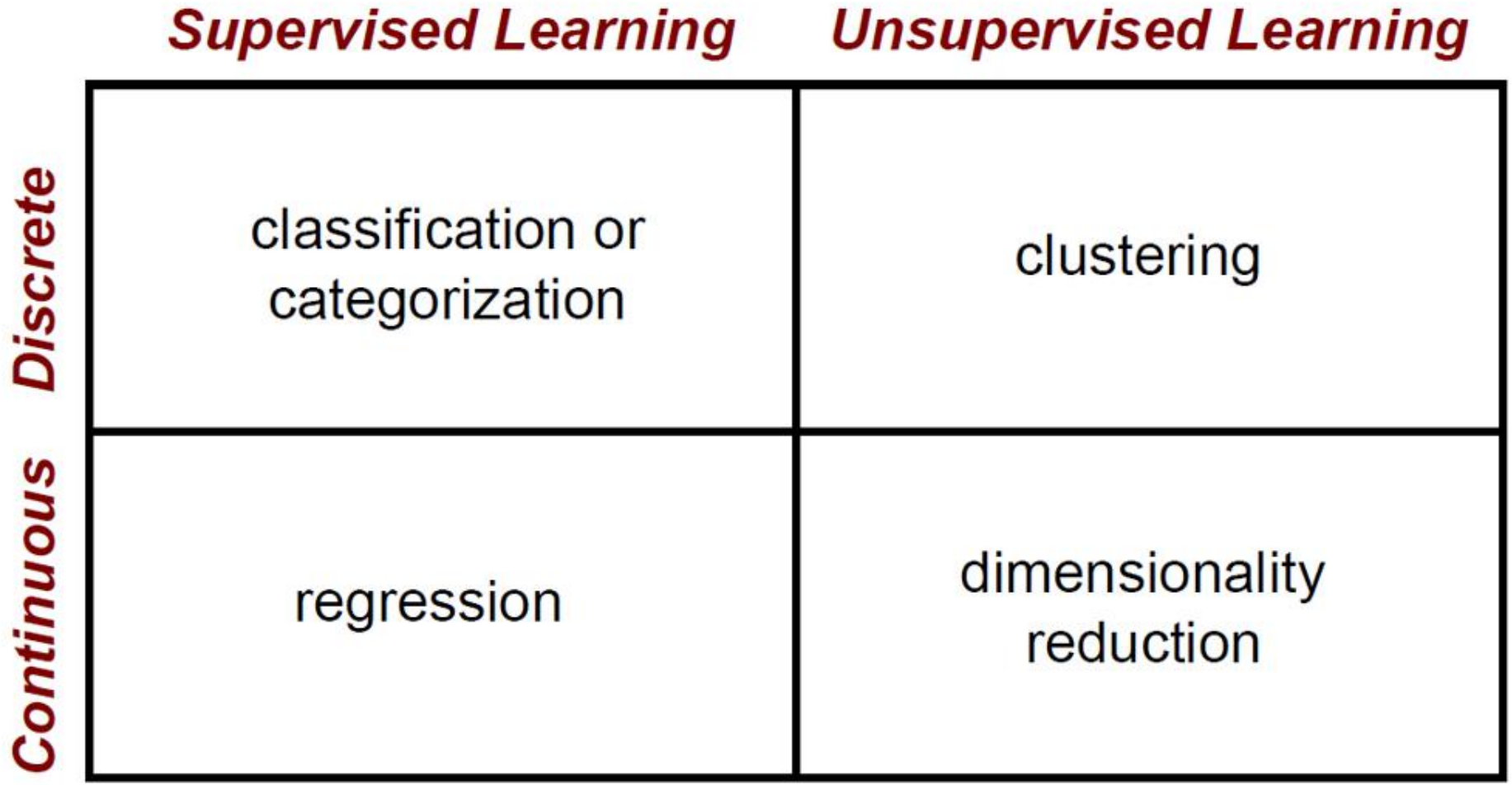
Supervised learning

- Spam vs not-spam (if labeled)
- Disease detection
- Loan approval -- prediction

Unsupervised learning

- Text clustering (topic modeling)
- Fraud detection (via outliers)
- Similarities across domain (e.g., clusters of grades)
- Suggesting similar images

TYPES OF ALGORITHMS



TYPES OF LEARNING

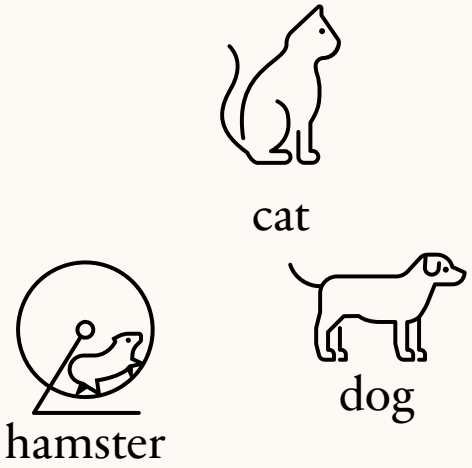
Offline learning

- We have data in advance
- Predictions do not change over time
- Data is independent and identically distributed
- *Examples*:
  - Database learning tasks
  - Image captioning
  - Machine translation

Online learning

- Model receives input, produces output, and *then* is told what the correct output was
- Assume predictions may change over time
- Data is related over time
- *Examples*:
  - Long-term applications
  - Personalization
  - Field robotics

HOW DO WE LEARN MODELS? (2)



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

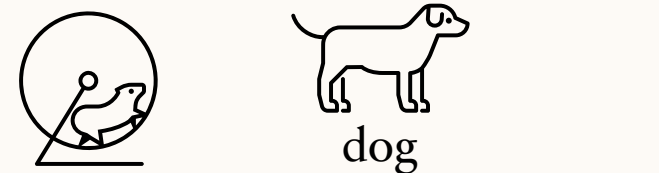
P(hamster|[line in this position],…)

P(dog|[line in this other position],…)

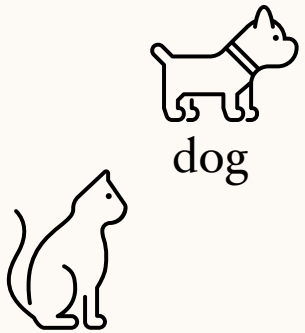
Have data with features extracted (and possibly labels)

Learn associations between features and labels

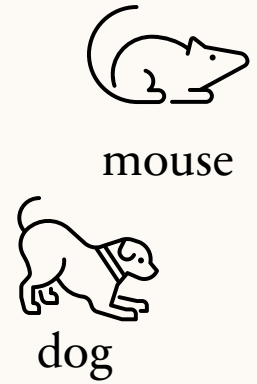
DIVIDING UP DATA FOR TRAINING



Training
~80%



Validation/Development
~10%



Testing
~10%

Why would we do this? To prevent overfitting!