Lecture 23: Artificial intelligence, machine learning, natural language processing, ...

- buzzwords, hype, real accomplishments, wishful thinking
  - big data, deep learning, neural networks, ...
- brief history
- examples
  - classification (spam detection)
  - prediction (future prices)
  - recommendation systems (Netflix, Amazon, Goodreads, ...)
  - games (chess, Go)
  - natural language processing (sentiment analysis, translation, generation)
- issues and concerns
  - fairness, bias
  - accountability and explainability
  - appropriate uses

- Beware: on this topic, I am even less of an expert than normal.

Revisionist history (non-expert perspective)

- 1950s, 1960s: naive optimism about artificial intelligence
  - checkers, chess, machine translation, theorem proving, speech recognition, image recognition, vision, ...
  - almost everything proved to be much harder than was thought
- 1980s, 1990s: expert or rule-based systems
  - domain experts createrules, computers apply them to make decisions
  - it's too hard to collect the rules, and there are too many exceptions
  - doesn't scale to large datasets or new problem domains
- 2010s: machine learning, big data, deep learning, ...
  - provide a "training set" with lots of examples correctly characterized
  - define "features" that might be relevant, or let program find them itself
  - write a program that "learns" from its successes and failures on the training data (basically by figuring out how to combine feature values)
  - turn it loose on new data
- 2020s: generative language models ???
  - near-human performance on some text understanding and generation tasks
  - GPT-3, DALL-E2, ...

Examples of ML applications (a tiny subset)

- classification
  - spam detection, digit recognition, optical character recognition, authorship, ...
  - image recognition, face recognition, ...
- prediction
  - house prices, stock prices, credit scoring, resume screening, ...
  - tumor probabilities, intensive care outcomes, ...
- recommendation systems
  - e.g., Netflix, Amazon, Goodreads, ...
- natural language processing (NLP)
  - language translation
  - text to speech; speech to text
  - sentiment analysis
  - text generation
- games
  - checkers, chess, Go

Types of learning algorithms

- supervised learning (labeled data)
  - teach the computer how to do something with training examples
  - then let it use its new-found knowledge to do it on new examples

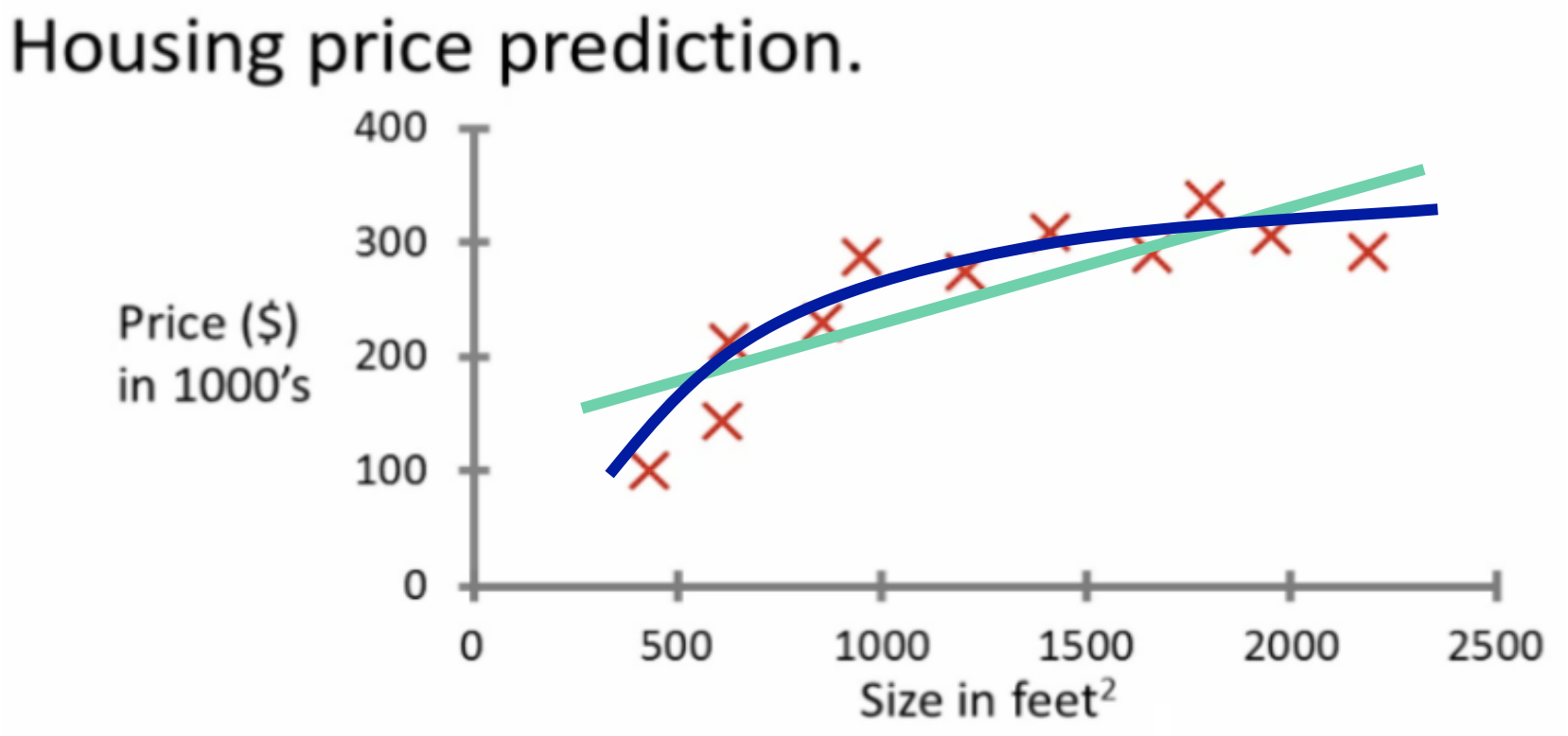
- unsupervised learning (unlabeled data)
  - let the computer learn how to do something without training data
  - use this to find structure and patterns in data

- reinforcement learning
  - some kind of "real world" system to interact with
  - feedback on success or failure guides/teaches future behavior

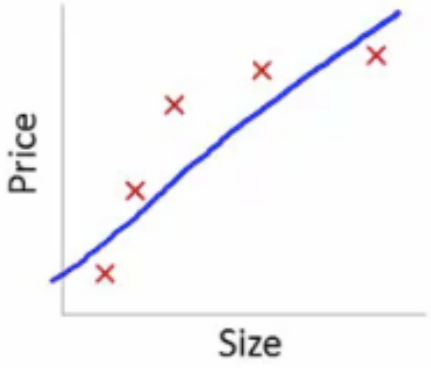
- recommender systems
  - look for similarities in likes and dislikes / behaviors / ...
  - use that to predict future likes / behaviors

Prediction example: house prices

- only one feature here: square footage
- straight line? ("linear regression")
- some kind of curve?

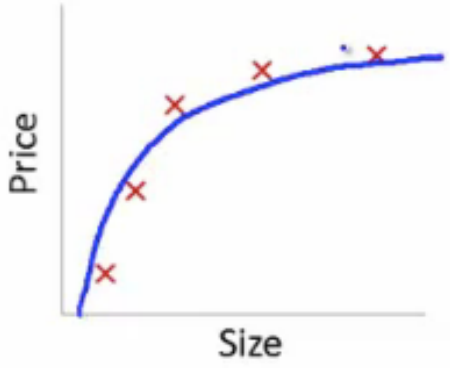


Over- and under-fitting



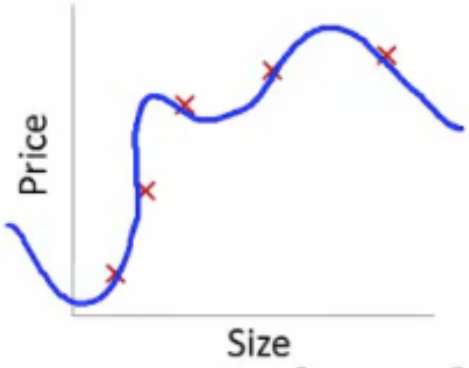
$\theta_0 + \theta_1 x$

High bias
(underfit)



$\theta_0 + \theta_1 x + \theta_2 x^2$

"Just right"

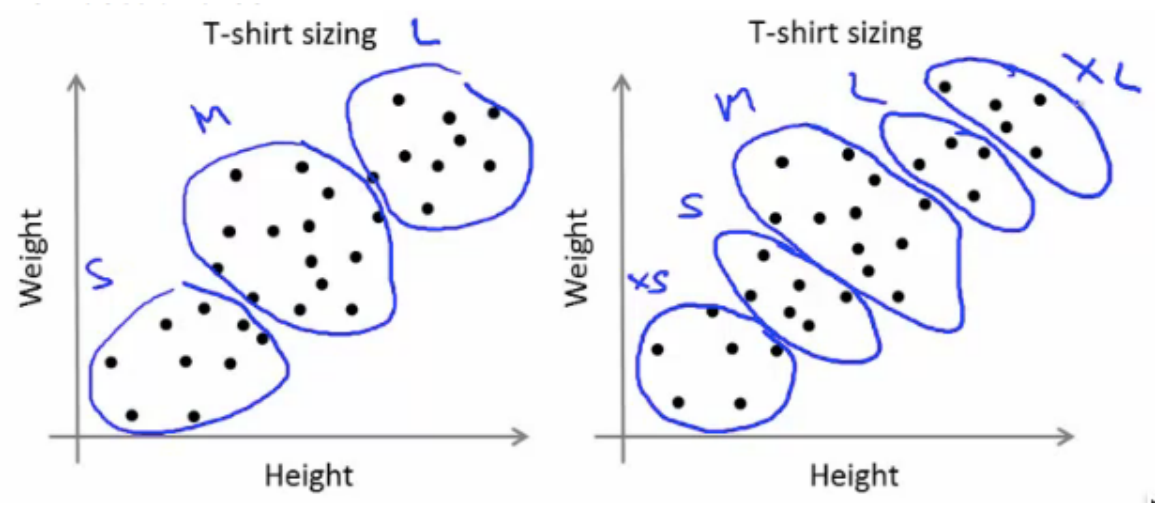


$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High variance
(overfit)

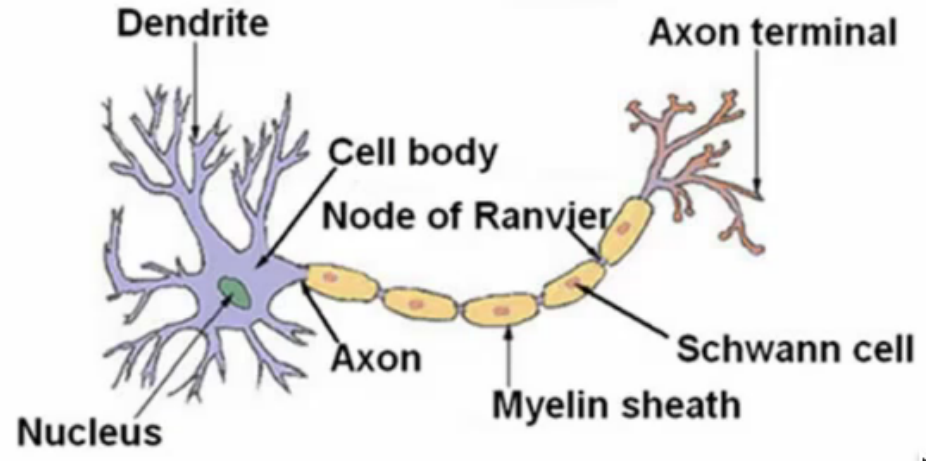
Clustering: learning from unlabeled data

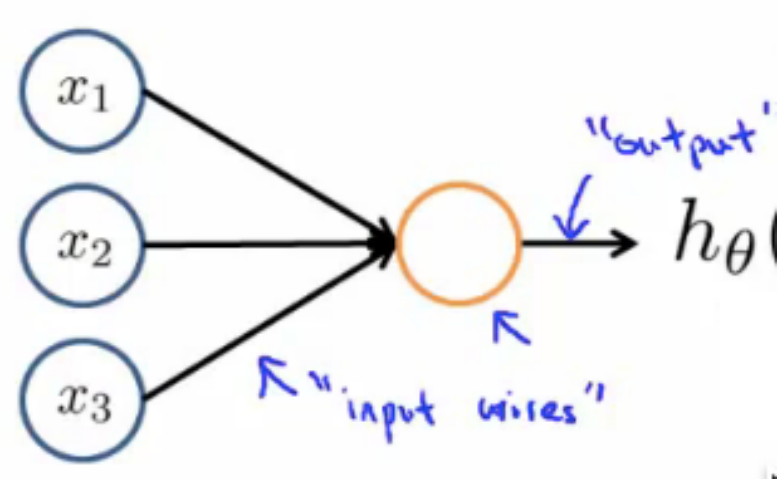
- contrast with supervised learning
  - supervised learning:
    given a set of labels, fit a hypothesis to it
  - unsupervised learning:
    try and determine structure in the data
    clustering algorithm groups data together based on data features
- clustering is good for
  - market segmentation – group customers into different market segments
  - social network analysis – identify friend groups
  - topic analysis
  - authorship

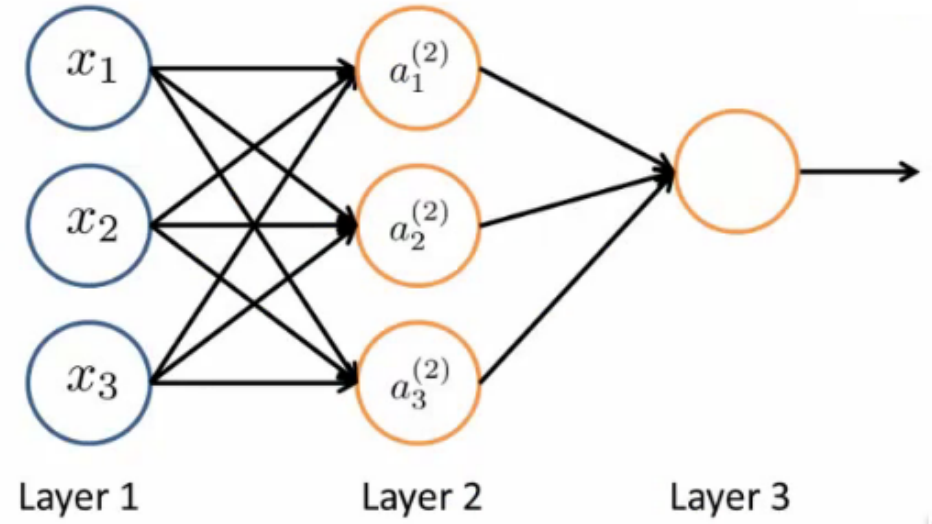


Neural networks, deep learning

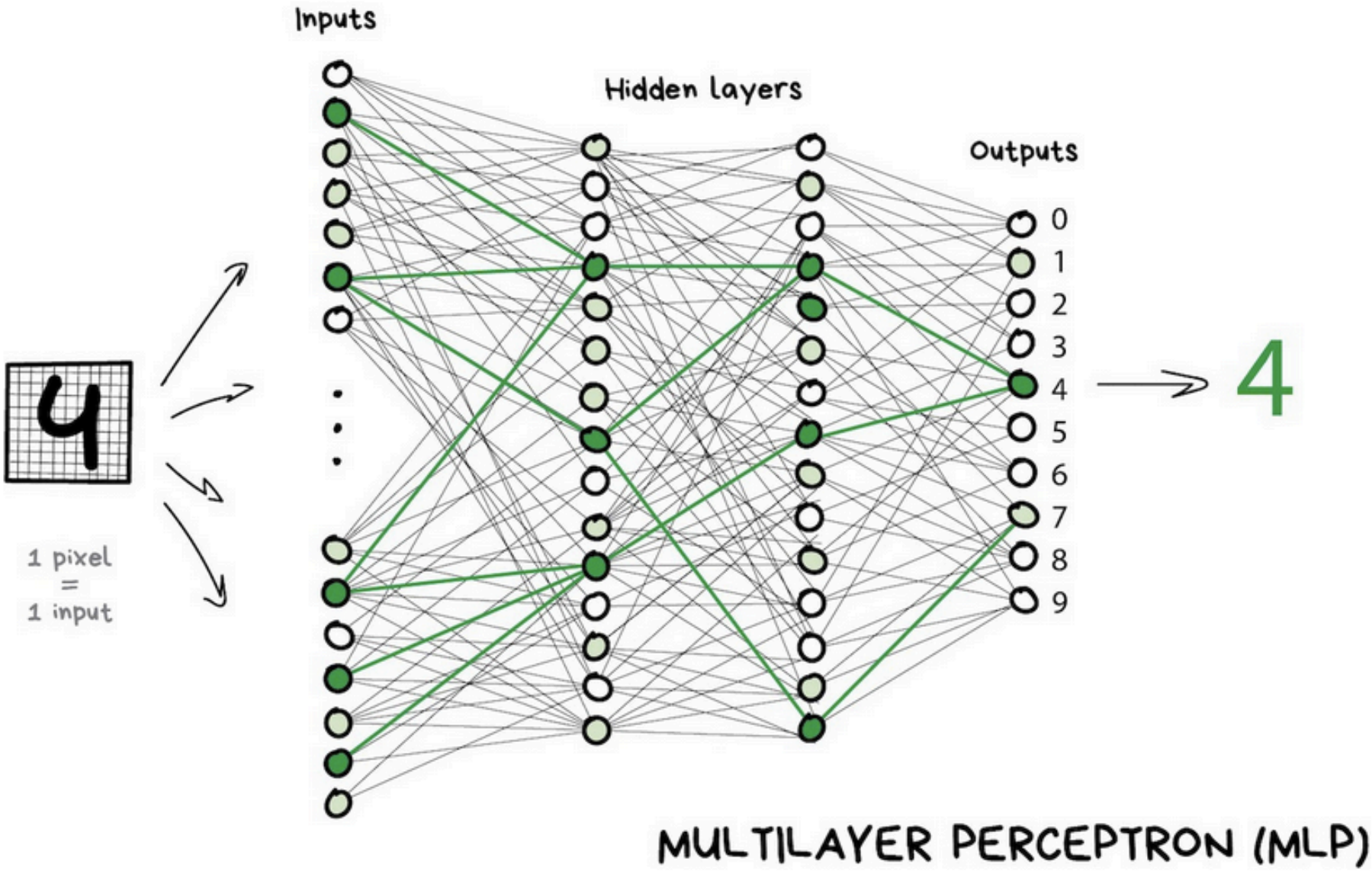
- simulate human brain structure
  with artificial neurons
  in simple connection patterns







Neural networks (from vas3k.com/blog/machine_learning)



GPT-3, ChatGPT: generative pre-trained transformers

- language models based on very large text corpus
- use deep learning to generate text that seems human-written
- models are proprietary
  - e.g., GPT-3 licensed by Microsoft

- ChatGPT is based on GPT-3 (chat.openai.com)

- tuned for conversational style
- can remember previous parts of conversaion

- very new: became available ~Dec 1

ML / AI issues

- algorithmic fairness
  - results can't be better than training data
  - if that has implicit or explicit biases, results are biased
  - can we detect and eliminate bias?
- accountability and explainability
  - what is the algorithm really doing?
  - can its results be explained
- appropriate uses?
  - prison sentencing
  - drone strikes
  - weapon systems
  - resume evaluation
  - medical decisions
  - ...
- to learn more:

  https://fairmlbook.org

More AI/ML issues

- what if it gets too good at faking humans?
  - deep fakes
  - text generation
  - generating problem set solutions

- training data is likely to contain bias, toxic language, stereotypes (e.g., gender, race, ...), and other potentially harmful material