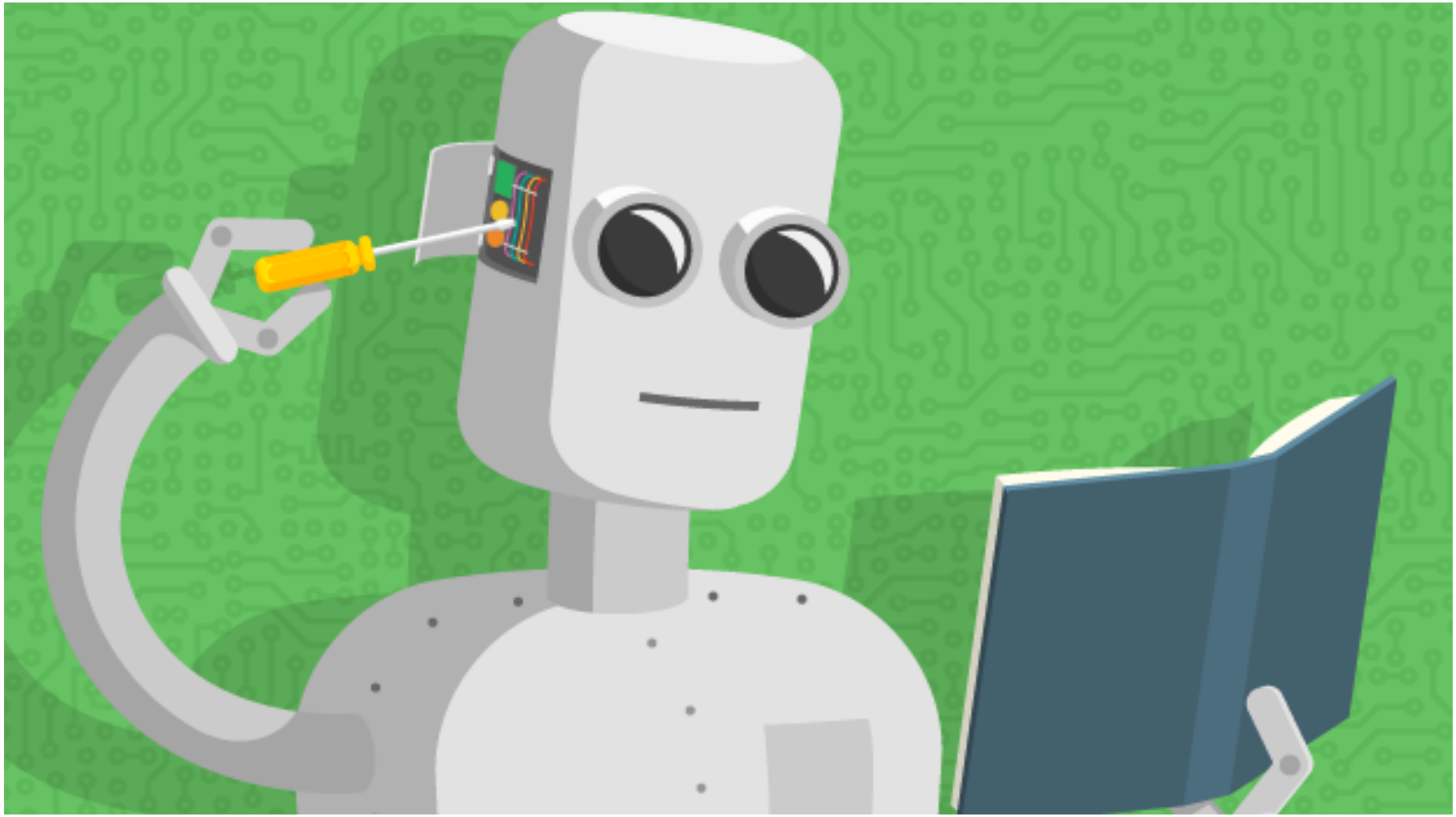
Machine Learning overview

Chapter 18, 21



What is learning?

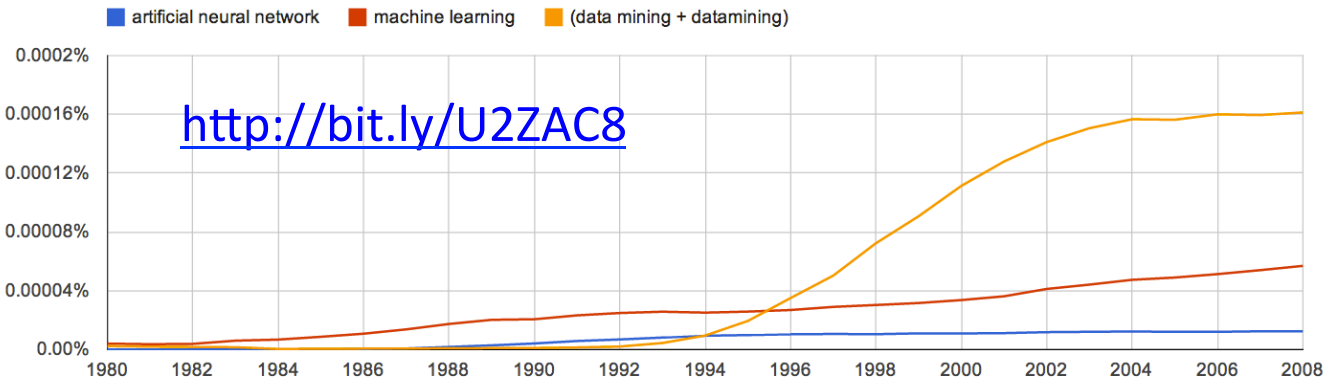
- "Learning denotes changes in a system that ... enable a system to do the same task more efficiently the next time" – Herbert Simon
- "Learning is constructing or modifying representations of what is being experienced" – Ryszard Michalski
- "Learning is making useful changes in our minds" – Marvin Minsky

Why study learning?

- Understand and improve efficiency of **human learning**
  - Use to improve methods for teaching and tutoring people (e.g., better computer-aided instruction)
- **Discover** new things or structure previously unknown
  - Examples: data mining, scientific discovery
- Fill in skeletal or **incomplete specifications in** a domain
  - Large, complex systems can't be completely built by hand & require dynamic updating to incorporate new information
  - Learning new characteristics expands the domain or expertise and lessens the "brittleness" of the system
- Build agents that can **adapt** to users, other agents, and their environment

AI & Learning Today

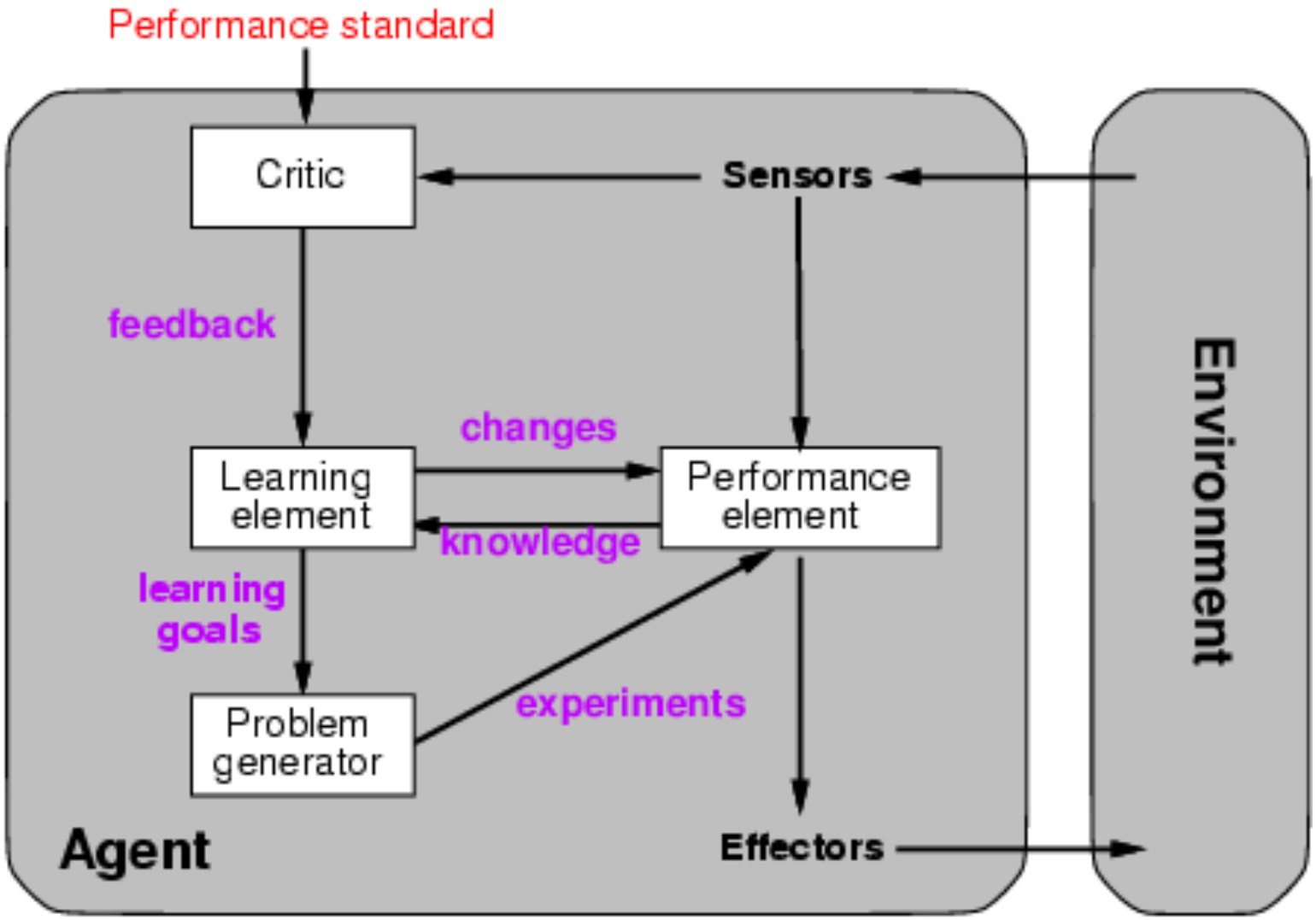
- Neural network learning was popular in the 60s
- In the 70s and 80s it was replaced with a paradigm based on manually encoding and using knowledge
- In the 90s, more data and the Web drove interest in new statistical machine learning (ML) techniques and new data mining applications
- Today, ML techniques and big data are behind almost all successful intelligent systems



Machine Leaning Successes

- Sentiment analysis
- Spam detection
- Machine translation
- Spoken language understanding
- Named entity detection
- Self driving cars
- Motion recognition (Microsoft X-Box)
- Identifying paces in digital images
- Recommender systems (Netflix, Amazon)
- Credit card fraud detection

A general model of learning agents



Major paradigms of machine learning

- **Rote learning**: 1-1 mapping from inputs to stored representation, learning by memorization, association-based storage & retrieval
- **Induction:** Use specific examples to reach general conclusions
- **Clustering**: Unsupervised discovery of natural groups in data
- **Analogy:** Find correspondence between different representations
- **Discovery**: Unsupervised, specific goal not given
- **Genetic algorithms:** *Evolutionary* search techniques, based on an analogy to *survival of the fittest*
- **Reinforcement** – Feedback (positive or negative reward) given at the end of a sequence of steps

What we will and won't cover

- We'll look at a few popular machine learning problems and algorithms
  - Take CMSC 478/678 Machine Leaning for more
  - Use online resources & experiment on your own
- We'll focus on when/how to use techniques and only touch on how/why they work
- We'll cover basic methodology and evaluation
- We'll use Weka platform for examples & demos
  - Great for exploration and learning

Types of learning problems

- **Supervised**: learn from training examples
  - Regression
  - Classification: Decision Trees, SVM
- **Unsupervised**: learn w/o training examples
  - Clustering
  - Dimensionality reduction
- Lots more we won't cover
  - Hidden Marrkov models
  - Learning to rank
  - Semi-supervised learning
  - Reinforcement learning
  - Active learning

Machine Learning Problems

	Supervised Learning	*Unsupervised Learning*
Discrete	classification or categorization	clustering
Continuous	regression	dimensionality reduction

Supervised learning

- Given training examples of inputs & corresponding outputs, produce "correct" outputs for new inputs
- Two main scenarios:
  - **Classification:** outputs typically labels (goodRisk, badRisk); learn a decision boundary that separates classes
  - **Regression:** aka "curve fitting" or "function approximation." Learn a continuous input-output mapping from (possibly noisy) examples

Unsupervised Learning

Given only *unlabeled* data as input, learn some sort of structure, e.g.:

- Cluster your Facebook friends based on similarity of posts and friends
- Find sets of words whose meanings are related (e.g., doctor, hospital)
- Induce N topics and the words that are common in documents that are about that dopic

Weka: Waikato Environment for Knowledge Analysis

Open source Java software for ML and datamining

http://cs.waikato.ac.nz/ml/weka/