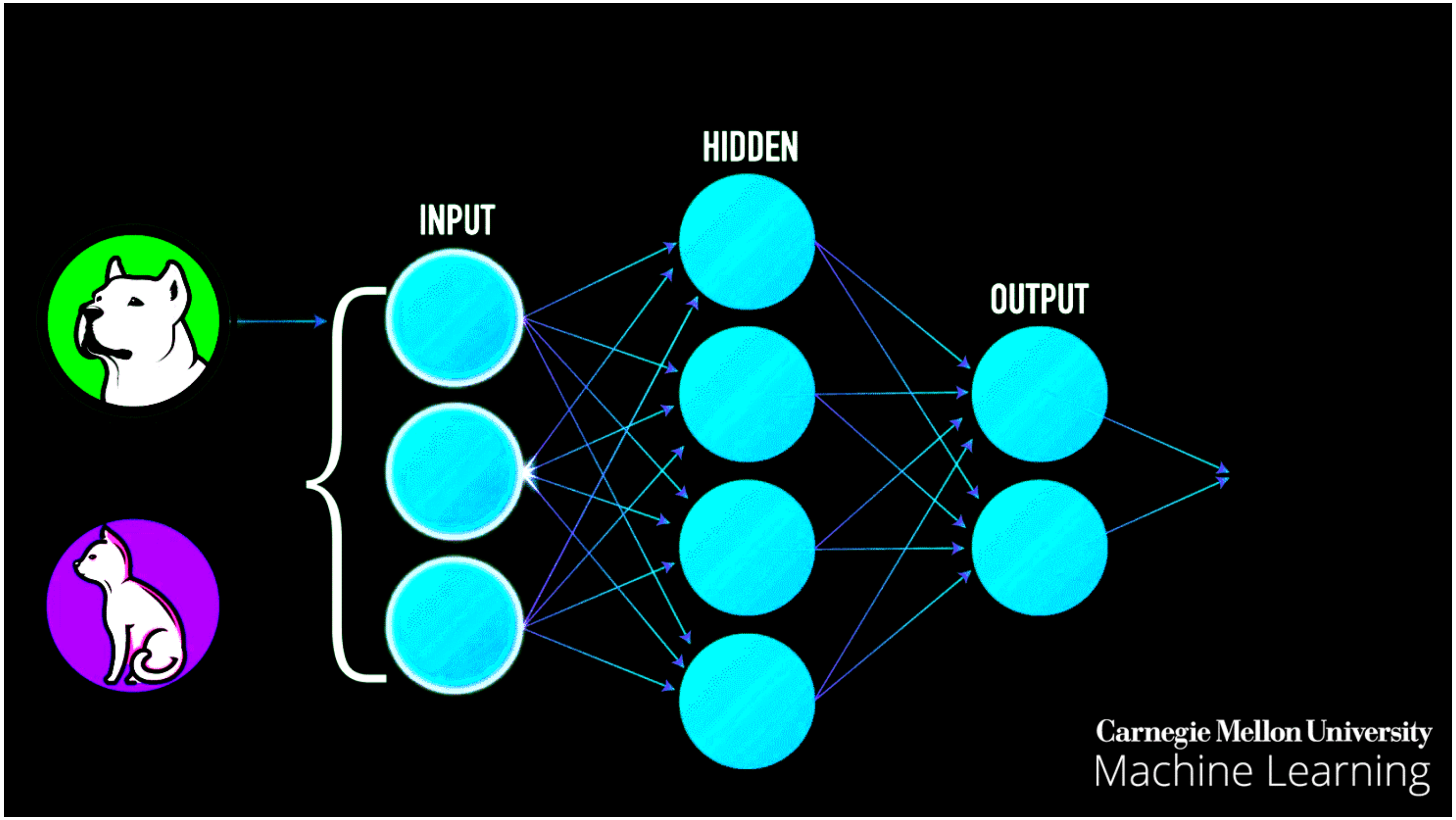
Lecture #1

Machine Learning Algorithms (INT417)

What is Learning?

What is Machine Learning?

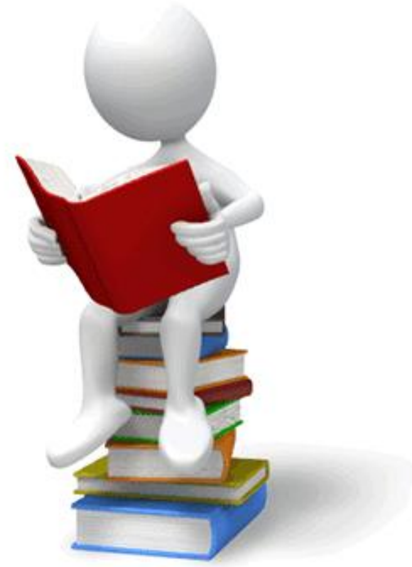


What Is Learning?

- **Rat Bait Shyness**
  - *Rats Learning to Avoid Poisonous Baits*

learning by memorization

- Suppose we would like to program a machine that learns how to filter spam e-mails.
  - The machine will simply *memorize* all previous e-mails that had been labeled as spam e-mails by the human user. When a new e-mail arrives, the machine will search for it in the set of previous spam e-mails. If it matches one of them, it will be trashed. Otherwise, it will be moved to the user's inbox folder
  - it lacks an important aspect of learning systems
    - the ability to label unseen e-mail messages.



Generalization OR Inductive Reasoning

- recognising similarity between different situations.
- bait shyness example presented previously, after the rats encounter an example of a certain type of food, they apply their attitude toward it on new, unseen examples of food of similar smell and taste.
- spam filtering task, the learner can scan the previously seen e-mails, and extract a set of words whose appearance in an e-mail message is indicative of spam. Then, when a new e-mail arrives, the machine can check whether one of the suspicious words appears in it, and predict its label accordingly.
- However, inductive reasoning might lead us to false conclusions.
- Example: “offer” keyword in mail -- spam

Inductive Bias

- The incorporation of ***prior knowledge*** that biases the learning mechanism is referred to as ***inductive bias*.**
- For example mail from my organization with any keyword.
- The stronger the prior knowledge (or prior assumptions) that one starts the learning process with, the easier it is to learn from further examples.
- However, the stronger these prior assumptions are, the less flexible the learning is – it is bound, a priori, by the commitment to these assumptions.

When Do We Need Machine Learning?

- Two aspects that learn and improve on the basis of their “experience”: the problem’s complexity and the need for adaptivity.
- **Tasks That Are Too Complex to Program.**
  - *Tasks Performed by Animals/Humans:*
    - driving, speech recognition, and image understanding.
    - “learn from their experience”
  - *Tasks beyond Human Capabilities:*
    - analysis of very large and complex data sets: astronomical data, turning medical archives into medical knowledge, weather prediction, analysis of genomic data, Web search engines, and electronic commerce.
- **Adaptivity.**
  - programs that decode handwritten text, where a fixed program can adapt to variations between the handwriting of different users; spam detection programs, adapting automatically to changes in the nature of spam e-mails; and speech recognition programs.

Types of Learning

- Supervised, Unsupervised and Reinforcement
- Active vs. Passive Learners
- Online vs. Batch Learning Protocol
- Eager Learner vs. Weak Learner

Active versus Passive Learners

- An active learner interacts with the environment at training time, say, by posing queries or performing experiments,
- while a passive learner only observes the information provided by the environment (or the teacher) without influencing or directing it.
- spam filter is usually passive

Online versus Batch Learning Protocol

- situations in which the learner has to respond online, throughout the learning process.
- settings in which the learner has to engage the acquired expertise only after having a chance to process large amounts of data.
- a stockbroker has to make daily decisions, based on the experience collected so far. He may become an expert over time, but might have made costly mistakes in the process. In contrast, in many data mining settings, the learner the data miner has large amounts of training data to play with before having to output conclusions.

Weak Learner Vs Eager Learner

– time of convergence

– Week Learner: KNN

– Fast Learner: SVM

Some successful applications of machine learning.

- **Learning to recognize spoken words.**
  - **All** of the most successful speech recognition systems employ machine learning in some form.
    - For example, the SPHINX system (e.g., Lee 1989) learns speaker-specific strategies for recognizing the primitive sounds (phonemes) and words from the observed speech signal.
    - Neural network learning methods (e.g., Waibel et al. 1989) and methods for learning hidden Markov models(e.g., Lee 1989) are effective for automatically customizing to, individual speakers, vocabularies, microphone characteristics, background noise, etc.
- **Learning to drive an autonomous vehicle.**
  - Machine learning methods have been used to train computer-controlled vehicles to steer correctly when driving on a variety of road types.
    - For example, the **ALVINN** system (Pomerleau 1989) has used its learned strategies to drive unassisted at 70 miles per hour for 90 miles on public highways among other cars.

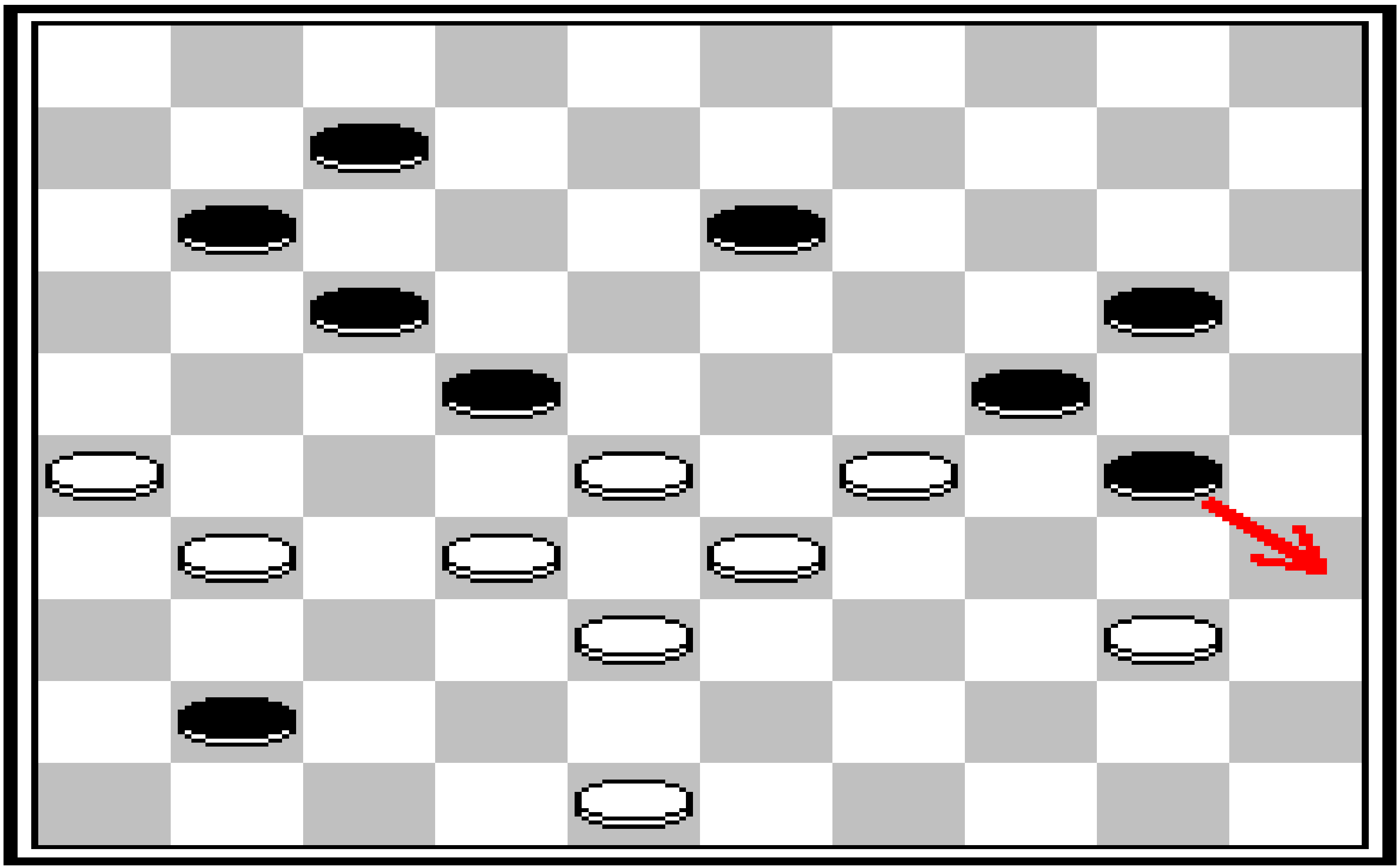
Some successful applications of machine learning.

- **Learning to classify new astronomical structures.**
  - Machine learning methods have been applied to a variety of large databases to learn general regularities implicit in the data.
    - For example, decision tree learning algorithms have been used by **NASA** to learn how to classify celestial objects from the second Palomar Observatory Sky Survey (Fayyad et al. 1995). This system is now used to automatically classify **all** objects in the Sky Survey, which consists of three terra bytes of image data.
- **Learning to play world-class backgammon.**
  - The most successful computer programs for playing games such as backgammon are based on machine learning algorithms.
    - For example, the world's top computer program for backgammon, **TD-GAMMON**(T esauro 1992, 1995). learned its strategy by playing over one million practice games against itself. It now plays at a level competitive with the human world champion.

Some disciplines and examples of their influence on machine learning.

- **Artificial intelligence**
  - Learning symbolic representations of concepts. Machine learning as a search problem. Learning as an approach to improving problem solving. Using prior knowledge together with training data to guide learning.
- **Bayesian methods**
  - Bayes' theorem as the basis for calculating probabilities of hypotheses. The naive Bayes classifier. Algorithms for estimating values of unobserved variables.
- **Computational complexity theory**
  - Theoretical bounds on the inherent complexity of different learning tasks, measured in terms of the computational effort, number of training examples, number of mistakes, etc. required in order to learn.
- **Control theory**
  - Procedures that learn to control processes in order to optimize predefined objectives and that learn to predict the next state of the process they are controlling.
- **Information theory**
  - Measures of entropy and information content. Minimum description length approaches to learning. Optimal codes and their relationship to optimal training sequences for encoding a hypothesis.
- **Philosophy**
  - Occam's razor, suggesting that the simplest hypothesis is the best. Analysis of the justification for generalizing beyond observed data.
- **Psychology and neurobiology**
  - The power law of practice, which states that over a very broad range of learning problems, people's response time improves with practice according to a power law. Neurobiological studies motivating artificial neural network models of learning.
- **Statistics**
  - Characterization of errors (e.g., bias and variance) that occur when estimating the accuracy of a hypothesis based on a limited sample of data. Confidence intervals, statistical tests.

WELL-POSED LEARNING PROBLEMS



if

Suite à l'attaque fautive (30-35) des Noirs, les Blancs forcent le gain...

WELL-POSED LEARNING PROBLEMS

- **A handwriting recognition learning problem:**
  - Task T: recognizing and classifying handwritten words within images
  - Performance measure ***P:*** percent of words correctly classified
  - Training experience E: a database of handwritten words with given classifications
- **A robot driving learning problem:**
  - Task T: driving on public four-lane highways using vision sensors
  - Performance measure P: average distance travelled before an error (as judged by human overseer)
  - Training experience E: a sequence of images and steering commands recorded while observing a human driver

Some Issues in Machine Learning

- What algorithms can approximate functions well (and when)?
- How does number of training examples influence accuracy?
- How does complexity of hypothesis representation impact it?
- How does noisy data influence accuracy?
- What are the theoretical limits of learnability?
- How can prior knowledge of learner help?
- What clues can we get from biological learning systems?
- How can systems alter their own representations?

Thank You !!!