Intelligent Systems: Recognition and Reasoning

James L. Crowley

ENSIMAG 2 — Second Semester 2019/2020
Lecture 1 — 5 Feb 2020

Intelligence: Recognition and Reasoning

Outline

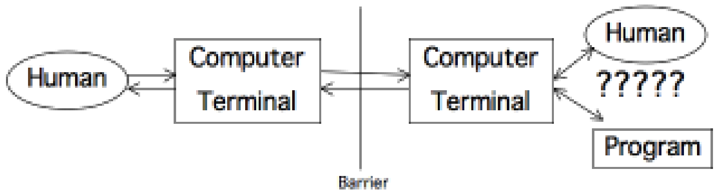
Class notes and exercises on the web:
http://www-prima.inrialpes.fr/Prima/Homepages/jlc/Courses/2019/ENSI2.SIRR/ENSI2.SIRR.html

Intelligence as a Description of Behavior

What do we mean by Intelligence? Alan Turing asked this question in 1950 and defined intelligence as a description of behavior.
According to Turing, intelligence is human level performance at tasks requiring perception, action and communication.

The Turing Test: an imitation game

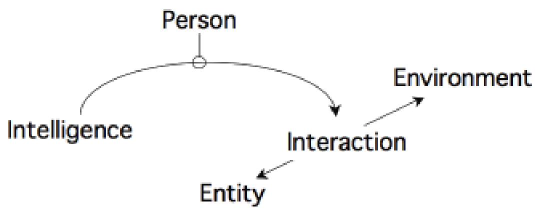
Alan Turing claimed that a machine would exhibit intelligence if it exhibited behavior that could not be distinguished from a person.



Turing proposed a test, in which a human observer interacts with an unknown agent over a teletype terminal. A machine (or program) is considered to be intelligent if the human observer cannot say whether it is a human or a machine.

Turing gave an important insight: Intelligence is NOT an intrinsic property of an agent. Intelligence is a "DESCRIPTION" of behavior.

Intelligence is a description of behavior.



Intelligence describes the interaction of an entity with its environment.
Intelligence is a description (a property assigned by an observer)
Intelligence describes an entity that interacts.

A modern definition defines intelligence as “Human level behavior at tasks requiring Perception, Action, and Interaction”.

Intelligence: Human-level performance at tasks requiring perception, action, communication, cognition or interaction

AI as a Scientific Discipline

Artificial Intelligence is the science and technology of intelligent systems.

Science: Science is the elaboration of theories and models that predict and explain phenomena (T Kuhn 76). Science is performed by communities who share paradigms (problems and problem solutions) and compete to publish in scientific journals and conferences.

Artificial Intelligence (AI): The science of artificial intelligent systems.

Scientific method is composed of empirical observation and objective documentation, followed by definition of concepts, and creation of theories and models. The value of a theory or model is in its ability to predict and explain phenomena. Theories and models are compared by scientific experiments. The model that best predicts the outcome of the experiment, and the theory that provide the simplest explanation, win. But all theories and models are only approximations to nature.

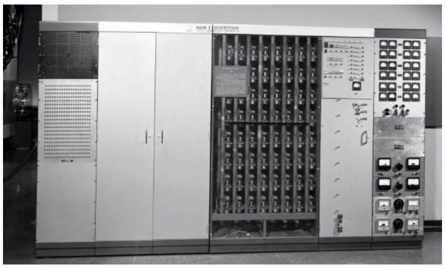
For sciences of the artificial, such as AI, the theories and models prescribe the design of systems systems, and explain their performance. For example, in electronics, theories and models explain and predict the relation between electric and magnetic fields and prescribe how to build a motor or a radio. There are generally many possible models for building a system. The key to comparison is performance evaluation. Thus, in lecture 2 of this course we will cover performance evaluation.

The modern scientific domain of AI emerged in the 1960s as a convergence of Cognitive Science, Logic, Planning, Pattern Recognition, Image Processing and other fields, driven by the emergence of Computer Science. The origin of the term "Artificial Intelligence" is generally credited to a Symposium at Dartmouth College in 1956, organized by John McCarthy, Marvin Minsky, and attended by the many AI pioneers, such as Arthur Samuel, Herb Simon, Allan Newell, etc.

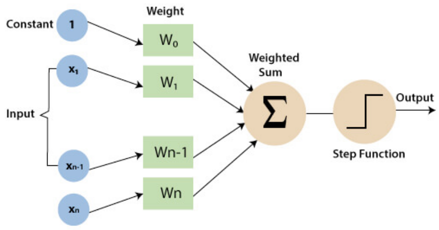
AI Pioneers at the Dartmouth Symposium (1956)

Machine Learning and Recognition

One pioneer that was NOT invited to the Dartmouth Symposium Frank Rosenblatt. In the 1950s, Rosenblatt invented a universal learning machine named the perceptron.



The Perceptron learning machine



The perceptron algorithm

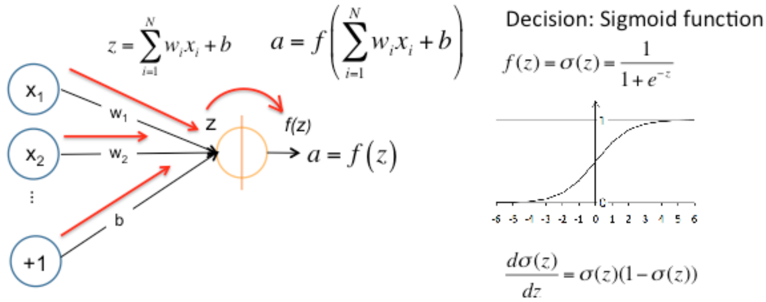
The perceptron was a Learning algorithm for a linear decision surface.
Problems: (1) Required labeled training data
(2) Could only classify linearly separable patterns
If the training data were not linearly separable, the algorithm would not terminate.

The perceptron was the original Artificial Neural Networks. We will start with the perceptron when we study Maximum Margin Classifiers such as the Support Vector Machine (lecture 6). This will be followed by a series of lectures on Artificial Neural Networks and Deep Learning (lectures 7 though 9) and a programming exercise

In 1969, Marvin Minsky published a text entitled "Perceptrons" that showed that the perceptron that showed that the learning algorithm would not terminate if the data could not be separated by a linear hyper-plane. The classic example was that a perceptron could not be programmed to imitate a simple logic gate such as the Exclusive OR. This destroyed the reputation of the perceptron learning algorithm. The dominant paradigm for building system that could learn to recognize was Bayes Rule.

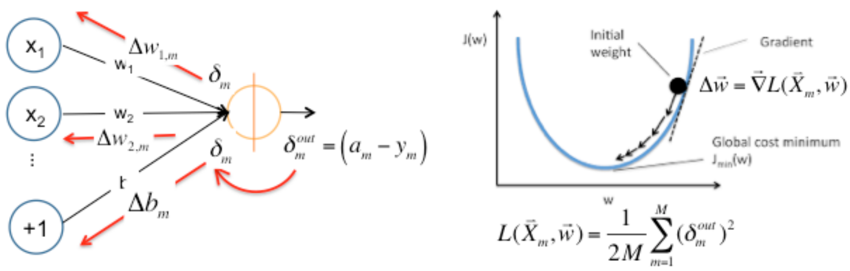
Artificial Neural Networks

During the 1980's, a small group of researchers continued to experiment with perceptrons. They found that problems with training with non-separable data could be overcome by using a soft decision surface, such as a sigmoid.



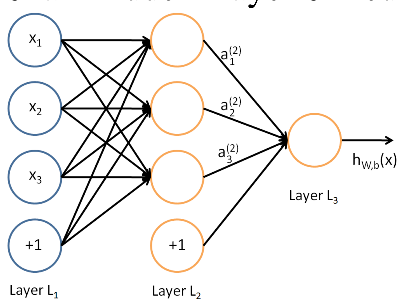
A feed forward neural network using a sigmoid decision functino

They found that the perceptron learning algorithm could be formulated as a form of optimization using gradient descent, and that gradient descent could be reformulated as a parallel algorithm, known as back propagation.



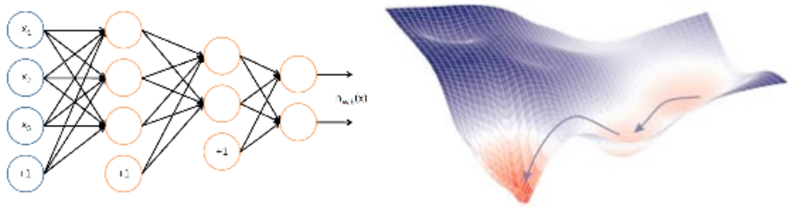
Back-propagation is a parallel form of Gradient Descent.

They found that limitations such as the inability to compute an XOR could be overcome by adding an additional "hidden" layer of neurons.



They were able to demonstrate that artificial neural networks were able to provide solutions to a number of problems in recognition for speech and computer vision that could not be solved by symbolic computing or expert systems.

Unfortunately attempt to generalize neural networks in the 1980s encountered several barriers.



Noise in the training data leads to local minima in the objective funtion.

1. Real-world problems led to networks with thousands (or millions) of parameters to tune, leading to excessively high cost in computation.
2. Training required labeled training data, and the required amount of data grew exponentially with the number of layers leading to prohibitive cost in data collection and labeling.
3. Real training data was noisy, leading to an loss function with many local minima, making training very unreliable. System performance was impossible to reproduce.
4. Neural networks are black boxes. No one can explain why they work or when they do not work.

By 1990, Neural networks were largely abandoned in favor of mathematically sound Bayesian approaches that we will study in lectures 3 to 6.

Bayesian Techniques for Recognition

Consider two classes of events A and B.

Let P(A) be the probability that an event belongs to class A
Let P(B) be the probability that an event belongs to class B
Let P(A, B) be the probability that the event is in both A and B.

Bayes rule tells us that conditional probability is the fraction of events that are B that are also A and B: $P(A \mid B) \equiv \frac{P(A,B)}{P(B)}$

From this we can write: $P(A \mid B)P(B) = P(A,B)$

Similarly $P(B \mid A) \equiv \frac{P(A,B)}{P(A)}$ so that $P(B \mid A)P(A) = P(A,B)$

This gives the common definition of Bayes Rule:

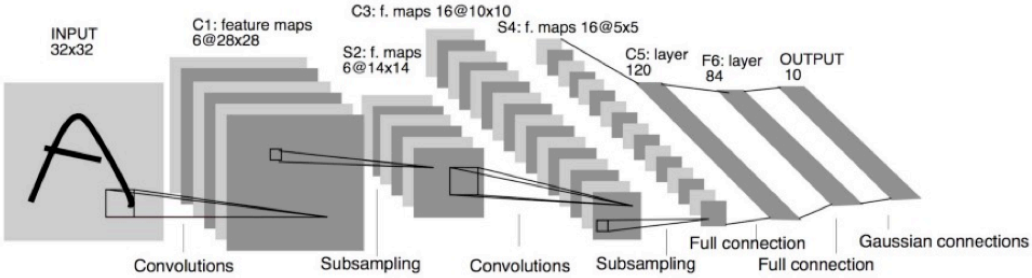
$$\boxed{P(A \mid B)P(B) = P(A,B) = P(B \mid A)P(A)}$$

We can use this to construct systems that learn to assign events to classes with the lowest possible probability of error.

We will review Bayesian methods for machine learning in lectures 3, 4 and 5. In lecture 3 we will review Bayes Rule and probability theory. In lecture 4 we will see that a simple "frequency based" or "statistical" interpretation of Bayes rule can lead to some very practical algorithms for building non-parametric systems for learning and classification. In lecture 5 we will show how an axiomatic approach to probability can lead to powerful non-supervised methods for learning, clustering and data mining such as the algorithms K-means and Expectation-Maximization followed by methods based on Support Vector machines in lecture 6.

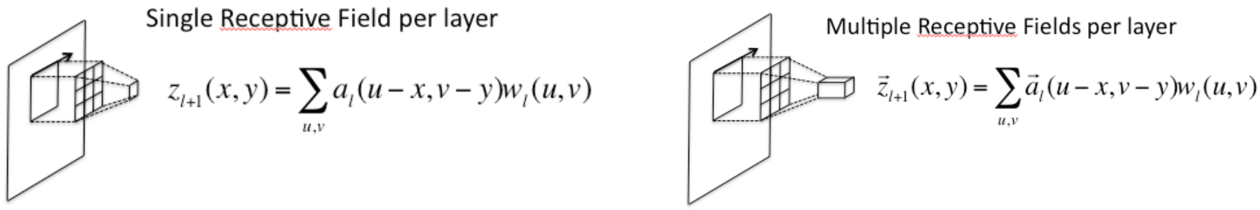
Convolutional Neural Networks

Although Neural Networks were largely abandoned, in the 1990's a small group of researchers continued to experiment. Using ideas from computer vision, Yann LeCun build a series of Convolutional Neural Networks. In 1994, one of his networks, LeNet5 won a competition for the best technique to recognize hand written characters on checks. This led to a commercial system for processing checks.



The LeNet 5 convolutional neural network - for recognizing hand-written digits.

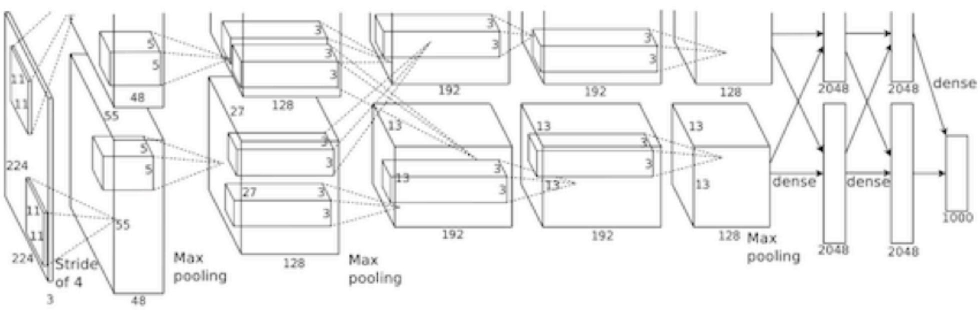
Convolutional networks learn a 2-D pattern of weights, called a receptive field.



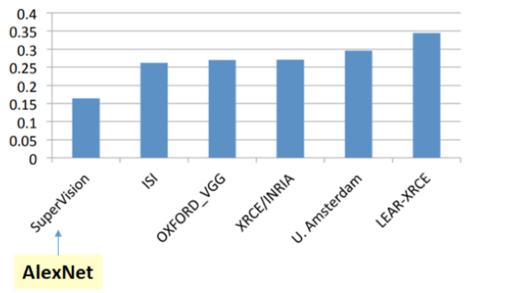
LeNet used back-propagation to learn several small receptive fields at each level.

In the early 2000's Computer Vision was increasingly driven by Challenge Based Research, in which researchers publish a data set and a base line algorithm and challenge the community to provide solutions with better performance.

The ImageNet Large Scale Visual Recognition Challenge was published in 2010. In 2012, the challenge was won by a network created by Alex Krizhevsky and Geoff Hinton based on the LeNet convolutional network. They called their network AlexNet.



AlexNet won by a large margin with an error of around 15%



This triggered a paradigm shift for Computer Vision, Speech Recognition, Machine Learning and (more recently) Artificial Intelligence.

So what changed between 1990 and 2010?

1. **Computing power**: Since 1948, as predicted by Moore's Law, computing power has doubled every 18 months, driven by the reduction in integrated circuit feature sizes (currently approx 10^{-5} meters (10 microns). In the 30 years since 1990, this has lead to a multiplication by 2^{20} in available computing power (aprox million fold growth!) With abundant GPUs and Cloud-based massively parallel Grid Computing the computing power is available.

2. **Data**: The internet and the World Wide Web has led to planetary scale data and the massive sharing of labeled training data. This is currently amplified by IOT and mobile computing.

3. **Reliability**: Improved understanding of optimization and learning algorithms have led to improved reliability and reproduce-ability.

We will study neural networks in lectures 7 through 12. You will do a programming exercise in March, using Keras and Pytorch to recreate a network that learns to recognize handwritten characters using the LeNet 5 architecture.

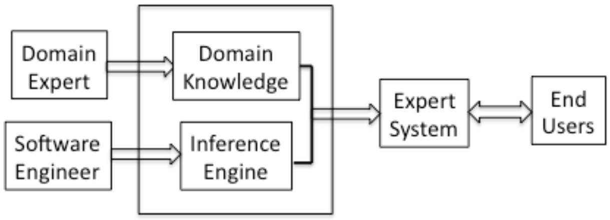
Since 2010 a major break through has occurred with the development of techniques based on Deep Learning. Deep Learning has been found to provide reliable solutions to longstanding problems perception and problem solving.

However, deep learning is only part of the solution. For human level performance at tasks requiring natural language requires explicit representation of knowledge. Natural language understanding offers access to the immense store of human knowledge coded in books and the internet, and to learn from and eventually teach humans.

Knowledge Representation and Reasoning.

Expert Systems

The field of AI gone through several “epochs”, each dominated by different paradigms (problems and problem solutions). In 1980, AI went through a revolution with the invention of Expert Systems. Expert systems use an "Inference Engine" to interpret an encoded form of Expert knowledge (a "Knowledge Base") in order to provide expert advice to users. Expert systems are constructed by a computer scientist (Knowledge engineer) working with a Domain Expert encode and test the "knowledge base".



The original expert system was the Mycin Antibiotic Therapy Advisor (1980). The system could interactively provide therapy advice for doctors to prescribe antibiotics. Other successful expert systems found rare-earth mineral deposits, helped configure Digital Equipment co. Vax Computers, and provided logistics planning for NASA.

AI researchers predicted that Expert Systems would come to dominate all of computer science. This lead to enormous investment and huge conference attendance. Expert systems were predicted to be a disruptive technology that would create an economic revolution. When the prediction failed to come true, funding and conference attendance dropped dramatically. However, research continued leading to a gradual convergence with cognitive science, and the emergence of symbolic computing techniques such as the semantic web and cognitive computing.

We will review expert systems in lecture 13. Much of the second half of the class will be devoted to symbolic reasoning and cognitive systems.

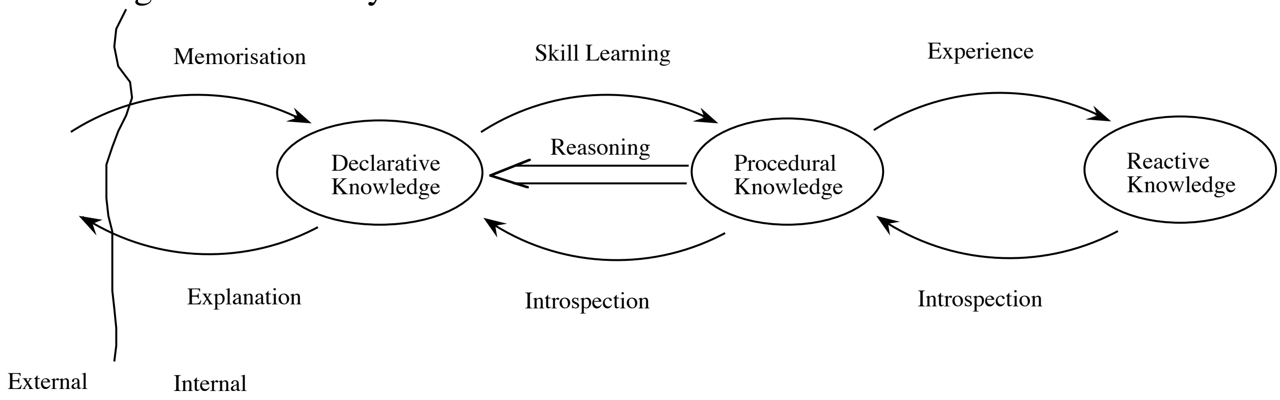
Symbolic reasoning and cognitive systems

We will see that schema systems can be used to represent both observed and abstract phenomena (concepts) and that these can be attached to methods for reasoning with Frames (lectures 15 and 16). We will see that relations between phenomena are fundamental to reasoning and study reasoning with spatial and temporal relations (lectures 13 and 17). Relations between phenomena are used to define a state based model and are fundamental both for planning and for Narrative and causal reasoning.

Kinds of Knowledge

In his 1980 Turing award lecture, Allen Newell ended the debate by defining knowledge as Competence. Anything that enabled the solution of problems.

Cognitive Psychologists identify different categories of knowledge representation.
A common cognitive model organized Declarative, Procedural and Reactive knowledge as a hierarchy.



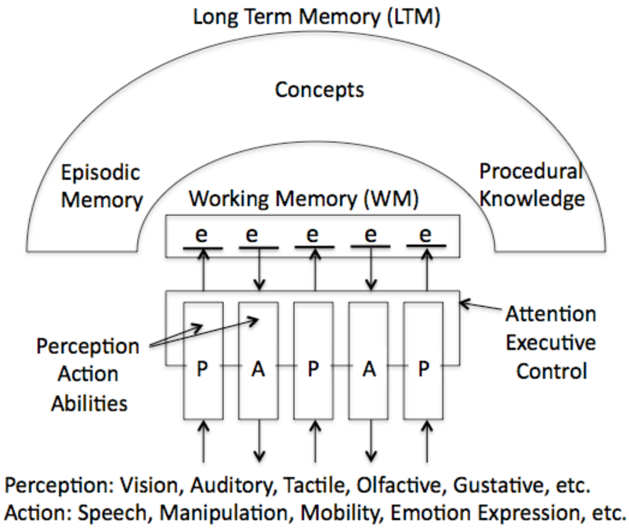
Declarative: A symbolic expression of competence.
Declarative knowledge is abstract
Declarative knowledge is used to communicate and to reason.
Declarative knowledge must be interpreted to be used.
Procedural: A series of steps to solve a problem.
Reactive: Automatic response to stimulus; conditioned reflex

Concepts, Working Memory and Cognition

Modern cognitive scientists study intelligence as interaction different forms of memories. Most models posit a cognitive architecture composed of a limited working memory interacting with perception and action based on episodic, procedural and conceptual memories.

Symbolic descriptions used for such models are like the mathematics used in other sciences. They describe functionality for interacting components. Actual systems may be implemented with a variety of programming tools, including Neural Networks.

A common cognitive model used for study of comprehension centers reasoning on processing of concepts in working memory.



(inspired by Rasmussen, Card-Moran-Newell, Anderson, Kintsch, many others)

Most models of human cognitive models share a number of common elements:

- Perception: Transforms and combines sensory stimuli to Phenomena
- Short Term Perceptual Memory: Temporary buffer holding recent stimuli
- Action: Activation patterns for muscle groups.
- Working Memory: 7+/-2 memory slots (perceived or remembered)
- Long Term Memory

Long-term memory (LTM) refers to memory structures used in several different cognitive abilities:

- Episodic Memories: recordings of significant sensory experiences
- Semantic Memory: Abstract representations for sensory experiences
- Procedural Memory: Sequences of operations to accomplish goals
- Spatial memory (Spatial relations between places)

Course Overview

Part 1 – Recognition and Machine Learning

1) Supervised learning and Performance Evaluation
2) Bayesian Learning, non-parametric methods.
3) Non-supervised learning with EM and K-Means
4) Support Vector Machines
5) Artificial Neural Networks, Back Propagation, and Architectures.

Part 2 – Reasoning

1) Knowledge Based Systems
2) Schema Systems, Frames, Structured Knowledge
3) Temporal and Spatial Reasoning
4) Planning and problem solving
5) Narrative and Causal Reasoning

Every week I will hand out exercise problems.
Exercises will NOT be graded.
However, the exam will be composed of (modified) versions of the exercises!
Do the exercises and the exam will be easy.
Ignore the exercises and the exam will be very hard.

Exercises may be done individually or in a group.
Exercises should be done within 2 weeks of assignment.

Completed exercises should be COPIED INTO AN EMAIL including the names of all persons who contributed to the solution. It is acceptable to work the exercise on paper and send a photo by email. The photo must be easily readable.

Feedback will be returned by email. Please allow at least 2 weeks for feedback.