Neural Networks
Lecture Notes
Introduction

Marcel van Gerven

1 Learning Goals

After studying the lecture material you should:

1. be able to place neural networks in their historical context
2. know the different network architectures.
3. know what is meant by input vector, input activation, activation function, objective function, weight vector, output.
4. understand the threshold activation functions.
5. be able to explain the difference between rate coding and temporal coding.
6. be able to recognize the equivalent circuit diagram as an implementation of the H&H model.
7. understand what is meant by distributed versus localist representations.
8. know what is meant by cybernetics.
9. understand what is meant by Hebbian learning.
10. know the difference between the three kinds of learning paradigms.

2 Notes

2.1 Properties of Artificial Neural Networks

An artificial neural network (ANN) is a system of interconnected processing units (artificial neurons) which exchange messages between each other. The connections have numeric weights $\mathbf{w}$ that can be tuned based on experience to maximize a certain *objective function*, making ANNs capable of learning.

Neural networks come in various flavours. Important distinguishing properties of neural networks are:

1. Neuron types (what values can artificial neurons take?)
2. Network architecture (how are artificial neurons connected?)
3. Activation function (how is neuron input translated to neuron output?)
4. Learning algorithm (how are network weights learned?)

Neuron types can be e.g. *binary* or *continuous-valued*. Network architecture can be *feed-forward*, *recurrent*, *undirected* or *modular*. Let $\mathbf{x} = (x_1, \ldots, x_N)^T$ be the inputs to some neuron y. Let $\mathbf{w} = (w_1, \ldots, w_N)^T$ be a weight matrix such that w_i denotes the weight between input x_i and output y. Then, the input *activation* is given by $a = \sum_i w_i x_i$, which can be written in vector form as $a = \mathbf{w}^T \mathbf{x}$. The output of neuron y is then given by $y = f(a)$, where f is the *activation function*. Several activation functions have been developed over the years. The LTU makes use of a (unipolar) *threshold function*, which is specified by

$$f(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{if } a < 0 \end{cases}$$

In order to modify network weights, we can make use of three kinds of learning paradigms:

- *Supervised learning*: Learn to predict an output when given an input vector
- *Unsupervised learning*: Discover a good internal representation of the input
- *Reinforcement learning*: Learn to select an action to maximize payoff

2.2 A Brief History of Artificial Neural Networks

An artificial neural network (ANN) is computational model that consists of idealized artificial neurons. ANNs are used to estimate or approximate highly non-linear functions that can depend on a large number of inputs. They are inspired by biological neural networks in two respects [Hay94]: First, knowledge is acquired by the network through a learning process. Second, interneuron connection strengths referred to as (synaptic) weights are used to store the knowledge. Artificial neural networks have been around for over seventy years [MP43] but have fallen in and out of favour several times throughout the course of their history.

Neural network theory, also known as *connectionism*, builds on several classical ideas. *Associationism* is a theory developed by Aristotle which assumes that mental processes operate by the association of one mental state with its successor states. Aristotle identified four laws by which associations might come about:

The law of contiguity. Things or events that occur close to each other in space or time tend to get linked together in the mind.

The law of frequency. The more often two things or events are linked, the more powerful will be that association.

The law of similarity. If two things are similar, the thought of one will tend to trigger the thought of the other.

The law of contrast. On the other hand, seeing or recalling something may also trigger the recollection of something completely opposite.

Connectionism is also strongly rooted in *empiricism*, as developed by John Locke in the 18th century. It states that human knowledge is derived from sensory experience and it is the association of these experiences that lead to thought. Therefore, human cognition is governed by physical laws and can be studied empirically (e.g. through observations). As such, it is also rooted in *materialism*, as developed by de la Mettrie, which assumes that matter is the fundamental substance in nature, and that all phenomena, including mental phenomena and consciousness, are identical with material interactions.

Theories that link psychological states such as associations to neural processing started to emerge in the 19th century. Spencer was convinced that understanding the biological nervous system was crucial for understanding psychological states [Spe55]. He postulated that connections between neurons are related to connections between thoughts and concepts. James, in turn, developed the *law of neural habit*: When two elementary brain-processes have been active together or in immediate succession, one of them, on reoccurring, tends to propagate its excitement into the other [Jam90]. This already starts to take the form of a theory which is implementable in computational models.

In the early 20th century, neuroscience started to take off, with Santiago Ramón y Cajal defending the neuron doctrine and Golgi defending the reticular theory [DB07]. While Cajal was right, as neurons are separated by the synaptic cleft, the existence of gap junctions that allow molecules to directly pass between neurons at least saves some of Golgi's theory.

In the early 20th century, behaviourism, as developed by Watson, Skinner, Thorndike & Hull took center stage. It maintains that psychology should concern itself with observable events rather than introspection in order to make testable predictions. Thorndike's *law of effect*, which states that the effect of an event alters the connections between event and response is especially interesting as it forms the basis for reinforcement learning.

Connectionism was first coined by Donald Hebb in the 1940's. It refers to the modeling of mental or behavioral phenomena as the emergent processes of interconnected networks of simple units (i.e. artificial neural networks). In his book, The Organization of Behaviour, he put forward an explicit learning rule by stating [Heb49] *"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."* This rule is known as *Hebbian learning* and is nicely summarized by the statement: *"Cells that fire together, wire together."*

Connectionism also relies heavily on the notion of *distributed representations*, which assumes a many-to-many relationship between two types of representation (e.g. concepts and neurons). That is (1) each concept is represented by many neurons, and (2) each neuron participates in the representation of many concepts. The brain as a distributed system was already put forward by Hughlings Jackson in 1869. It is also related to notions of *equipotentiality* (all cortical areas can substitute for each other) and *mass action* (Reduction in learning is proportional to the amount of tissue distroyed) put forward by Lashley in his search of the engram.

Connectionism can be contrasted with *computationalism* as developed by Putnam, Fodor, Chomsky and Pinker. It is a specific form of cognitivism which assumes that the brain works via formal symbol manipulation. Here, *cognitivism* refers to the study in psychology which focuses on mental processes in contrast to observable behavior, as in behaviorism. Computationalism contrasts with connectionism as the latter focuses on understanding cognition not via formal symbol manipulation but rather via parallel distributed processing between simple (sub-symbolic) processing elements. Though, both theories are not necessarily incompatible, it led to polarization between two camps and the development of various *post-cognitivist theories* (e.g Dreyfus, Bateson, Maturana, Varela) that renounced computationalism, emphasizing sub-symbolic reasoning, embodiment, et cetera.

The mid-twentieth century also saw the rise of *cybernetics*, which is the study of control and communication in the animal and the machine, as developed by Norbert Wiener. The work by Ashby and Conant led to the *Good Regulator Theorem*, which states that every good regulator of a system must be a model of that system. For the brain, as a regulator for survival, iit implies that it must learn a model of its environment.

The first artificial neural network was developed by McCulloch & Pitts [**?**] and strongly depended on earlier work by Alan Turing. They developed a mathematical model of networks of artificial neurons (*linear threshold units*; LTUs) that can account for high level cognitive tasks. It was founded on logic (implementation of Boolean functions) and they showed that any statement within propositional logic could be represented by a network of such simple processing units (cf. Turing Machines) given suitably chosen weights. Hence, if any number can be computed by an organism, it is computable by these definitions, and vice versa.

Note that the artificial neurons used by McCulloch & Pitts are far removed from the properties of real neurons. For instance, action potential generation as described by Hodgkin & Huxley [HH52] is not modeled using such neurons. Instead they make use of *rate coding*, where continuous input activations reflect firing rates that are used for computation. This contrasts with *temporal coding*, where neurons are assumed to make use of the timing of individual action potentials for computation. The latter is used in *spiking neuron models* that aim to approximate the behaviour of the Hodgkin & Huxley model. The H&H model can be described in terms of an equivalent circuit diagram.

Over the past decades various neural network models have been developed. Alan Turing was one of the pioneers with his little-known work on unorganised machines. The first generation of neural networks are known as perceptrons. They were one of the first models which were able to efficiently learn how to adapt their weights. The perceptron built on work by various other pioneers in the field. The second generation of neural networks were models such as the multi-layer perceptron, which extended the capabilities of perceptrons, Hopfield networks and Boltzmann machines, which arose from statistical physics, and radial basis function (RBF) networks and self-organizing maps (SOMs), which can be used for function approximation and clustering. Third-generation neural networks that focus on deep and recurrent neural network learning have become extremely popular in the last decade due to breakthroughs in ANN training. Spiking neural networks are also heavily studied, especially in computational neuroscience.

3 Reading material

Nice historical overviews on the early days of neural networks are given in [AR98, Med98].

References

[AR98] J A Anderson and E Rosenfeld. *Talking Nets: An Oral History of Neural Networks*. 1998.

[DB07] Juan A De Carlos and José Borrell. A historical reflection of the contributions of Cajal and Golgi to the foundations of neuroscience. *Brain Res. Rev.*, 55(1):8–16, 2007.

[Hay94] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan, New York, 1994.

[Heb49] Donald O Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

[HH52] Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.*, 117(4):500, 1952.

[Jam90] Willliam James. *The Principles of Psychology*. Dover, New York, 1890.

[Med98] DA Medler. A brief history of connectionism. *Neural Comput. Surv.*, pages 1–56, 1998.

[MP43] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5(4):115–133, 1943.

[Spe55] Herbert Spencer. *The Principles of Psychology*. Longman, 1855.