UNIVERSITY OF LEEDS

Class: Machine Learning

Neural Networks: Perceptron

Instructor: Matteo Leonetti

Learning outcomes

UNIVERSITY OF LEEDS

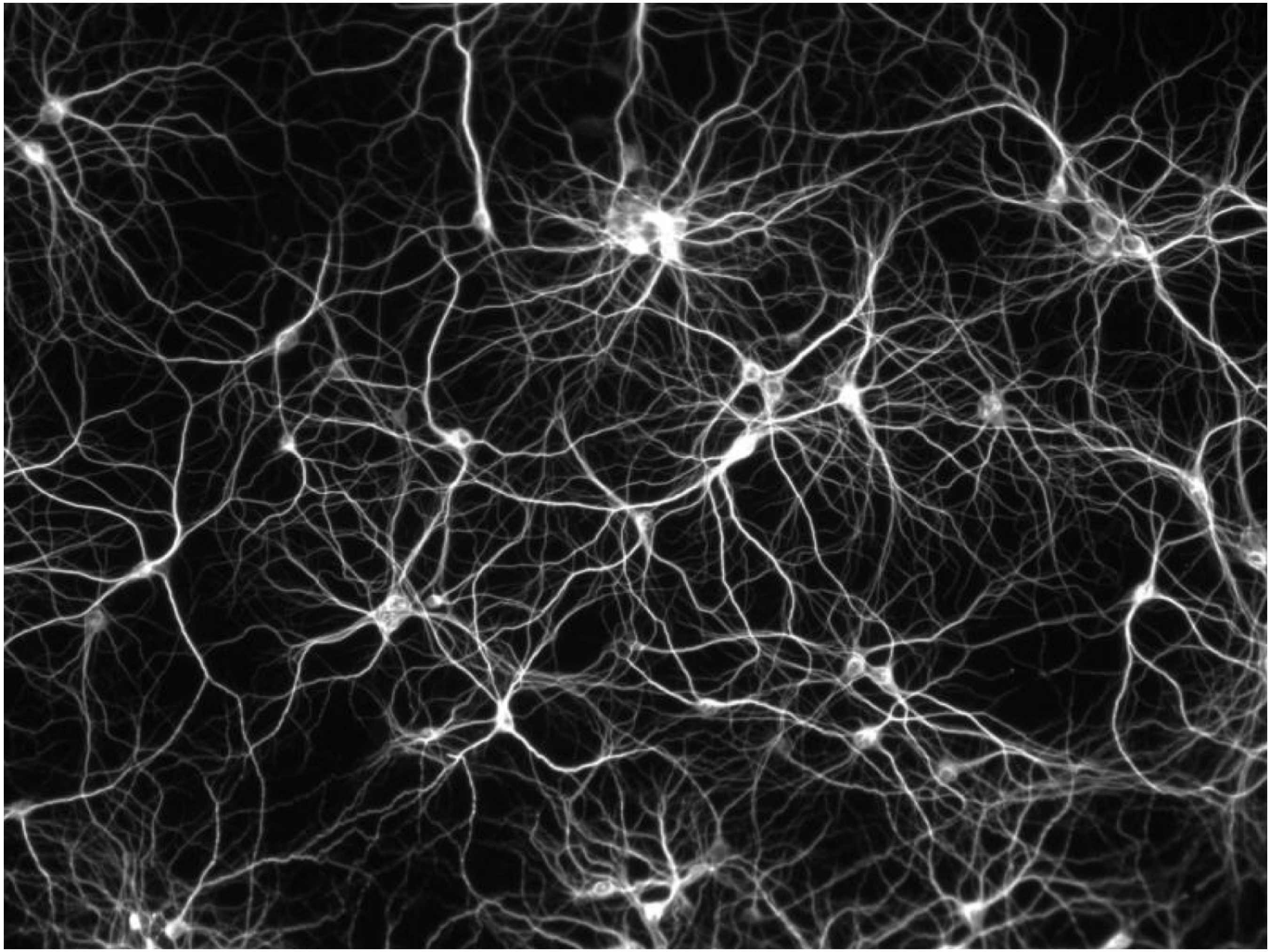
- Describe the biological principles that inspired neural networks.
- Draw the diagram of the McCulloch and Pitts's neuron.
- Distinguish between generative and discriminative learning models.

Demo

https://cs.stanford.edu/people/karpathy/convnetjs/

Our brain

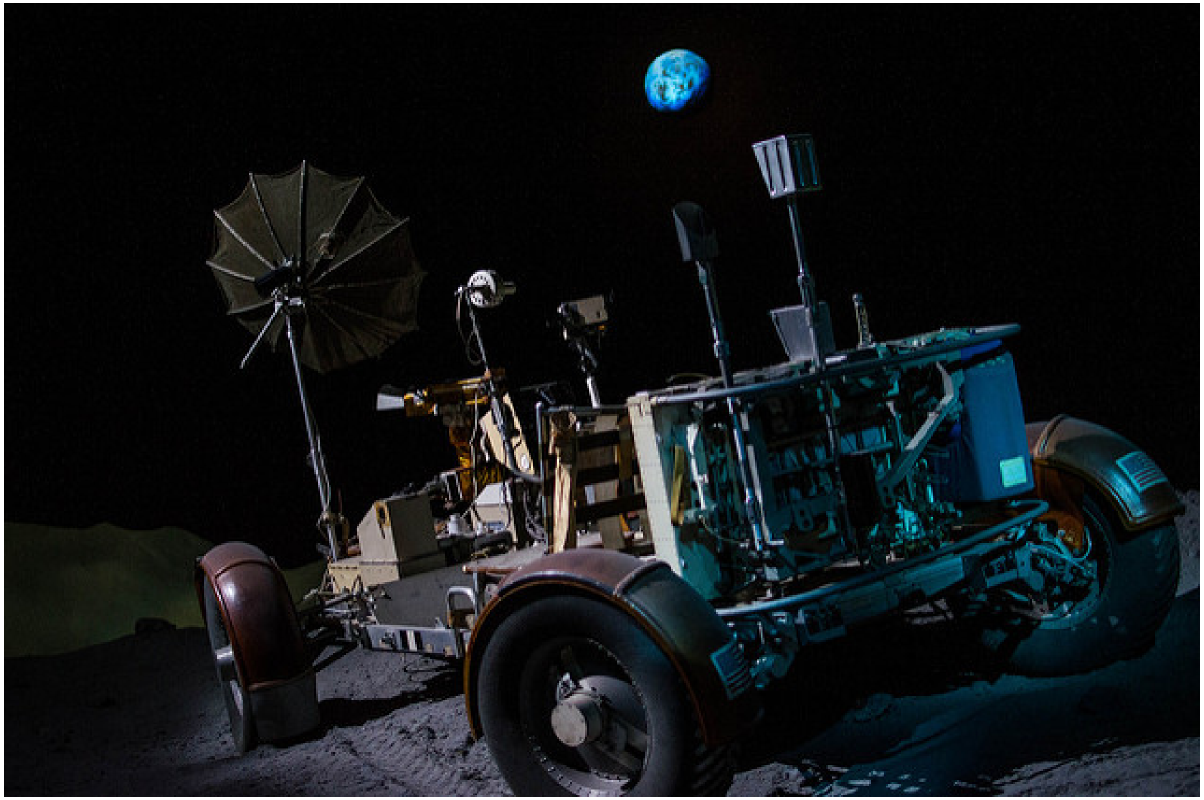
UNIVERSITY OF LEEDS



Neurons

10^{11} neurons 10^{15} synapses

Enough fibres to cover the distance to the moon and back

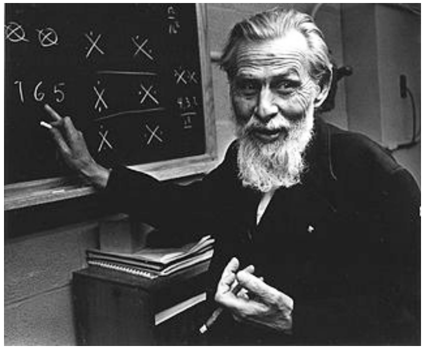


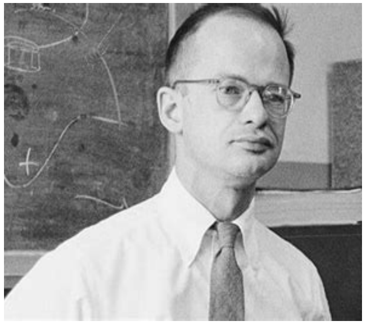
History

UNIVERSITY OF LEEDS

- 1943 – McCulloch and Pitts

  A Logical Calculus of the Ideas Immanent in Nervous Activity





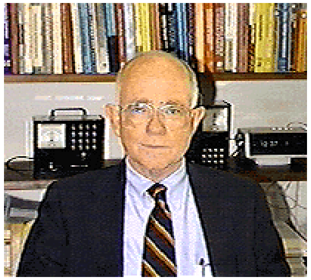
- 1949 Donald Hebb

  The Organization of Behavior

- 1958 Frank Rosenblatt
  The perceptron: A probabilistic model for information storage and organization in the brain.

History

UNIVERSITY OF LEEDS




- 1962, Widrow and Hoff

  Adaptive Switching Circuits

- *1965 Alexey Ivakhnenko*

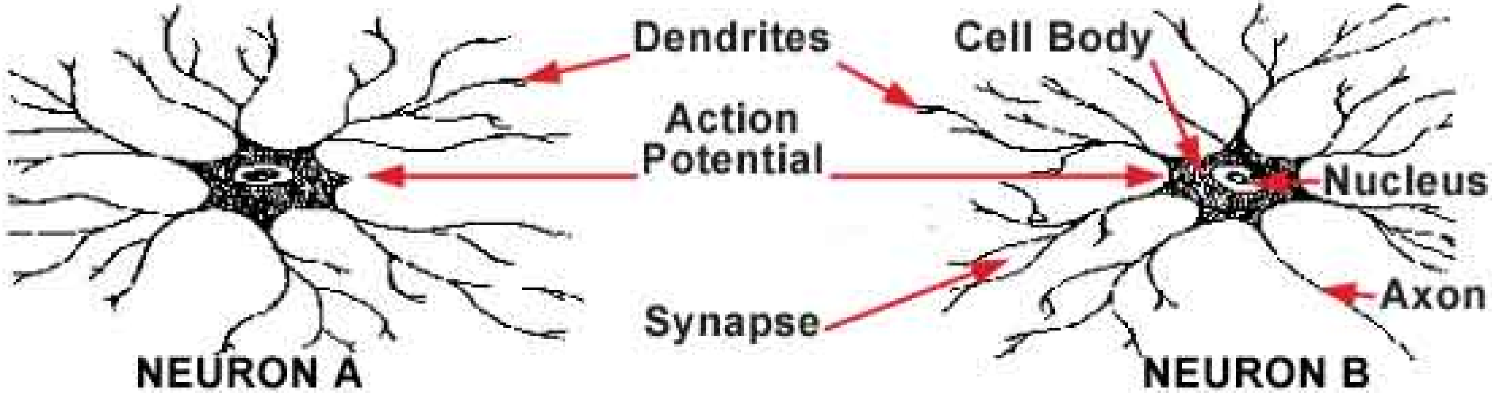
- 1975, Paul Werbos

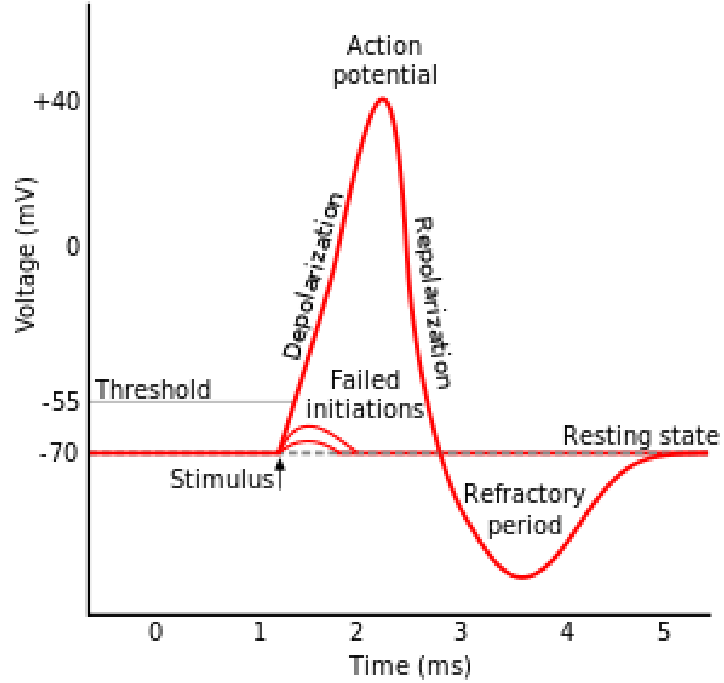
  Backpropagation

- …

- 2009- now, Deep Learning: Dayan (Max Planck Tubingen), Schmidhuber (IDISA), Hinton (Google, U Toronto) , Bengio (U Montreal), LeCun (FAIR), … and more!

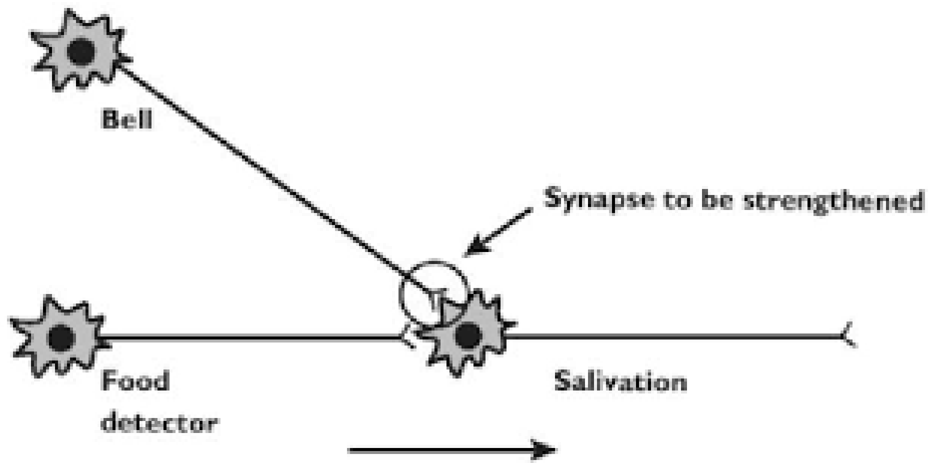
Firing


Dendrites
Cell Body
Action Potential
Nucleus
Synapse
Axon
NEURON A
NEURON B


Action potential
+40
0
-55
-70
Voltage (mV)
Depolarization
Repolarization
Threshold
Failed initiations
Resting state
Stimulus
Refractory period
0
1
2
3
4
5
Time (ms)

Hebbian learning

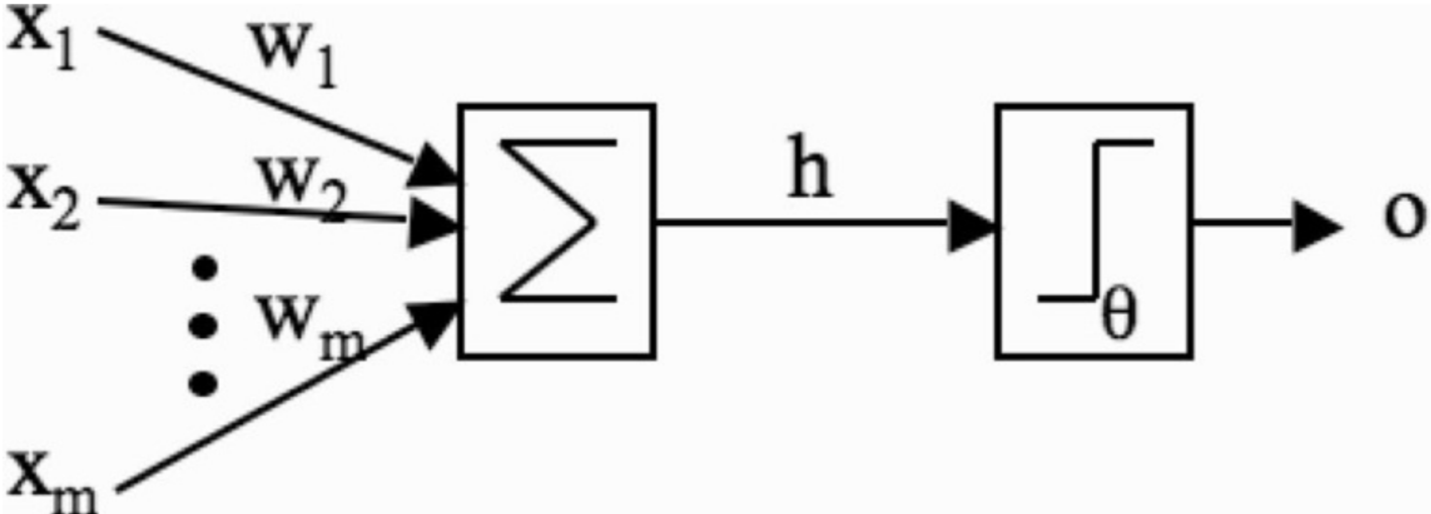
UNIVERSITY OF LEEDS



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

Donald Hebb (1949)

McCulloch and Pitts Neuron



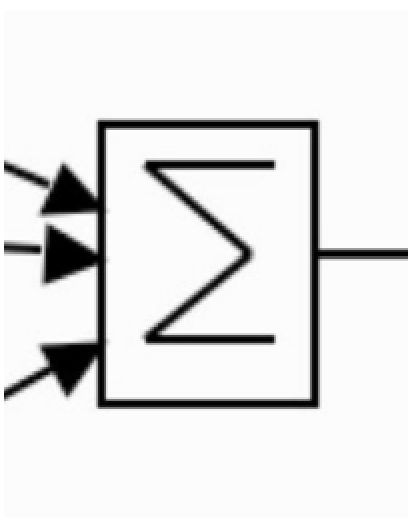
$$h_w(\boldsymbol{x}) = \sum_i w_i x_i = \boldsymbol{w} \cdot \boldsymbol{x}$$

$$o(h_w) = \begin{cases} 1 & \text{if} \quad h_w > \theta \\ 0 & \text{if} \quad h_w \leq \theta \end{cases}$$

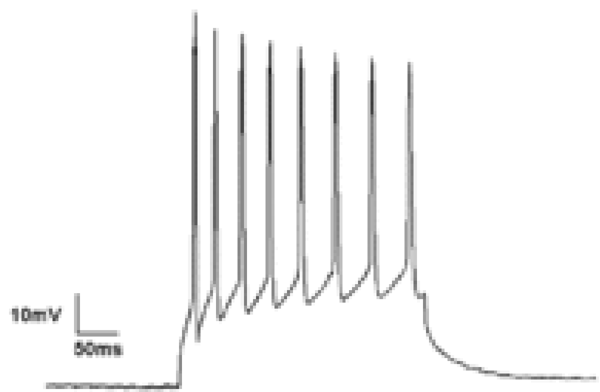
Activation function

Model Critique

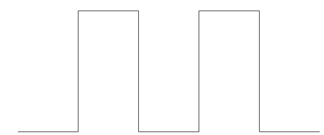
?



Single threshold vs spike train

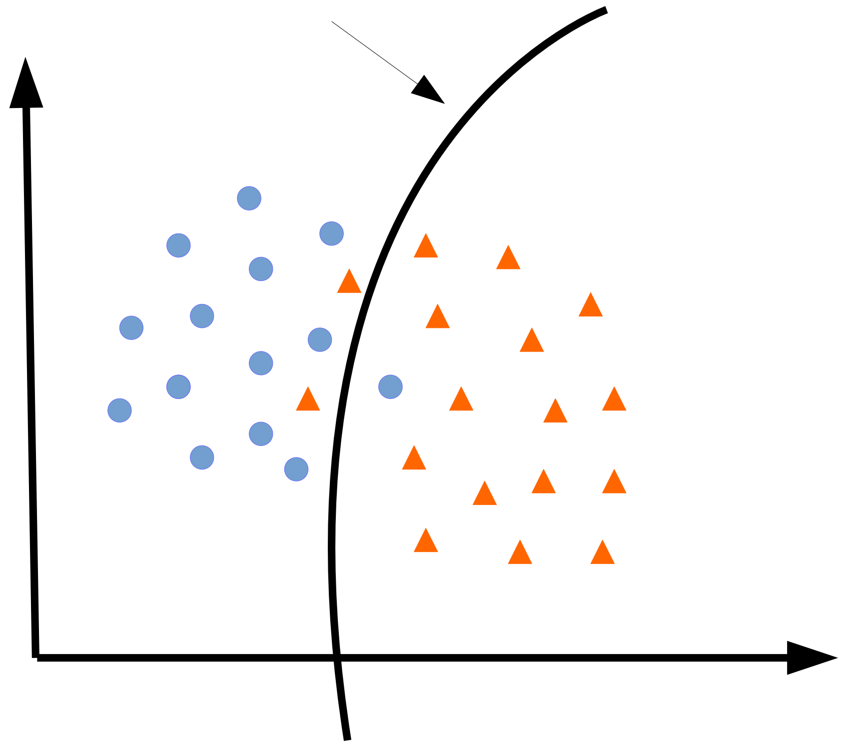


No clock → asynchronous



A function approximator

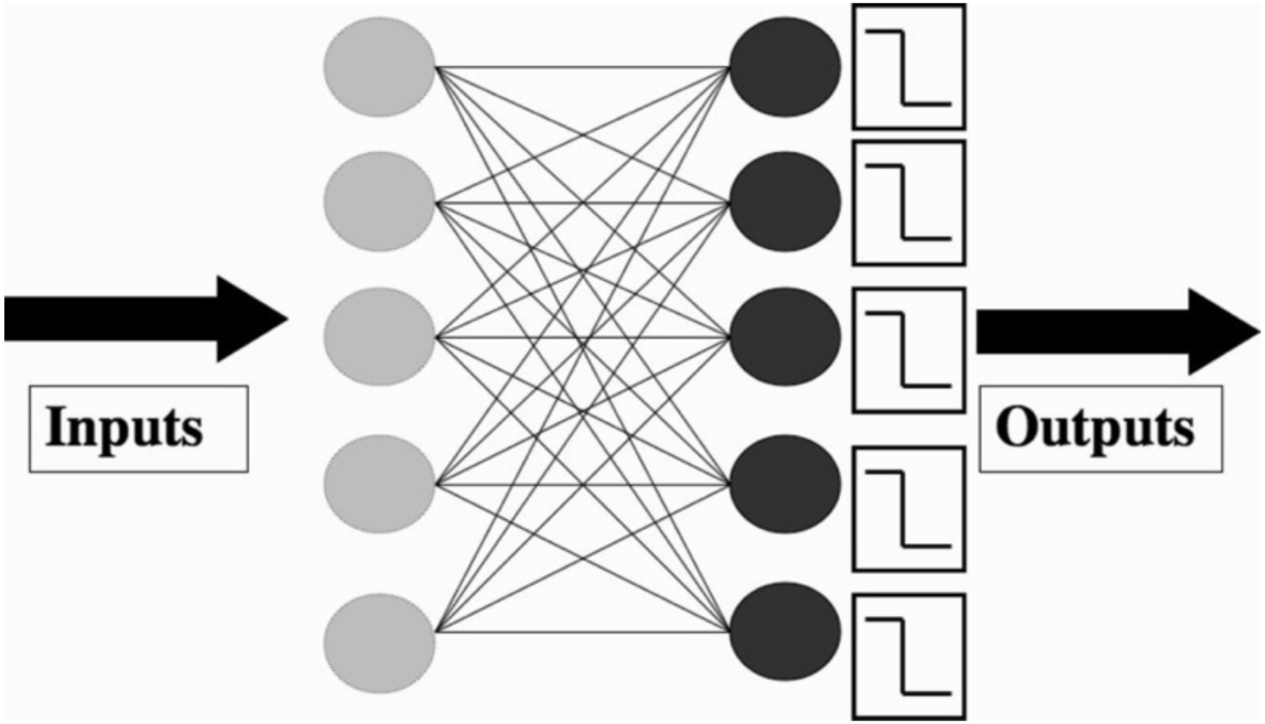
Decision boundary to model
(approximate) → regression



Generative vs Discriminative model

Perceptron

UNIVERSITY OF LEEDS



Training the perceptron

Learning happens through **optimisation**.

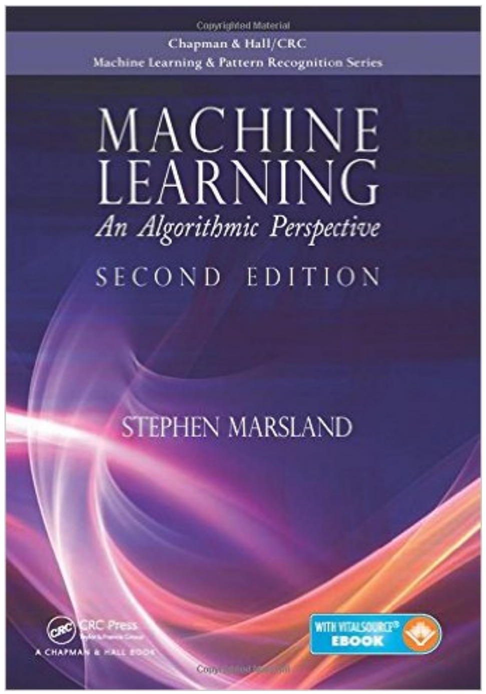
We define an error function, and then an optimisation algorithm finds the parameters that obtain the minimum error.

For example, the error function is the total number of mistakes:

$$E(\boldsymbol{X}) = \sum_{\vec{x}_n \in \boldsymbol{X}} |y_n - t_n|$$

Where y_n is the output of the perceptron on point n, and $t_n \in \{0,1\}$ is the desired class, and $\boldsymbol{X}$ is the dataset.

Conclusion



Chapter 3, up to 3.3