The Perceptron

Foundations of Data Analysis

04/09/2020

History of Perceptron

- Frank Rosenblatt
- 1928-1969

invented perceptron algorithm

History of Perceptron

- Marvin Minsky
- 1927-2016

stated that a simple perceptron is limited to linearly separable problems

Perceptron Learning Algorithm

- First neural network learning model in the 1960's

Perceptron Learning Algorithm

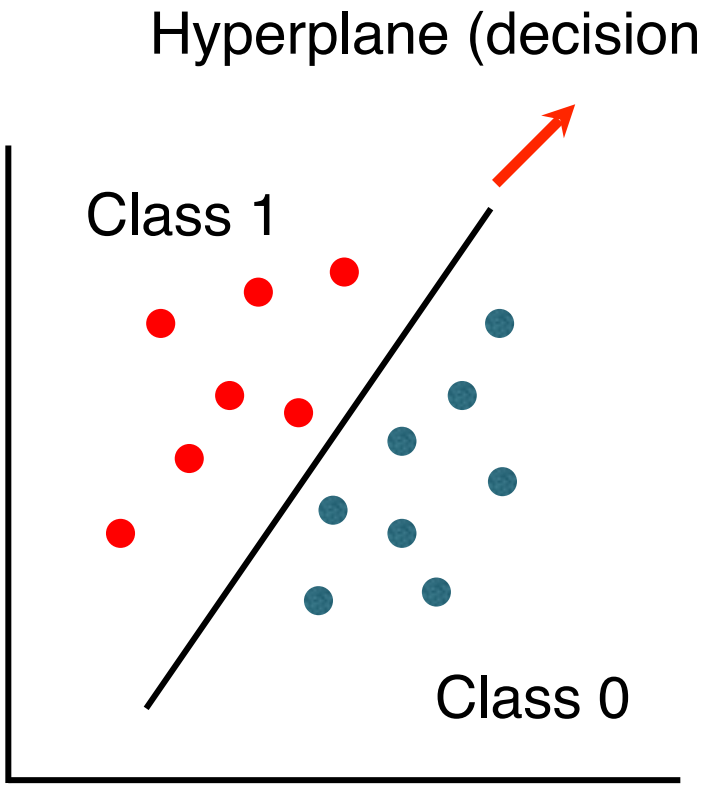
- First neural network learning model in the 1960's
- Simple and limited (single layer model)

Perceptron Learning Algorithm

- First neural network learning model in the 1960's
- Simple and limited (single layer model)
- Basic concepts are similar to multi-layer models

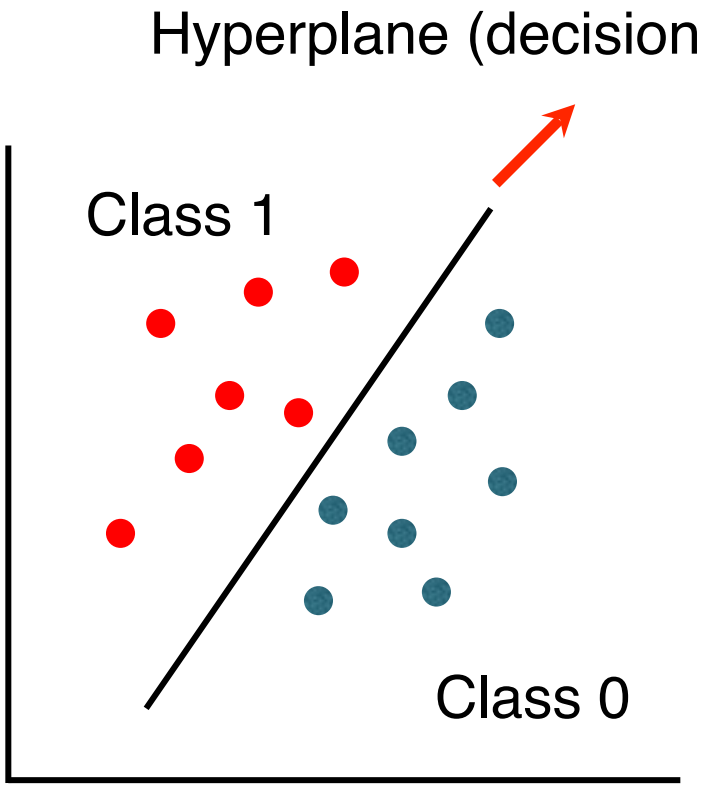
What is Perceptron?

The goal of perceptron algorithm is to find a hyperplane that separates a set of data into two classes.



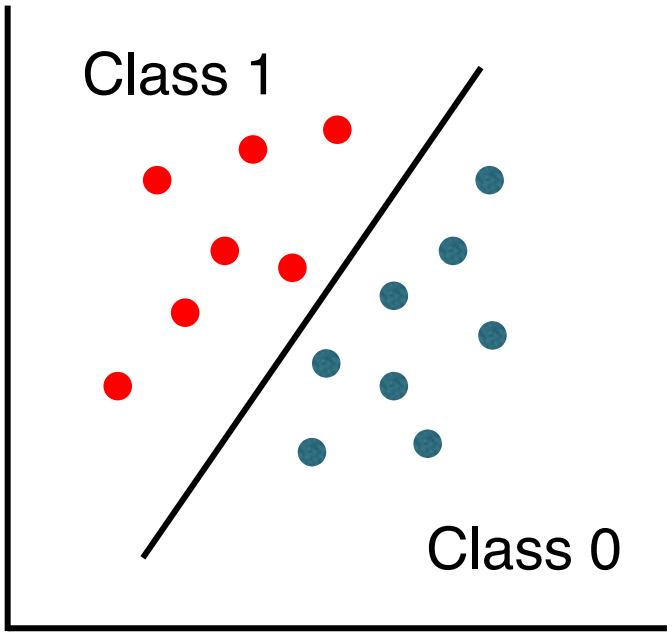
What is Perceptron?

The goal of perceptron algorithm is to find a hyperplane that separates a set of data into two classes.



- Binary classifier
- Supervised learning

Perceptron

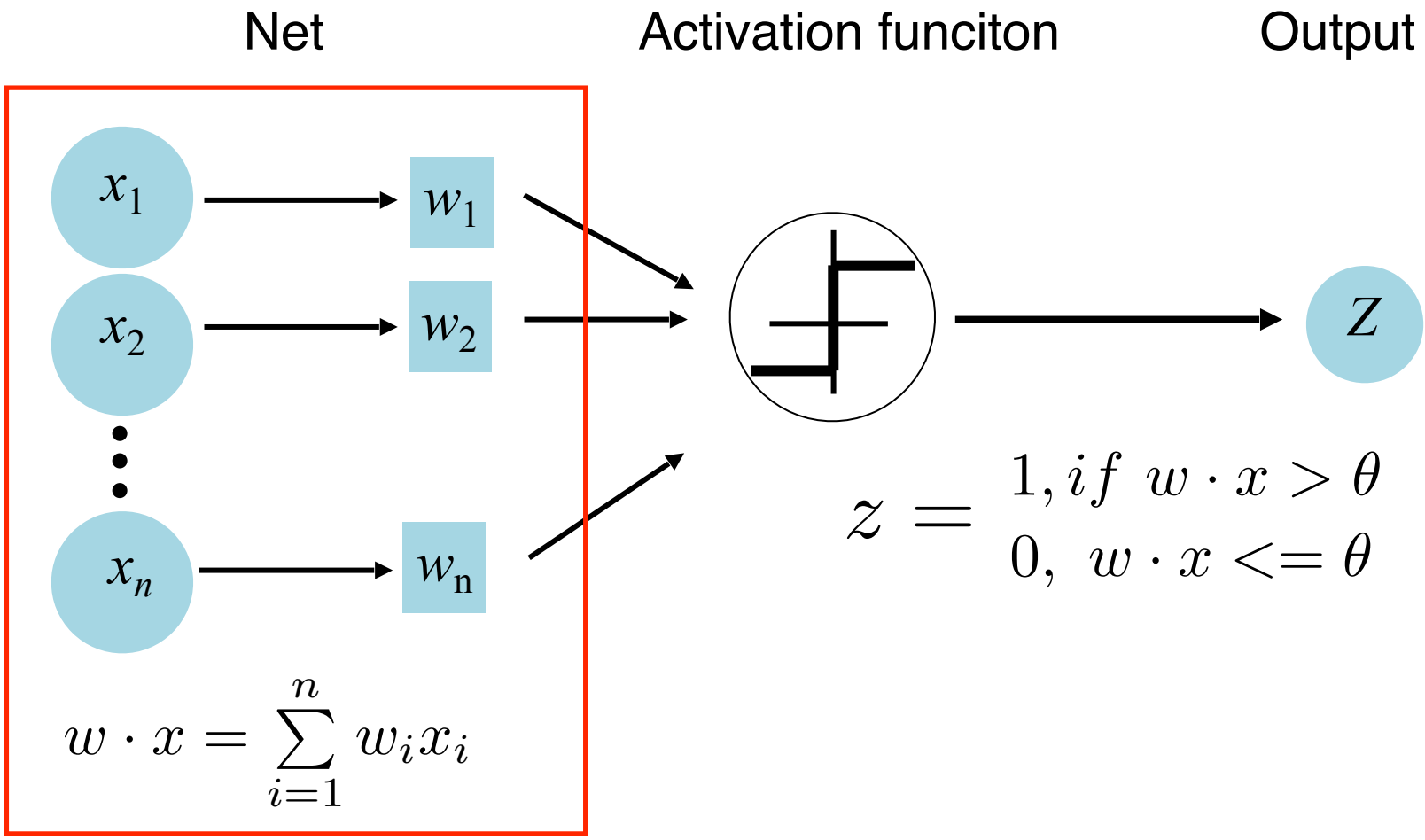


bias term

$$f(x) = \begin{matrix} 1, if\ w \cdot x + \theta > 0 \\ 0, otherwise \end{matrix}$$

$$w \cdot x = \sum_{i=1}^{n} w_i x_i \quad \text{(dot product)}$$

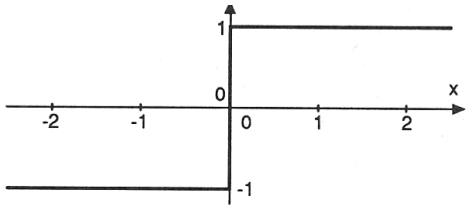
Perceptron



- Learning weights such that an objective function is minimized

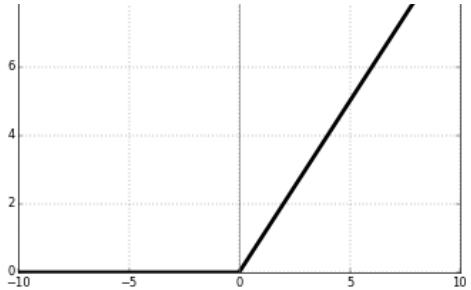
Activation Function

Outputs the label given an input or a set of inputs.



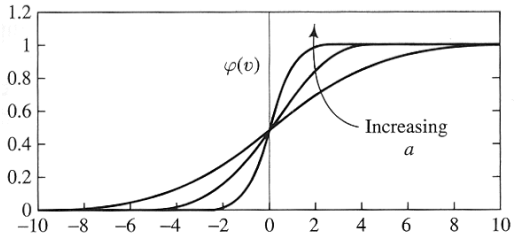
Step function

$$f(x) := \operatorname{sgn}(x) = \begin{cases} 1 & if \quad x \geq 0 \\ -1 & if \quad x < 0 \end{cases}$$



ReLU (rectified linear unit)

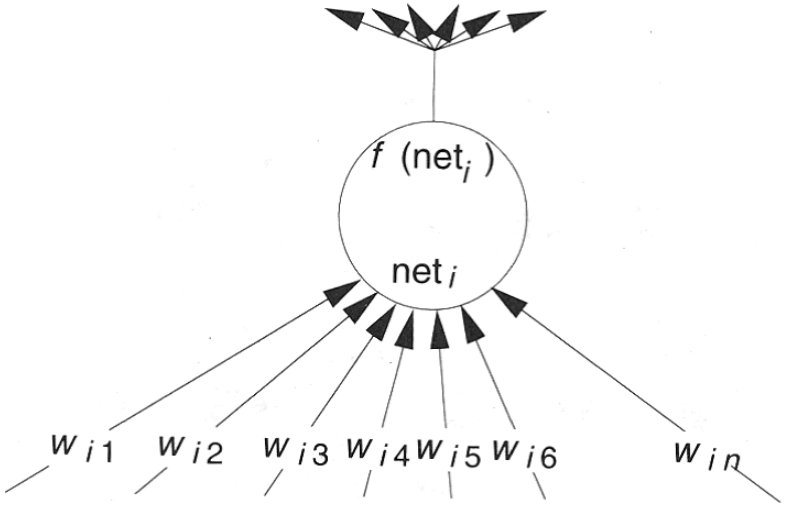
$$f(x) = max(0, x)$$

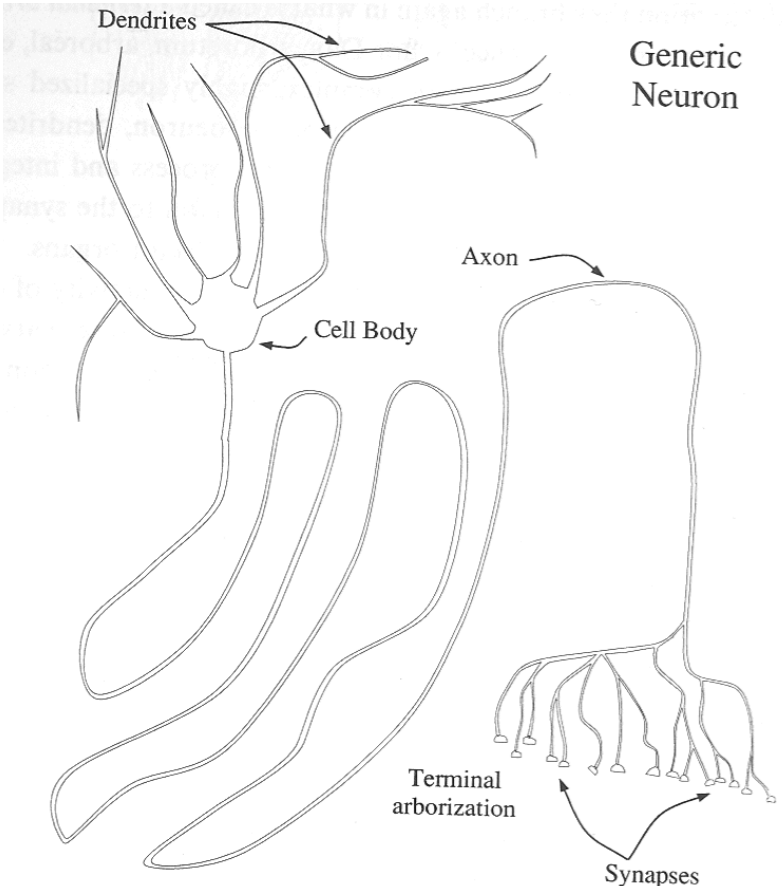


Sigmoid function

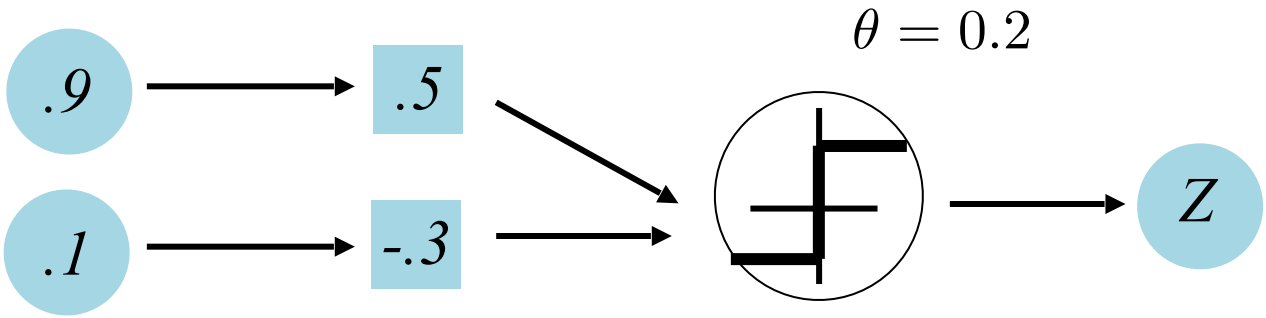
$$f(x) := \sigma(x) = \frac{1}{1 + e^{(-ax)}}$$

Perceptron as a Single Layer Neuron

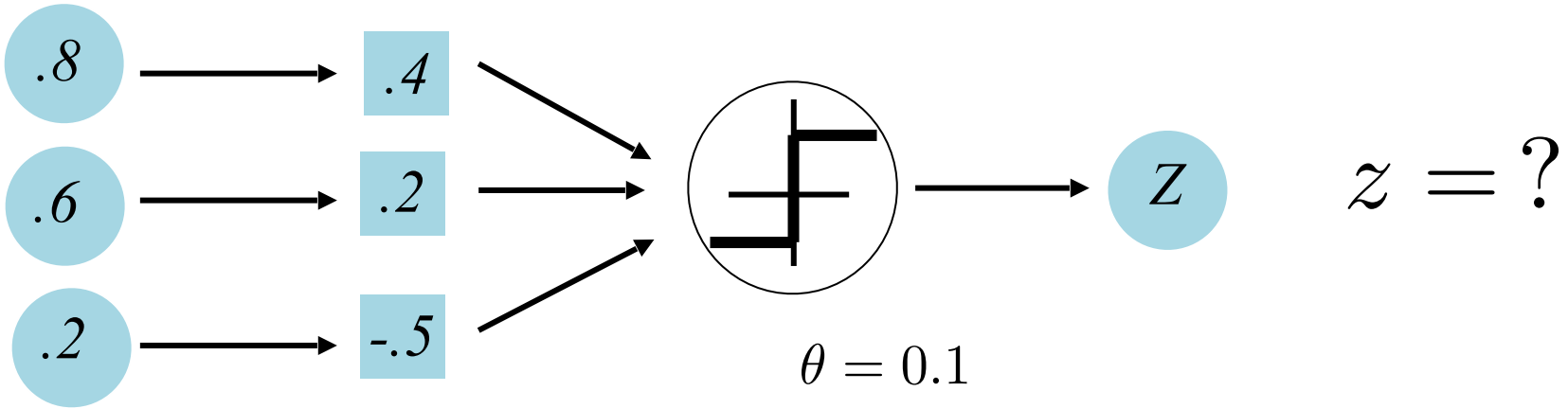




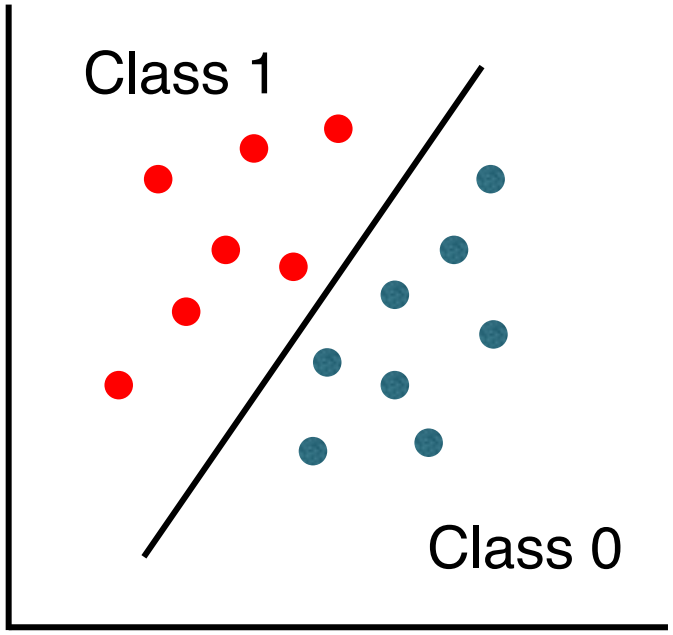
Examples



$$z = \begin{cases} 1, if\ w \cdot x > \theta \\ 0,\ w \cdot x <= \theta \end{cases}$$



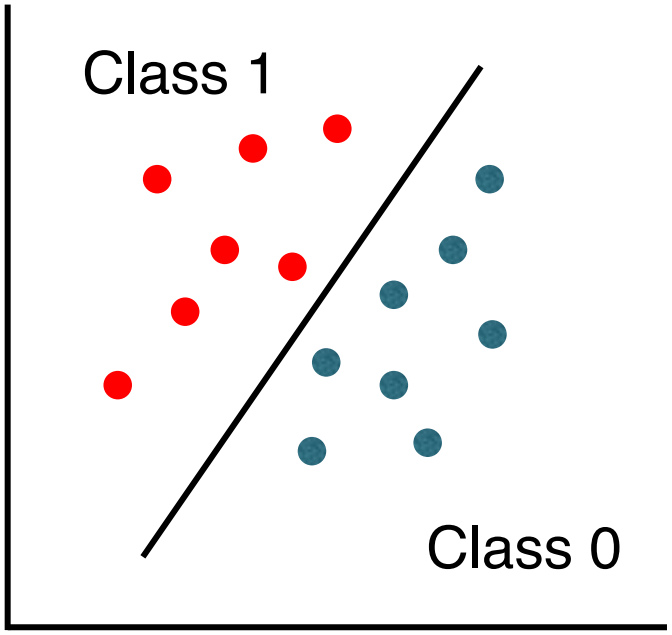
How to Learn Perceptron?



$$f(x) = \begin{matrix} 1, if\ w \cdot x + \theta > 0 \\ 0, otherwise \end{matrix}$$

w, θ are unknown parameters

How to Learn Perceptron?

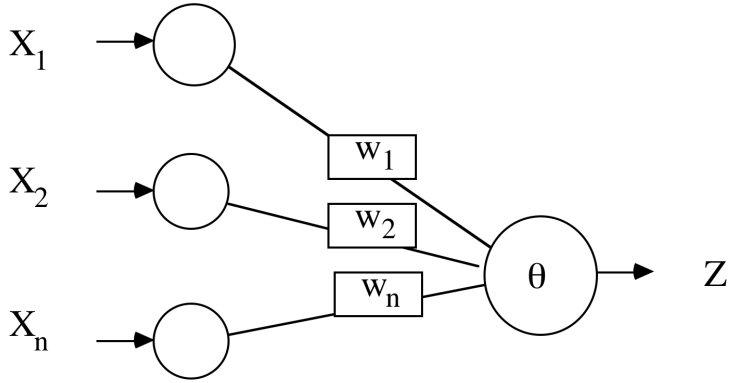


$$f(x) = \begin{matrix} 1, if\ w \cdot x + \theta > 0 \\ 0, otherwise \end{matrix}$$

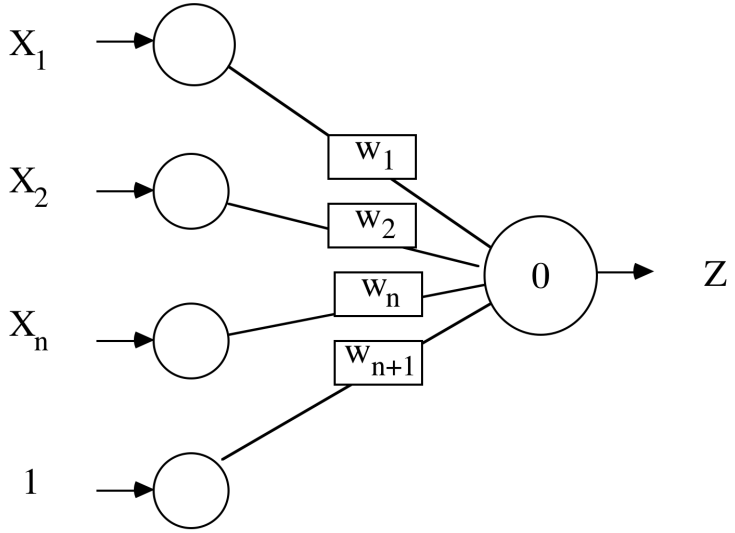
w, θ are unknown parameters

- In supervised learning the network has its output compared with known correct answers
  - Supervised learning
  - Learning with a teacher

Weight Versus Threshold



Do you need to adjust Theta? Yes, in most cases



where $w_{n+1} = -\theta$

Perceptron Learning Rules

- Consider linearly separable problems
- How to find appropriate weights
- Look if the output result *o* belongs to the desired class has the desired value *d* (give labels)

$$w^{new} = w^{old} + \triangle w \quad \triangle w = \eta \sum_i (d - o) x_i$$

η is called the learning rate, with $0 < \eta \leq 1$

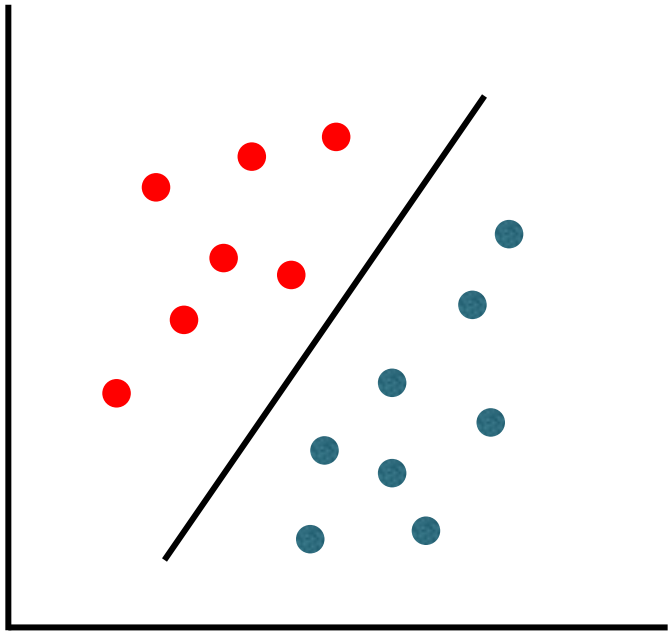
Perceptron Convergence Theorem: Guaranteed to find a solution in finite time if a solution exists

Perceptron Learning Rules

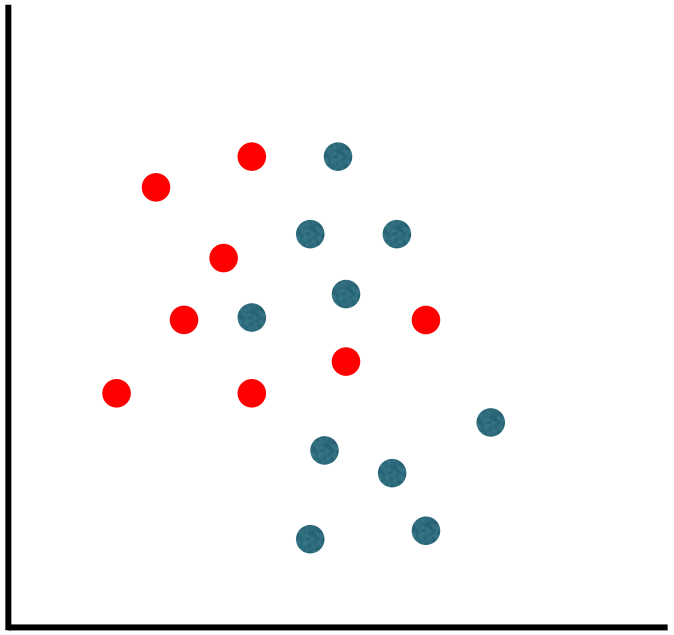
- The algorithm converges to the correct classification
  - if the training data is linearly separable

- When assigning a value to η we must keep in mind two conflicting requirements
  - Averaging of past inputs to provide stable weights estimates, which requires small η
  - Fast adaptation with respect to real changes in the underlying distribution of the process responsible for the generation of the input vector $\boldsymbol{x}$, which requires large η

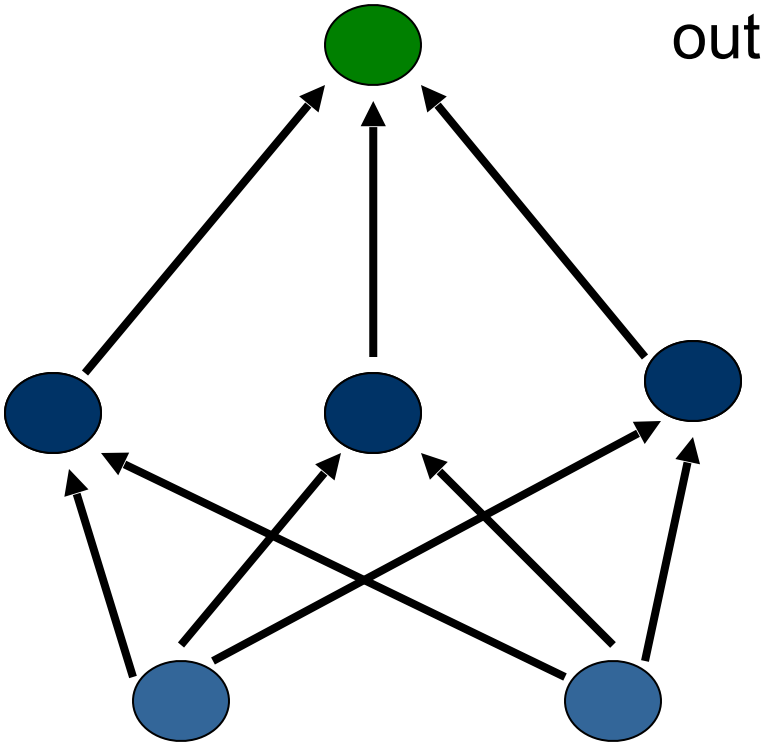
Linear Separability



Limited Functionality of Hyperplane



Multilayer Network



output layer

$$o_1 = \mathrm{sgn}(\sum_{i=0}^{n} w_{1i} x_i)$$

hidden layer

$$o_2 = \mathrm{sgn}(\sum_{i=0}^{n} w_{2i} x_i)$$

input layer