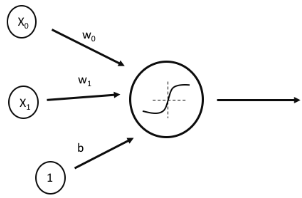
PERCEPTRON

Dr. Hilman F. Pardede

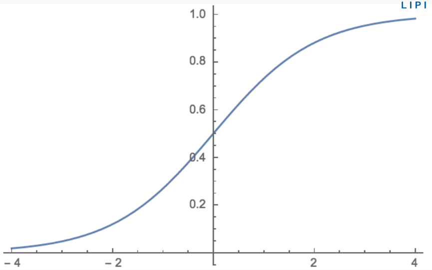
Research Center for Informatics
Indonesian Institute of Sciences

- Perceptron = a linear classifier
  - The parameters that are often called weights (w)
- real-valued constants (can be positive or negative)
- A perceptron calculates 2 quantities:
  - A weighted sum of the input features
  - This sum is then thresholded by the T(.) function



- Perceptron: a simple artificial model of human neurons
- weights = "synapses"
- threshold = "neuron firing"

- Input: by x_0, x_1, vector notation $\mathbf{x}$
- Edge parameters: $\mathbf{w}$ + bias term b



- A single neuron implements:

$$o(x;\theta) = \phi(\sum_i w_i x_i + b) = \phi(\mathbf{w}^\top \mathbf{x}) \tag{1}$$

$$\phi(\mathbf{x}) = logit(\mathbf{w}^T \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^\top \mathbf{x}}} \tag{2}$$

- Bounded from 0 to 1. Smooth, positive function.

- Output is $o_c(\mathbf{x}) = P(y = c|\mathbf{x})$
- $o_1(\mathbf{x}) = P(y = 1|\mathbf{x}) = \phi(\mathbf{w}^\top \mathbf{x}) = logit(\mathbf{w}\top \mathbf{x})$
- $o_0 = (1 - o_1)$

- Model:
  - $o_\theta = logit(\mathbf{w}^\top x + b) = \frac{1}{1+e^{-\mathbf{w}^\top \mathbf{x}+b}}$
  - $\theta : \{\mathbf{w}, b\}$
- Loss function:
  - $L(\mathbf{x}, y; \theta) = -\sum_c 1_{(y=c)} \log\ o_c = -log\ o_y$
  - Log for numerical stability and math simplicity
- Gradient of L wrt $\mathbf{w}$ and b:
  - $\frac{\partial}{\partial w} L(.)$ for both w_0 and w_1
  - $\frac{\partial}{\partial b} L(.)$

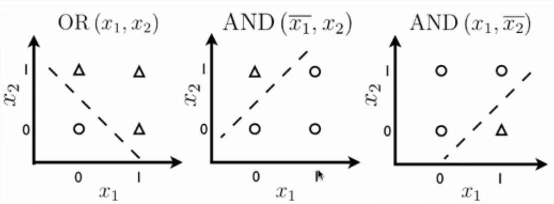
- Given a set of n examples $\{(\mathbf{x}, y)\}$.
- GD Update rule:
  - repeat until convergence

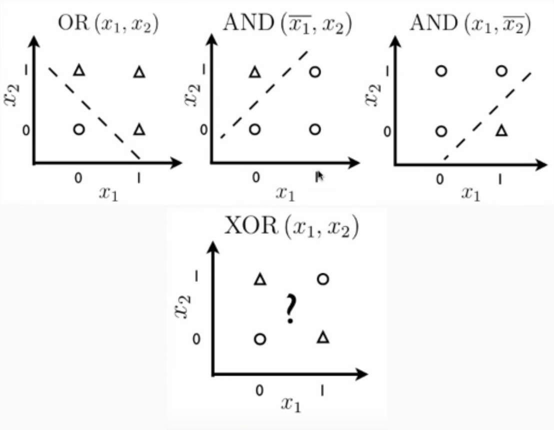
$$w \leftarrow w - \alpha \frac{\partial}{\partial w} L(\mathbf{x}, y; \mathbf{w}, b)$$

$$b \leftarrow b - \alpha \frac{\partial}{\partial b} L(\mathbf{x}, y; \mathbf{w}, b)$$

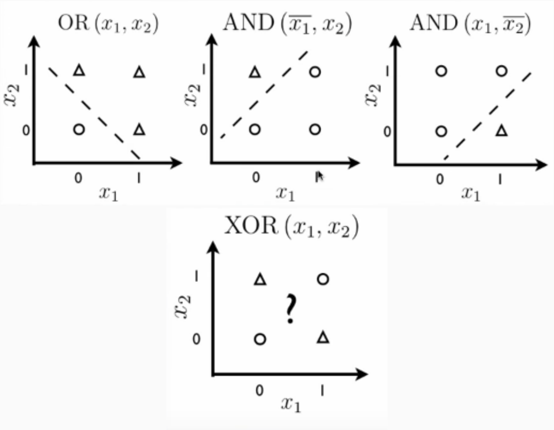
  - α - learning rate

LIMITATION OF PERCEPTRON





- it cannot solve linearly not separable data



- it cannot solve linearly not separable data
- In shallow architecture such as SVM, transforming the features into high dimensional space is one way