UNIVERSITY OF LEEDS

Class: Machine Learning

Neural Networks: Perceptron

Instructor: Matteo Leonetti

Learning outcomes

UNIVERSITY OF LEEDS

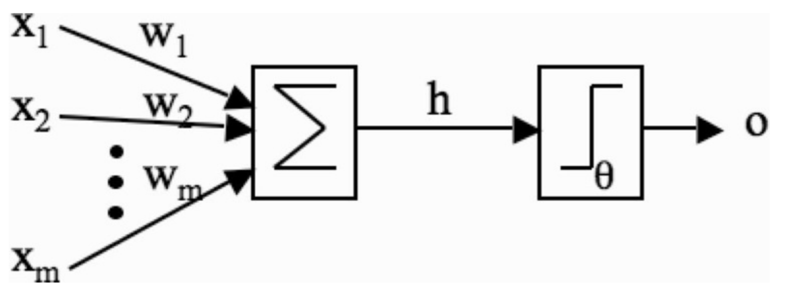
- Define linear separability.
- Justify whether a given error function is suitable for gradient descent.

Recap

We want to apply gradient descent:

$$x_{t+1} = x_t - \eta \nabla f(x_t)$$

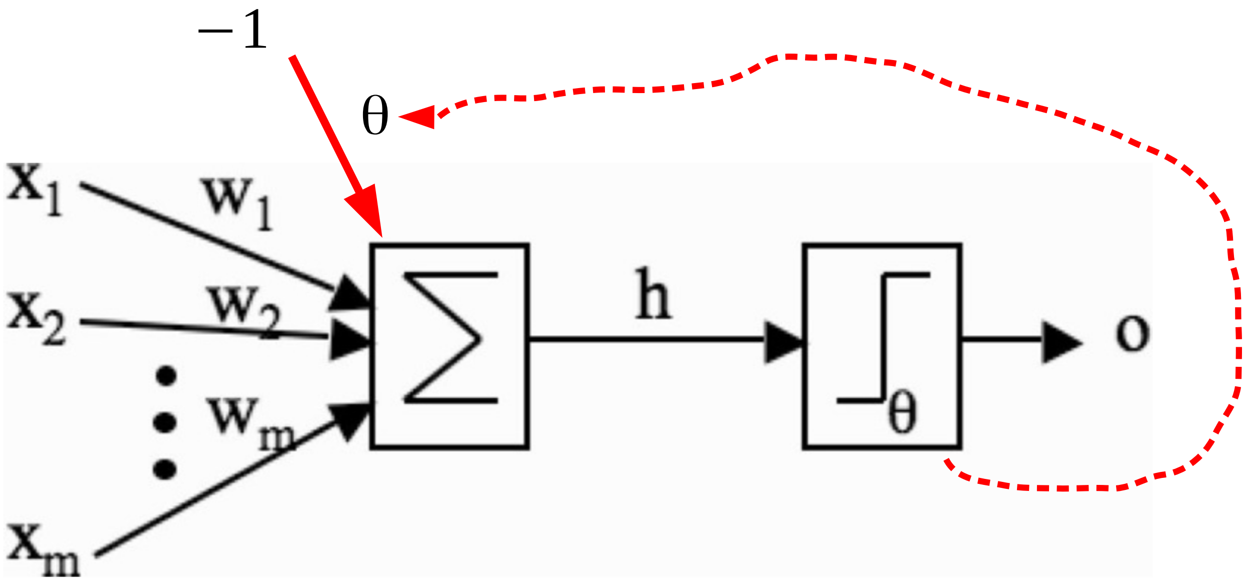
To the parameters of a perceptron:



So as to minimise an error (or loss) function, such as:

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

Bias input



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

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

$$w_1 x_1 + w_2 x_2 + \cdots + w_n x_n - 1 \cdot \theta > 0$$

$$\longrightarrow$$

$$h_w(\boldsymbol{x}) = \sum_i w_i x_i - \theta$$

$$\boldsymbol{x}_{new} = \langle \boldsymbol{x}, -1 \rangle \qquad \boldsymbol{w}_{new} = \langle \boldsymbol{w}, \theta \rangle$$

$$h_w(\boldsymbol{x}_{\boldsymbol{new}}) = \boldsymbol{w}_{\boldsymbol{new}} \cdot \boldsymbol{x}_{\boldsymbol{new}}$$

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

Question

UNIVERSITY OF LEEDS

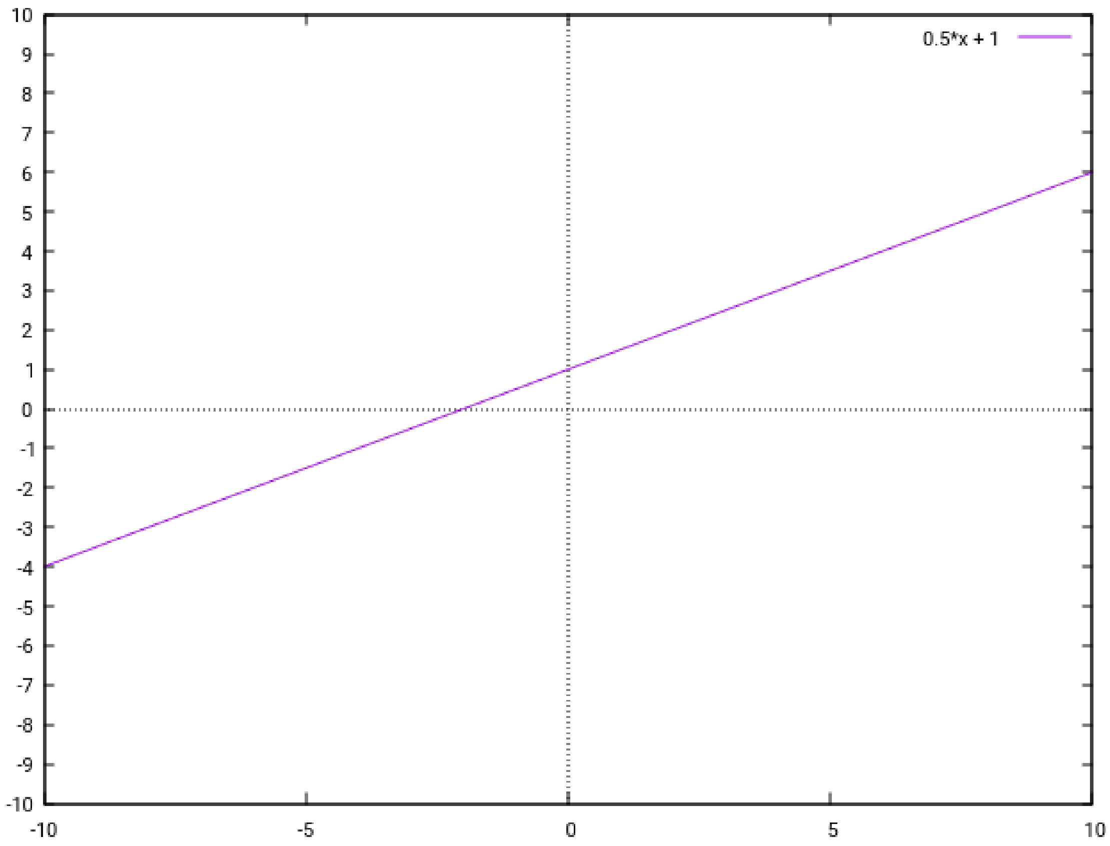
The decision boundary of the perceptron is the function below:

$$f(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} = 0$$

Plot the following function: $\frac{1}{2}x - y + 1 = 0$

Solution

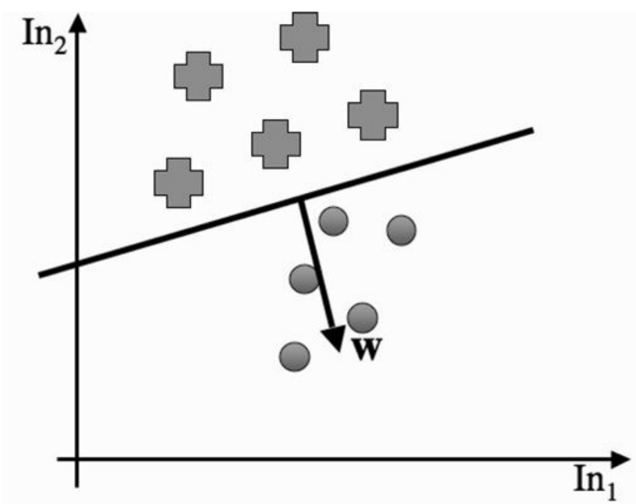
Plot the following function: $\frac{1}{2}x - y + 1 = 0$



Linear separability

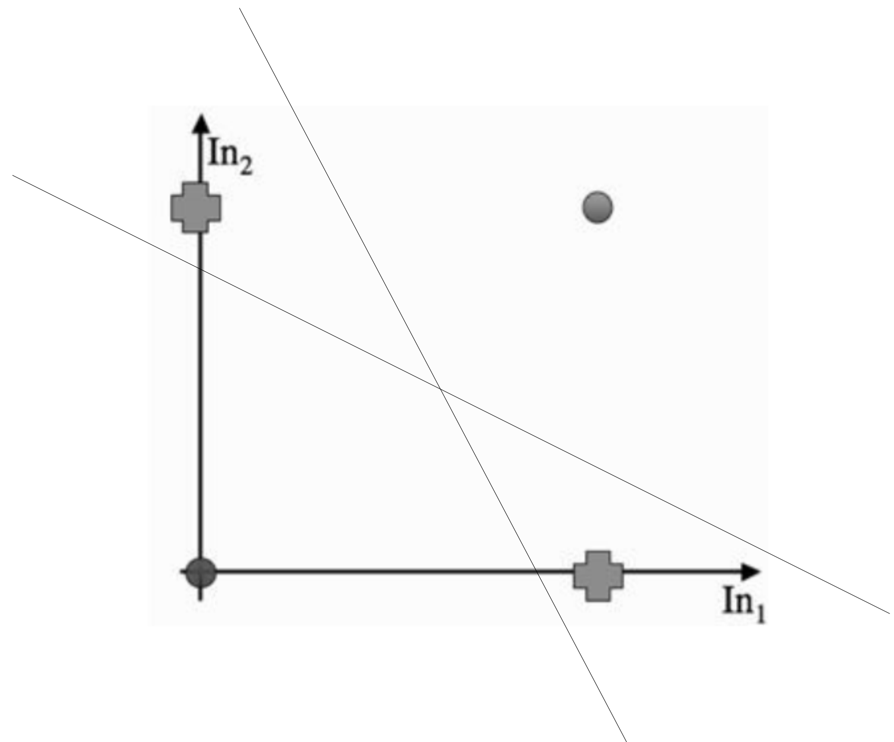
We have established that the decision boundary is a hyperplane.

$$h_w(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0$$

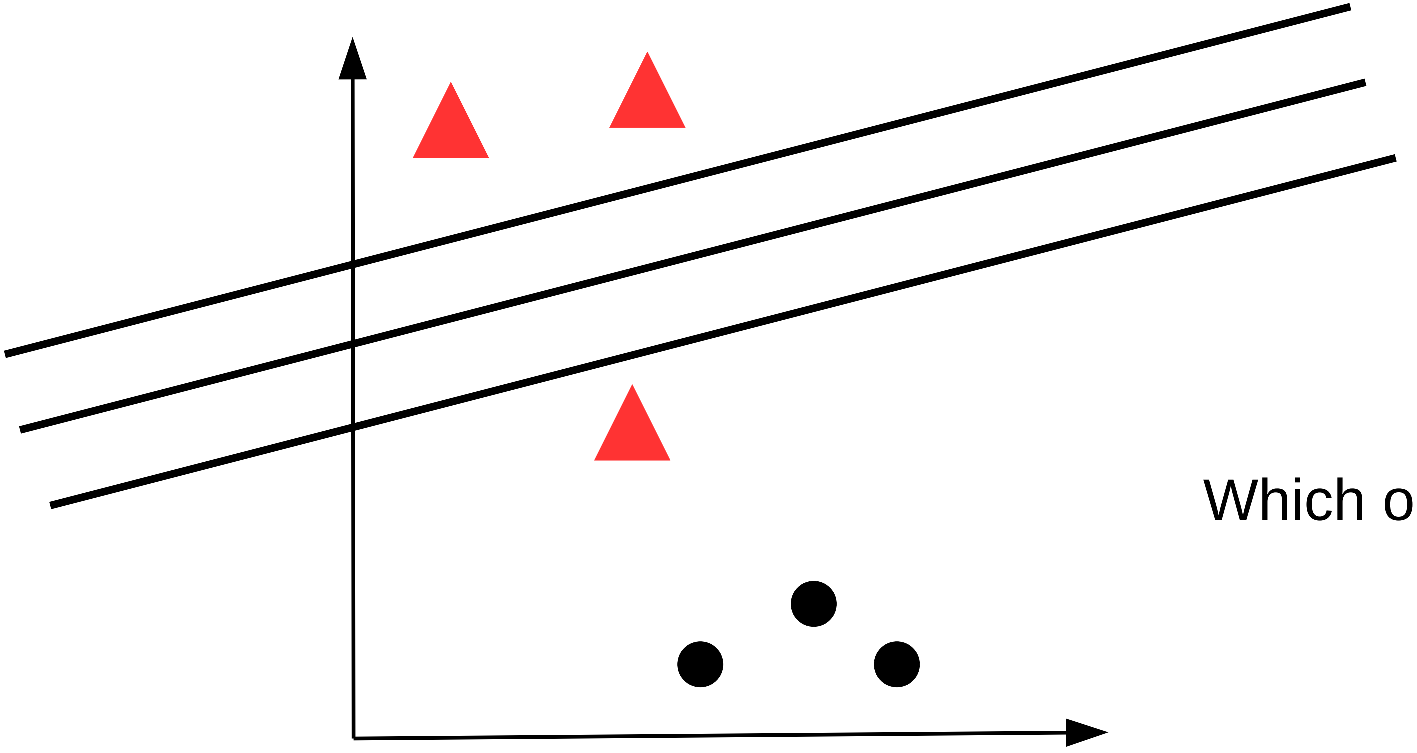


XOR

Not linearly separable!



Number of mistakes as Error



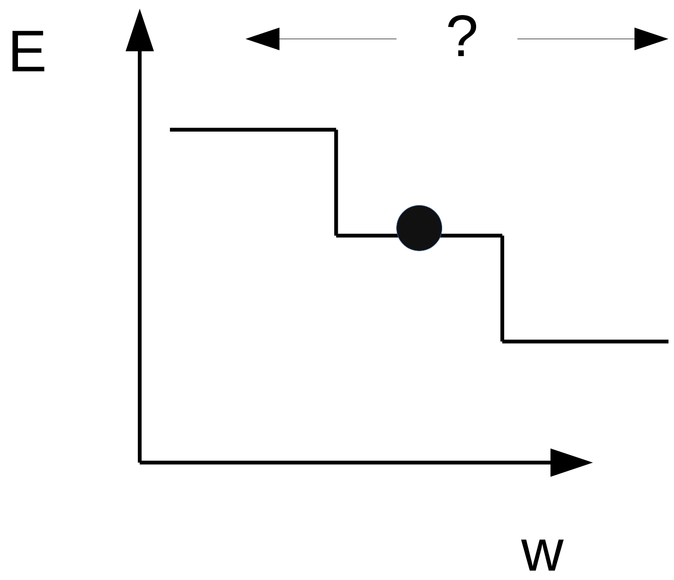
$$h_{\boldsymbol{w}}(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0$$

Which one is *better*?

Number of mistakes

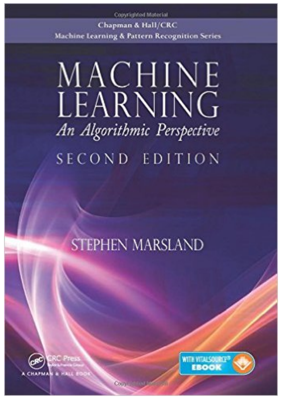
$$E(\boldsymbol{X}) = \sum_{\vec{x}_n \in \boldsymbol{X}} |y_n - t_n|$$

Number of mistakes on the dataset. Piecewise constant → no gradient.



There is no local information on the direction of improvement

Conclusion



Section 3.3 & 3.4