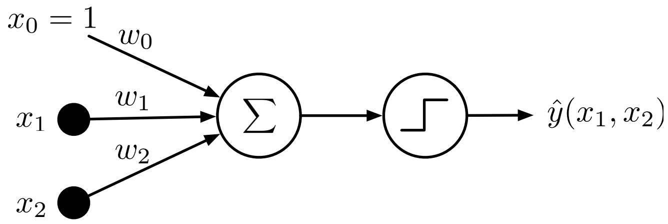
Exercise Sheet

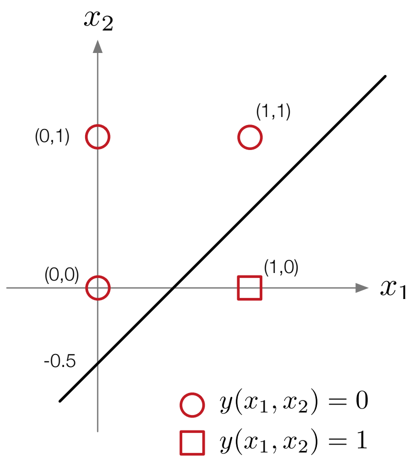
Learning Goals

- Perceptron
- Logistic Regression

1. ●○○ For a single perceptron, find an assignment to the parameters w_0, w_1, w_2 such that the perceptron implements the boolean function $y(x_1, x_2) = x_1 \land \neg x_2$ for binary variables x_1 and x_2. (Use the Heaviside step function $\varphi(x) = \max(sign(x), 0)$ as activation function.)



Solution:



1.

$w^T x + w_0 = 0$ line for all points x_1, x_2 on a line

$w^T x + w_0 > 0$ for all points x_1, x_2 in the direction of $\vec{w}$

$w^T x + w_0 < 0$ for all points $x_1 x_2$ in the opposite direction of $\vec{w}$

- From normal form to general form of line.

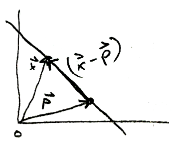
$\vec{n} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ← orientation of line $\quad \vec{p} = \begin{pmatrix} 0 \\ -0.5 \end{pmatrix}$ ← point on line

$\vec{n} \cdot (\vec{x} - \vec{p}) = 0$

$\Leftrightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix} \begin{pmatrix} x_1 - 0 \\ x_2 + 0.5 \end{pmatrix} = 0$

$\Leftrightarrow x_1 - x_2 - 0.5 = 0$

$\Leftrightarrow \underbrace{1 x_1}_{w_1} \underbrace{- 1 x_2}_{w_2} \underbrace{- 0.5}_{w_0} = 0$



1) for any two points $\vec{x}_1, \vec{x}_2$ on the line

$\vec{w}^T x_1 + w_0 = w^T x_2 + w_0 = 0$

$\Leftrightarrow \vec{w}^T (x_1 - x_2) = 0$

$\vec{w} \perp (x_1 - x_2)$

A single perceptron computes the sum $\sum w_i \cdot x_i = w_0 \cdot x_0 + w_1 \cdot x_1 + w_2 \cdot x_2$ (without applying the heaviside function). The weights of the perceptron are just the coefficients from the equation of the line: $w_0 = -0.5$, $w_1 = 1$, $w_2 = -1$. Through the heaviside function, the perceptron maps any (x_1, x_2) with $\sum_i w_i \cdot x_i > 0$ to 1 and any (x_1, x_2) with $\sum_i w_i \cdot x_i \leq 0$ to 0. (all points in the direction of $\mathbf{w}$ map to 1, points on the other side of the line to 0).

As a check, compute the output of the perceptron $\hat{y}(x_1, x_2) = \max(\text{sgn}(w_0 + w_1 x_1 + w_2 x_2), 0)$: it holds $\hat{y}(1,0) = \max(\text{sgn}(0.5) = 1,\ 0) = 1$ and $\hat{y}(0,1) = \max(\text{sgn}(-1.5) = -1,\ 0) = 0$, $\hat{y}(1,1) = \max(\text{sgn}(-0.5) = -1,\ 0) = 0$ and $\hat{y}(0,0) = \max(\text{sgn}(-0.5) = -1,\ 0) = 0$.

2. ●●○ Apply the perceptron training algorithm, as described in the lecture notes, on the following four data points to learn the parameters of a perceptron (use the heaviside step function as defined in 1).
Normally, the algorithm would pick a data point randomly, update the parameters and repeat until convergence. Here, go through the data points in the given order only once. The learning rate is $\eta = 0.4$ and the weights are initialized with $\mathbf{w} = (w_0, w_1, w_2) = (0.5, 1, -1)$. In each iteration (parameter update from a data point) list: $\mathbf{w}^T\mathbf{x}$, $\varphi(\mathbf{w}^T\mathbf{x})$, error, $\Delta\mathbf{w}$ and the new parameter vector $\mathbf{w}$.

x_1	x_2	$y(x_1, x_2)$
0	0	0
0	1	0
1	0	1
1	1	0

Solution:

We start with a learning rate $\eta = 0.4$ and initial weights $\mathbf{w}^{(0)} = (0.5, 1.0, -1.0)$.

In each iteration, we calculate:

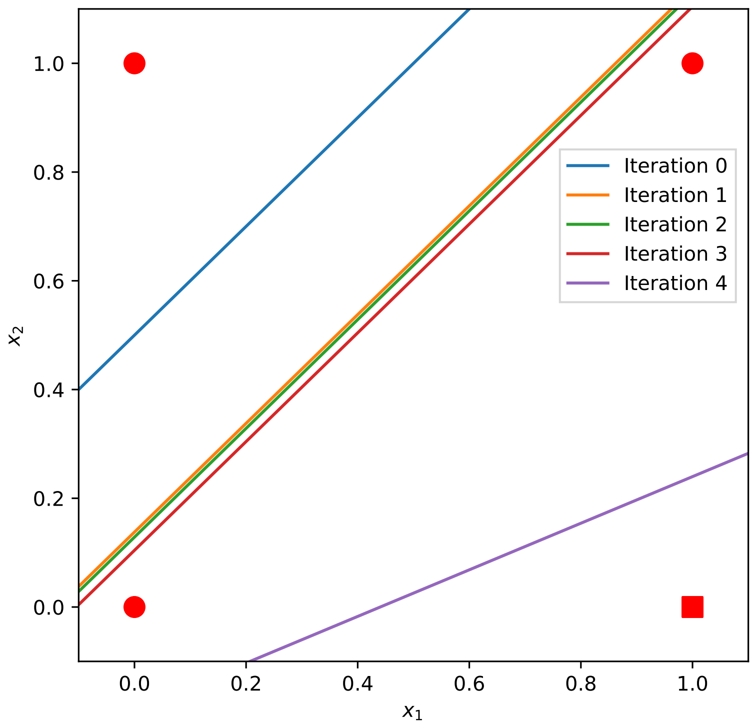
$$\text{err} = y(\mathbf{x}) - \varphi(\mathbf{w}^T\mathbf{x})$$

$$\Delta\mathbf{w} = \eta \cdot \text{err} \cdot \mathbf{x}$$

and update the parameters:

$$\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} + \Delta\mathbf{w}$$

i	x_0	x_1	x_2	$y(\mathbf{x})$	$\mathbf{w}^T\mathbf{x}$	$\varphi(\mathbf{w}^T\mathbf{x})$	err	Δw_0	Δw_1	Δw_2	w_0	w_1	w_2
0											0.5	1.0	−1.0
1	1	0	0	0	0.5	1.0	−1.0	−0.4	0.0	0.0	0.1	1.0	−1.0
2	1	0	1	0	−0.9	0.0	0.0	0.0	0.0	0.0	0.1	1.0	−1.0
3	1	1	0	1	1.1	1.0	0.0	0.0	0.0	0.0	0.1	1.0	−1.0
4	1	1	1	0	0.1	1.0	−1.0	−0.4	−0.4	−0.4	−0.3	0.6	−1.4



3. Implement a Logistic Regression Classifier. The jupyter notebook attached to this exercise sheet contains a skeleton implementation of a LogReg Classifier. Complete the implementations of the functions `fit(...)` and `predict(...)`.