Mathematics of Learning – Worksheet 10

- The exercise sheets will be uploaded every Monday. Solution sketches will be uploaded one week later.
- You can hand in your own solutions via StudOn and we correct them - this is not mandatory. Please hand in small groups of 2-3 students.
- For questions, please use the forum on StudOn since other students may have similar questions. If you have a more personal question about the exercises please send an email to ehsan.waiezi@fau.de or lars.weidner@fau.de respectively.

Exercise 1 [Descent directions, gradients, niveau lines]
Consider a function $f\colon \mathbb{R}^n \to \mathbb{R}$, to be continuously differentiable. We define a descent direction for f at a point $\tilde{x} \in \mathbb{R}^n$ as all vectors $\theta \in \mathbb{R}^n$, such that $\exists \rho > 0$ such that $f(\tilde{x} + \lambda\theta) < f(\tilde{x})$ for all $0 < \lambda < \rho$ (*"if we make a tiny (ρ) or an even smaller (λ) step in that direction (θ), the value of f gets smaller"*). Prove or disprove:

a) The set of descent directions for f at a point $\tilde{x}$ with $\nabla f(\tilde{x}) \neq 0$ is the set $\{\theta \in \mathbb{R}^n : \langle \nabla f(\tilde{x}), \theta \rangle < 0\}$.

b) For $\nabla f(\tilde{x}) \neq 0$, the minimization problem

$$\min_{\theta \in \mathbb{R}^n : ||\theta||_2 = 1} \lim_{\lambda \to 0} \frac{f(\tilde{x} + \lambda\theta) - f(\tilde{x})}{\lambda}$$

is solved by $-\frac{\nabla f(\tilde{x})}{||\nabla f(\tilde{x})||_2}$.

Exercise 2 [Derivative of logistic activation function].
Let $\psi\colon \mathbb{R} \to \mathbb{R}$ the logistic activation function for a perceptron defined as

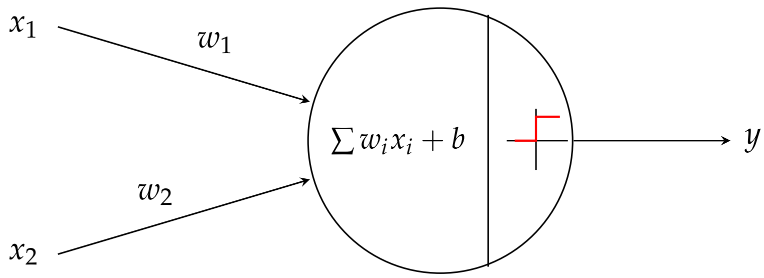
$$\psi(t) := \frac{1}{1 + e^{-t}}$$

Show that the derivative ψ' of ψ can be computed as:

$$\psi'(t) = \psi(t)(1 - \psi(t)).$$

Exercise 3 [Implementation of perceptrons for binary logic functions].

Let $f_\theta\colon \mathbb{R}^2 \to \mathbb{R}$ be a parametrized map realized by the following binary perceptron that maps two inputs $\vec{x} = (x_1, x_2)$ to an output y:



Here, $\theta \in \mathbb{R}^3$ is the vector of free parameters with $\theta := (w_1, w_2, b)$, where w_1, w_2 are the weights of the respective inputs and b is the bias of the perceptron. We assume that the activation function of the perceptron is the *Heavyside step function* $H\colon \mathbb{R} \to \{0, 1\}$ defined as :

$$H(x) \;:=\; \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x \geq 0. \end{cases}$$

Implement a Python function `perceptron(x, theta)` that return a output `y`, which is either 0 or 1. Use this function to implement a family of perceptrons, which realize the following binary logic functions:

AND			OR			XOR			NAND			NOR		
x_1	x_2	y	x_1	x_2	y	x_1	x_2	y	x_1	x_2	y	x_1	x_2	y
0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
1	0	0	1	0	1	1	0	1	1	0	1	1	0	0
0	1	0	0	1	1	0	1	1	0	1	1	0	1	0
1	1	1	1	1	1	1	1	0	1	1	0	1	1	0

Hint: One of the binary logic functions cannot be realized by a simple perceptron. Explain the reasons for this and suggest an alternative realization.