Authors:

- Mhd Jawad Al Rahwanji - 7038980 - mhal00002@stud.uni-saarland.de
- Christian Singer - 7039059 - chsi00002@stud.uni-saarland.de

Exercise 4.3 - Gradient Descent

a) We have the following:

$$f(\mathrm{x}) = f(x_1, x_2) = \frac{1}{1 + exp(-(x_1^2 + x_2^2))},$$

$$\epsilon = 0.2, \text{ and}$$

$$\nabla f(\mathrm{x}) = \nabla f(x_1, x_2) = \left(\frac{2x_1 exp(-(x_1^2 + x_2^2))}{(1 + exp(-(x_1^2 + x_2^2)))^2}, \frac{2x_2 exp(-(x_1^2 + x_2^2))}{(1 + exp(-(x_1^2 + x_2^2)))^2}\right)$$

The gradient descent iterations:

Iteration	**x**	***f*(x)**	***f*(x)**
0	(1, -1)	**0.88**	(0.21, -0.21)
1	(0.958, -0.958)	0.86	(0.22, -0.22)
2	(0.914, -0.914)	0.84	(0.24, -0.24)
3	(0.866, -0.866)	**0.81**	(0.25, -0.25)

Note. We notice $f(\mathrm{x}_0) > f(\mathrm{x}_3)$

A higher ϵ would lead to faster convergence but at a small risk of slightly overshooting. Whereas, a much larger ϵ, say > 1, might lead to never finding the minimum.

After 21 iterations of gradient descent, an $\hat{x}$ was achieved where for 3 consecutive iterations the value of $f(\mathrm{x})$ exhibited no change > 1e-2. A ϵ equal to 0.2 was used to avoid overshooting and slowly but surely reaching the minimum.

$$\hat{x} = (0.162, \; -0.162)$$

Our $\hat{x}$ is not the global minimum because we stopped early but had we continued a $\hat{x}$ = (0, 0) is achievable using gradient descent. The function is convex, therefore, it has a guaranteed global minimum at (0, 0) where $f((0, 0))$ = 0.5.