Deep Learning in Computer Vision

Exercise no. 1

Solution by Fabian Späh, Minchen Zhu

Exercise 1.3 (10 + 10 pts)

a

Given $E(\theta) = 2\theta_1^2 + 4\theta_2 + max(0, \theta_2 + \theta_3), \quad \theta^{[0]} = [2, 1, 0]^T, \quad \tau = 0.5$

Perform 2 steps of Gradient Descent:

$$\nabla E(\theta) = \begin{bmatrix} \partial_1 E(\theta) \\ \partial_2 E(\theta) \\ \partial_3 E(\theta) \end{bmatrix} = \begin{bmatrix} 4\theta_1 \\ 4 + \begin{cases} 1 & \text{, if } \theta_2 + \theta_3 > 0 \\ 0 & \text{, if } \theta_2 + \theta_3 < 0 \end{cases} \\ \begin{cases} 1 & \text{, if } \theta_2 + \theta_3 > 0 \\ 0 & \text{, if } \theta_2 + \theta_3 < 0 \end{cases} \end{bmatrix}$$

Step 1:

$$\theta^{[1]} = \theta^{[0]} - \tau \nabla E(\theta^{[0]}) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 4\theta_1^{[0]} \\ 4+1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1.5 \\ -0.5 \end{bmatrix},$$

$E(\theta^{[1]}) = 2(-2)^2 + 4(-1.5) + 0 = 2$

Step 2:

$$\theta^{[2]} = \theta^{[1]} - \tau \nabla E(\theta^{[1]}) = \begin{bmatrix} -2 \\ -1.5 \\ -0.5 \end{bmatrix} - 0.5 \begin{bmatrix} 4\theta_1^{[1]} \\ 4+0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ -3.5 \\ -0.5 \end{bmatrix},$$

$E(\theta^{[2]}) = 2(2)^2 + 4(-3.5) + 0 = -6$

b

If the gradient of f exists, it holds for every τ, q that

$$f(\mathbf{p} + \tau q) = f(\mathbf{p}) + \tau \nabla f(\mathbf{p})^T \cdot q + r(\tau)$$

with $\lim_{\tau \to 0} \frac{r(\tau)}{\tau} = 0$. Hence we get for h

$$\begin{aligned} & h(\mathbf{p} + \tau q) = g(f(\mathbf{p} + \tau q) = g(f(\mathbf{p}) + \tau \nabla f(\mathbf{p})^T \cdot q + r(\tau)) \\ = \; & g(f(\mathbf{p}) + \tau(\nabla f(\mathbf{p})^T \cdot q + \tfrac{r(\tau)}{\tau})) \\ = \; & g(f(\mathbf{p})) + \tau g'(f(\mathbf{p})) \cdot (\nabla f(\mathbf{p})^T \cdot q + \tfrac{r(\tau)}{\tau}) + r'(\tau) \\ = \; & h(\mathbf{p}) + \tau g'(f(\mathbf{p})) \cdot \nabla f(\mathbf{p})^T \cdot q + \tau g'(f(\mathbf{p})) \tfrac{r(\tau)}{\tau} + r'(\tau) \; . \end{aligned}$$

It clearly is

$$\lim_{\tau \to 0} \frac{\tau g'(f(\mathbf{p})) \frac{r(\tau)}{\tau} + r'(\tau)}{\tau} = \lim_{\tau \to 0} g'(f(\mathbf{p})) \cdot \frac{r(\tau)}{\tau} + \frac{r'(\tau)}{\tau} = 0$$

and thus

$$h(\mathbf{p} + \tau q) = h(\mathbf{p}) + \tau \nabla h(\mathbf{p})^T \cdot q + r''(\tau)$$

with $r''(\tau) = \tau g'(f(\mathbf{p})) \frac{r(\tau)}{\tau} + r'(\tau)$ and $\nabla h(\mathbf{p}) = g'(f(\mathbf{p})) \cdot \nabla f(\mathbf{p})$ which was to show.