Theoretical part
1) 8 / 8
2) 7 / 7
Total 15 / 15
Practical part
1) 5 / 5
2) 2 / 5
3) 4 / 5
Total 11 / 15

Deep Learning for Visual Recognition - Assignment 3

Mayara E. Bonani, Guillaume Rouvarel, Arash Safavi, Vardeep Singh, Cüneyt Erem

December 7, 2020

Theoretical Part

a) Fundamentals of Unconstrained Optimization

1.

$f(x) = 100(x_2 - x_1^2)^2 - (1 - x_1)^2$

* Gradient $\nabla f(x)$:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}$$ ✓

* Hessian $H(f(x))$:

$$H(f(x)) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$ ✓

2.

* Critical Point $\rightarrow$ Gradient $\nabla f(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\nabla f(x) = \begin{pmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Solving the equation above, we obtain:

$$200(x_2 - x_1^2) = 0 \rightarrow x_2 = x_1^2$$
$$-400x_1(x_2 - x_1^2) - 2(1 - x_1) = -400x_1(x_2 - x_2) - 2(1 - x_1) = 0$$
$$1 - x_1 = 0 \rightarrow x_1 = 1 \text{ and } x_2 = 1$$ ✓

$H(f(x))$ is a symmetric 2×2 matrix. If the $h_{1,1}$ of the matrix and the determinant of the matrix are both greater than zero, then the hessian matrix is positive definite and f has a relative minimum at the point.

* We show that the Hessian matrix the point $(1,1)^T$ is positive definite.

$$H(f(x)) = \begin{pmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix} = \begin{pmatrix} -400 + 1200 + 2 & -400 \\ -400 & 200 \end{pmatrix} = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix}$$

$$det(H(f(x))) = \begin{vmatrix} 802 & -400 \\ -400 & 200 \end{vmatrix} = 400 > 0 \quad \checkmark$$

$$802 > 0 \quad \checkmark$$

3.

$f(x) = 8x_1 + 12x_2 + x_1^2 + 2x_2^2$

* Critical Point $\rightarrow$ Gradient $\nabla f(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 8 + 2x_1 \\ 12 - 4x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \checkmark$$

Thefore we obtain that the point $x_1 = -4$ and $x_2 = 3$ is the only stationary point. $\checkmark$

To prove that it is a saddle point, we show that the determinant of the hessian matrix is smaller than zero.

* Hessian $H(f(x))$:

$$det(H(f(x))) = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix} = \begin{vmatrix} 2 & 0 \\ 0 & -4 \end{vmatrix} = -8 < 0 \quad \checkmark$$

b) Case Study

a) The error surface between the times corresponding to iteration 100^{th} and iteration $10,000^{th}$ looks like linear by being stuck in a flat plateau of the training error. $\checkmark$

b) I would choose the property of receiving a higher weight when eigenvectors have a smaller value with respect to the directional second derivative. In our case the learning would be accelerated and the processing time shortened. $\checkmark$