Introduction to Optimization

Week 3 Assignment : Solution

Part I [*10marks*]

1. Let's consider the function f defined over $\mathbb{R}^2$ by $f(x,y) = \sqrt{x^2+y^2+\pi}$

   a. Explain (simply) why $\sqrt{\pi}$ is the minimum of this function.[*1mark*]

   Since x^2 and y^2 are always non-negative, the smallest possible value of x^2+y^2 is 0, which occurs when $x=0$ and $y=0$. Therefore, the minimum value of $f(x,y)$ is achieved when $x=0$ and $y=0$, resulting in

   $$f(0,0) = \sqrt{0^2+0^2+\pi} = \sqrt{\pi}.$$

   b. Compute the gradients (partial derivatives) of f with respect to x and y.[*2marks*]

   Recall that for a function $u(x)$,

   $$\frac{d}{dx}\sqrt{u(x)} = \frac{1}{2\sqrt{u(x)}}\frac{du}{dx}(x) = \frac{u'(x)}{2\sqrt{u(x)}}$$

   Therefore, the partial derivatives of f with respect to x and y are:

   $$\frac{\partial f}{\partial x} = \frac{2x}{2\sqrt{x^2+y^2+\pi}} = \frac{x}{\sqrt{x^2+y^2+\pi}} \tag{1}$$

   and

   $$\frac{\partial f}{\partial y} = \frac{2y}{2\sqrt{x^2+y^2+\pi}} = \frac{y}{\sqrt{x^2+y^2+\pi}} \tag{2}$$

   c. Find the pair (x,y) for which f is minimum and deduce that the minimal value of $f(x,y)$ is indeed $\sqrt{\pi}$.[*2marks*]

   To find the minimum of f, we set the partial derivatives equal to 0:

   $$\frac{\partial f}{\partial x} = 0 \quad \text{and} \quad \frac{\partial f}{\partial y} = 0.$$

   Solving these equations using (1) and (2), we find that $x=0$ and $y=0$. Therefore, the minimal value of $f(x,y)$ is $\sqrt{\pi}$, as shown earlier.

2. **Questions :**

   a. Rewrite the cost function using a matrix form.[*1mark*]
   Using

   $$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad \omega = \begin{bmatrix} \omega_0 \\ \omega_1 \end{bmatrix}.$$

   where $\mathbf{y}$ is the vector of outputs, $\mathbf{X}$ is the design matrix, and ω is the parameter vector, the cost function $L(x;\omega)$ can be rewritten in matrix form as:

   $$L(x;\omega) = \frac{1}{m}\|\mathbf{y} - \mathbf{X}\omega\|_1,$$

b. Compute the gradients with respect to ω_0 and ω_1.[*4marks*]

To compute the gradients with respect to ω_0 and ω_1, we keep the original form of the cost function $L(x;\omega)$ using the definition of the 1-norm:

$$L(x;\omega) = \frac{1}{m}\sum_{i=1}^{m} |y_i - \omega_0 - \omega_1 x_i|.$$

Now, let's compute the gradients:
Using $\frac{d}{dx}|u| = \frac{u}{|u|}\frac{du}{dx}$, we have for ω_0:

$$\frac{\partial L}{\partial \omega_0} = \frac{1}{m}\sum_{i=1}^{m} \frac{y_i - \omega_0 - \omega_1 x_i}{|y_i - \omega_0 - \omega_1 x_i|} \cdot \frac{\partial}{\partial \omega_0}(y_i - \omega_0 - \omega_1 x_i).$$

Moreover,

$$\frac{\partial}{\partial \omega_0}(y_i - \omega_0 - \omega_1 x_i) = -1$$

Thus,

$$\frac{\partial L}{\partial \omega_0} = \frac{-1}{m}\sum_{i=1}^{m} \frac{y_i - \omega_0 - \omega_1 x_i}{|y_i - \omega_0 - \omega_1 x_i|}$$

Similarly, for ω_1:

$$\frac{\partial L}{\partial \omega_1} = \frac{1}{m}\sum_{i=1}^{m} \frac{y_i - \omega_0 - \omega_1 x_i}{|y_i - \omega_0 - \omega_1 x_i|} \cdot \frac{\partial}{\partial \omega_1}(y_i - \omega_0 - \omega_1 x_i).$$

And,

$$\frac{\partial}{\partial \omega_1}(y_i - \omega_0 - \omega_1 x_i) = -x_i$$

Therefore,

$$\frac{\partial L}{\partial \omega_1} = \frac{-1}{m}\sum_{i=1}^{m} \frac{x_i(y_i - \omega_0 - \omega_1 x_i)}{|y_i - \omega_0 - \omega_1 x_i|}$$

Part II- Python Implementation. [*10marks*]

Find the solution in the Jupiter notebook.

3. We will see if we can find the results of question 1 using the gradient descent.

   a. Use the results of question 1.b) to find the couple (x, y) that minimizes $f(x, y)$, this time, with the gradient descent.[*3marks*]

   b. Plot the values of x and y during the iterative process.[*1mark*]

   c. Compute $f(x_{final}, y_{final})$ and compare with the result of question 1.[*1mark*]

4. Implementation of the question 2. You will use the dataset and the jupyter notebook provided with the assignment.

   a. Read and visualize the dataset with a scatter plot.[*0.5marks*]

   b. Use the results of question 2 to implement the gradient descent and minimize the cost function defined.[*3marks*]

   c. Print the values of ω_0 and ω_1 at the end of the optimization.[*0.5marks*]

   d. Plot the evolution of the error during the iterations.[*1mark*]