Homework 2

Q1
10 Points

In typical gradient descent, we take steps using a constant step size η, so that:

$$\theta^{t+1} = \theta^t - \eta \nabla_\theta f\left(\theta^t\right).$$

In the following, assume that f is an arbitrary differentiable function. For very very **small** η, what will generally be true?
(a) $f\left(\theta^t\right) \geq f\left(\theta^{t+1}\right)$
(b) $f\left(\theta^t\right) \leq f$
(c) cannot say

Describe conc

Save Answer

Q2
10 Points

In typical gradient descent, we take steps using a constant step size η, so that:

$$\theta^{t+1} = \theta^t - \eta \nabla_\theta f\left(\theta^t\right).$$

In the following, assume that f is an arbitrary differentiable function. For very very **big** η, what will generally be true?
(a) $f\left(\theta^t\right) \geq f\left(\theta^{t+1}\right)$
(b) $f\left(\theta^t\right) \leq f\left(\theta^{t+1}\right)$
(c) cannot say

Describe concisely your reasoning behind the choice.

Save Answer

Q3
10 Points

In typical gradient descent, we take steps using a constant step size η, so that:

$$\theta^{t+1} = \theta^t - \eta \nabla_\theta f\left(\theta^t\right).$$

In the following, assume that f is an arbitrary differentiable function. Grady would like to pick a perfect step size on every step and proposes a new update rule that selects η^* to be the value of step-size η that decreases the objective as much as possible in the direction $\nabla_\theta f(\theta)$ and then uses η^* as the step size:

$$\eta^* = \arg\min_\eta f\left(\theta^t - \eta \nabla_\theta f\left(\theta^t\right)\right)$$
$$\theta^{t+1} = \theta^t - \eta^* \nabla_\theta f\left(\theta^t\right)$$

For Grady's ru

(a) $f\left(\theta^t\right) \geq f$

(b) $f\left(\theta^t\right) \leq f$

(c) cannot say

Describe conc

Save Answer

Q4
10 Points

Robots want to find a location for a meeting that minimizes the sum of squared distances from the rooms of a group of friends (assume everyone is in their room) to the location of the meeting. Assume for now that they can host the gathering at any location in the hallway. Assuming that the robots live in a 1-dimensional hallway, pose this problem as an (unconstrained) optimization problem. Assume there are n robots (1 robot per friend) and the i-th friend is located at location l_i. Denote the location of the meeting by p. What is the objective as a function of p ? Write it down.

Hint: If you are having trouble deriving the general case, try the $n = 2$ case first.

Save Answer

Q5
10 Points

Robots want t [illegible] the sum of squared dista [illegible] ume everyone is in [illegible] ume for now that they [illegible] allway. Assuming that the robots live in a 1-dimensional hallway, determine the gradient (write down/show your computation). Use it to find the optimal location p^{opt} for the meeting by setting the gradient equal to zero and solving.

Save Answer

Q6

10 Points

Robots want to find a location for a meeting that minimizes the sum of squared distances from the rooms of a group of friends (assume everyone is in their room) to the location of the meeting. Assume for now that they can host the gathering at any location in the hallway. Assuming that the robots live in a 1-dimensional hallway, mark all of the following that are True / False for this particular objective function (i.e., the objective function you derived in Q4). Provide the reasoning behind your choice:

(a) There is necessarily a unique location that minimizes the objective function.

(b) The optimization problem may have local minima that are not global minima.

(c) The optima[illegible] the rooms.

(d) There is ne[illegible]ent descent conve[illegible]

Save Answer

Q7
10 Points

Let's consider gradient descent when the dimension of the input is 2.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

So, we have $f(\theta)$ and we are trying to find the values of θ_1 and θ_2 that minimize it. Suppose

$$f(\theta) = -3\theta_1 - \theta_1\theta_2 + 2\theta_2 + \theta_1^2 + \theta_2^2$$

If we started at $\theta = (1, 1)$ and took a step of gradient descent with step-size 0.1, what would the next value of θ be?
Enter a tuple (No need to show any calculation).

Save Answer

Q8
10 Points

Let's consider ... nput is 2.

So, we have $f(\theta)$ and we are trying to find the values of θ_1 and θ_2 that minimize it. Suppose

$$f(\theta) = -3\theta_1 - \theta_1\theta_2 + 2\theta_2 + \theta_1^2 + \theta_2^2$$

What is $f([1.2, 0.7])$?
Enter a numerical value.

Save Answer

Q9

10 Points

Let's consider gradient descent when the dimension of the input is 2.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

So, we have $f(\theta)$ and we are trying to find the values of θ_1 and θ_2 that minimize it. Suppose

$$f(\theta) = -3\theta_1 - \theta_1\theta_2 + 2\theta_2 + \theta_1^2 + \theta_2^2$$

If we started at $\theta = (1, 1)$ and took a step of gradient descent with step-size 1.0, what would the next value of θ be?
Enter a tuple (No need to show any calculation).

Save Answer

Q10

10 Points

Let's consider ... nput is 2.

So, we have $f(\theta)$ and we are trying to find the values of θ_1 and θ_2 that minimize it. Suppose

$$f(\theta) = -3\theta_1 - \theta_1\theta_2 + 2\theta_2 + \theta_1^2 + \theta_2^2$$

What is $f([3, -2])$?
Enter a numerical value.

Save Answer