Course › Unit 0. Course Overview, Homework 0, Project 0 (1 week) › Homework 0 › 9. Gradients and Optimization

Audit Access Expires May 11, 2020
You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

9. Gradients and Optimization

Multivariable Calculus Review: Simple Gradient

1.0/1 point (graded)

Let

$$\begin{array}{rccl} f: & \mathbb{R}^d & \to & \mathbb{R} \\ & \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix} & \mapsto & f(\theta). \end{array}$$

denote a **differentiable** function. The **gradient** of f is the vector-valued function

$$\begin{array}{rccl} \nabla_\theta f: & \mathbb{R}^d & \to & \mathbb{R}^d \end{array}$$

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_d \end{pmatrix} \mapsto \left.\begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \vdots \\ \frac{\partial f}{\partial \theta_d} \end{pmatrix}\right|_{\theta} .$$

Consider

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

Compute the gradient ∇f.

(Enter your answer as a vector, e.g., type **[2,x]** for the vector $\begin{pmatrix} 2 \\ x \end{pmatrix}$. Note the square brackets, and commas as separators. Enter **theta_i** for θ_i.)

$\nabla_\theta f(\theta) =$ 2*[theta_1, theta_2]

✓ **Answer:** [2*theta_1,2*theta_2]

STANDARD NOTATION

Solution:

$$\begin{aligned} f(\theta) &= \theta_1^2 + \theta_2^2 \\ \nabla f(\theta) &= \left.\begin{pmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \end{pmatrix}\right|_{\theta} = \begin{pmatrix} 2\theta_1 \\ 2\theta_2 \end{pmatrix} . \end{aligned}$$

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

Geometric Picture of the Function

3/3 points (graded)

As above, consider $f(\theta) = \theta_1^2 + \theta_2^2$. Let us visualize $f(\theta)$ as a surface on the (θ_1, θ_2)-plane. We will use the usual horizonal plane as the (θ_1, θ_2)-plane, and the vertical axis as the $f(\theta)$-axis.

Consider the level curves $\theta_1^2 + \theta_2^2 = K$ where $K > 0$ is some fixed real number.

What is the shapes of such a curve?

○ parabola

● circle

○ hyperbola

○ line

✓

Consider how the level curves $\theta_1^2 + \theta_2^2 = +K$ change as K increases from 0 to ∞. Does the graph (surface) of $f(\theta)$ have a global maximum, or global minimum, or neither?

○ global maximum

● global minimum

○ neither

✓

At each point $\theta = (\theta_1, \theta_2)$ in the (θ_1, θ_2)-plane, $f(\theta)$ decreases in the direction of...

○ $\nabla_\theta f(\theta)$

● $-\nabla_\theta f(\theta)$

✓

Solution:

The graph of $f(\theta)$ is a paraboloid that opens downwards. Its global maximum is at $\theta = (0,0)$. We see that $\nabla_\theta f(\theta) = (2\theta_1, 2\theta_2)^T$, and hence $-\nabla_\theta f(\theta)$ points towards the origin at all points θ.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Gradient ascent/descent methods are typical tools for maximizing/minimizing functions. Consider the function $L(x,\theta)$ where $\theta = [\theta_1, \theta_2, \ldots, \theta_n]^T$ and $x = [x_1, x_2, \ldots, x_n]^T$. Our goal is to select θ such to maximize/minimize the value of L while keeping x fixed.

Compute the Gradient

1/1 point (graded)

The gradient $\nabla_\theta L(x,\theta)$ is a vector with n components:

$$\nabla_\theta L(x,\theta) = \begin{pmatrix} \frac{\partial}{\partial \theta_1} L(x,\theta) \\ \vdots \\ \frac{\partial}{\partial \theta_n} L(x,\theta) \end{pmatrix}.$$

(Note that we are treating x as a constant and also differentiating w.r.t. to θ.)

Let

$$L(x,\theta) = \log\left(1+\exp\left(-\theta\cdot x\right)\right).$$

(Notice that here the $\log$ function is the natural algorithm.)

Evaluate the gradient $\nabla_\theta L(x,\theta)$. Which of the following is its j^{th} component?

- ○ $\dfrac{\exp\left(-\theta\cdot x\right)}{1+\exp\left(-\theta\cdot x\right)}$
- ● $\dfrac{-x_j\exp\left(-\theta\cdot x\right)}{1+\exp\left(-\theta\cdot x\right)}$
- ○ $\dfrac{-x_j}{1+\exp\left(-\theta\cdot x\right)}$

✓

Solution:

The derivative of $\log(x) = \frac{1}{x}$ and the derivative of $e^{cx} = ce^{cx}$. Applying these rules with the chain rule gives the correct answer.

Submit

You have used 1 of 1 attempt

Answers are displayed within the problem

Gradient Ascent or Descent

1/1 point (graded)

The direction of the derivative of a function gives us the direction of the largest change in the function as the independent variables vary.

In gradient ascent/descent methods, we make an educated guess about the next values of θ, with consecutive updates that will hopefully eventually converge to the global minimum of $L\left(x,\theta\right)$ (if it exists).

If

$$\theta' = \theta + \epsilon \cdot \nabla_{\theta} L\left(x,\theta\right)$$

where ϵ is a small positive real number, Which of the following is true?

● $L\left(x,\theta'\right) > L\left(x,\theta\right)$

○ $L\left(x,\theta'\right) < L\left(x,\theta\right)$

✓

STANDARD NOTATION

Solution:

Consider the one-dimensional case. If the gradient is positive, we obtain θ' by moving from θ in the positive direction. This increases $L\left(x,\theta\right)$. If the gradient is negative, we move in the negative direction, again increasing $L\left(x,\theta\right)$. This analysis extends to higher dimensions. Note that if we used the function above to continue updating θ, we would (in theory) maximize $L\left(x,\theta\right)$. Alternatively if our update rule was $\theta' = \theta - \epsilon \cdot \nabla_{\theta} L\left(x,\theta\right)$, we would minimize the function. There are more complications in higher dimensions, but this is the basic idea behind stochastic gradient descent, which forms the backbone of modern machine learning.

Submit

You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 0. Course Overview, Homework 0, Project 0 (1 week):Homework 0 / 9. Gradients and Optimization

Add a Post

Show all posts by recent activity

💬 I found this link very usefull (gradients) 2
Hi all, after watching the gradient introductory video I still had some doubts, so I browsed an...

💬 Gradient Ascent or Descent 1
The last one is intentionally confusing without obvious reason. In gradient descent -epsilon*g...

💬 Compute the gradient 3
In this exercise, it has marked the correct answer as wrong. I checked my answer with two co...

💬 Geometric Picture of function 1
Anyone can give me hint on this question : Consider how the level curves θ21+θ22=+K chang...

? Geometric Picture of the Function 3
What does it mean in English: level curves ?

☑ Gradient Ascent or Descent: sign of x 5
Do we assume x is positive?

💬 Gradient Ascent or Descent 5
Any hint? I am really lost on this one

💬 Compute the gradient 3
Don't understand why some are xj an others are just x

💬 [STAFF] Some typos in "Geometric Picture of the Function" 1
In "We will use the usual horizonal plane as the (θ1,θ2 -plane, and the vertical axis as the f(θ) ...

💬 In Gradient Ascent or Descent, both answer choices are the same 4
title

Learn About Verified Certificates