

Exercise

01

TUM Department of Informatics

Supervised by	Prof. Dr. Stephan Günnemann Informatics 3 - Professorship of Data Mining and Analytics
Submitted by	Marcel Bruckner (03674122) Julian Hohenadel (03673879) Kevin Bein (03707775)
Submission date	Munich, November 23, 2019

Optimization

Problem 1:

a)

$h(x) = g_2(g_1(x))$ is not convex in this context:

Since g_2 and g_1 are convex: $g_2'' \geq 0$ and $g_1'' \geq 0$ but g_2 can be decreasing so $g_2'(x)$ does not have to be positive $\forall x$.

$$[h(x)]' = g_2'(g_1(x)) * g_1'(x)$$

$$[h(x)]'' = [g_2'(g_1(x)) * g_1'(x)]'$$

$$[h(x)]'' = g_2''(g_1(x)) * (g_1'(x))^2 + g_2'(g_1(x)) * g_1''(x)$$

So $[h(x)]''$ can be ≤ 0 if $g_2'(g_1(x)) \leq 0$ ($(g_1'(x))^2$ is always ≥ 0)

As an example: $g_2(x) = -\frac{1}{2}x$ and $g_1(x) = x^2$ are convex $\implies h(x) = -\frac{1}{2}x^2$ and should also be convex.

But the second derivation of $h(x)$ is -1 and therefore not convex.

b)

$h(x) = g_2(g_1(x))$ is convex in this context:

Since g_2 and g_1 are convex: $g_2'' \geq 0$ and $g_1'' \geq 0$ also g_2 is non-decreasing so $g_2'(x) \geq 0$.

$$[h(x)]' = g_2'(g_1(x)) * g_1'(x)$$

$$[h(x)]'' = [g_2'(g_1(x)) * g_1'(x)]'$$

$$[h(x)]'' = g_2''(g_1(x)) * (g_1'(x))^2 + g_2'(g_1(x)) * g_1''(x)$$

So $[h(x)]''$ can only be ≥ 0 . ($(g_1'(x))^2$ is always ≥ 0)

c)

$h(x) = \max(g_1(x), \dots, g_n(x))$ is always convex:

$$\begin{aligned} h(\lambda x + (1 - \lambda)y) &= \max(g_1(\lambda x + (1 - \lambda)y), \dots, g_n(\lambda x + (1 - \lambda)y)) \\ &\leq \max(\lambda g_1(x) + (1 - \lambda)g_1(y), \dots, \lambda g_n(x) + (1 - \lambda)g_n(y)) \\ &\leq \max(\lambda g_1(x), \dots, \lambda g_n(x)) + \max(\lambda g_1(y), \dots, \lambda g_n(y)) \\ &= \lambda h(x) + (1 - \lambda)h(y) \end{aligned}$$

Problem 2:

a)

Minimum x^* of f is the partial derivative wrt. x_1 and x_2 :

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= x_1 + 2 \stackrel{!}{=} 0 \implies x_1 = -2 \\ \frac{\partial f}{\partial x_2} &= 2x_2 + 1 \stackrel{!}{=} 0 \implies x_2 = -\frac{1}{2}\end{aligned}$$

x^* is at $x_1 = -2$ and $x_2 = -\frac{1}{2}$.

b)

2 steps gradient descent with $x^{(0)} = (0, 0)$ and learning rate $\tau = 1$:

Gradient descent in general:

1) Take point $x^{(n)}$

2) compute $f'(x)$

3) $x^{(n+1)} = x^{(n)} - \tau * f'(x)$

First step:

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= x_1 + 2 = 2 \implies x_1^{(1)} = 0 - 1 * 2 = -2 \\ \frac{\partial f}{\partial x_2} &= 2x_2 + 1 = 1 \implies x_2^{(1)} = 0 - 1 * 1 = -1 \\ \implies x^{(1)} &= (-2, -1)\end{aligned}$$

Second step:

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= x_1 + 2 = 0 \implies x_1^{(2)} = -2 - 1 * 0 = -2 \\ \frac{\partial f}{\partial x_2} &= 2x_2 + 1 = -1 \implies x_2^{(2)} = -1 - 1 * -1 = 0 \\ \implies x^{(2)} &= (-2, 0)\end{aligned}$$

c)

It will with $x_1 = -2$ but it won't with x_2 because it alternates between -1 and 0 .

To solve this problem a learning rate $0 < \tau < 1$ would be needed to stop the alternation of x_2 and help to converge to $x_2 = -\frac{1}{2}$.

Problem 3:

Problem 4:

Appendix

We confirm that the submitted solution is original work and was written by us without further assistance.
Appropriate credit has been given where reference has been made to the work of others.

Munich, November 23, 2019, Signature Marcel Bruckner (03674122)

Munich, November 23, 2019, Signature Julian Hohenadel (03673879)

Munich, November 23, 2019, Signature Kevin Bein (03707775)