# Machine Learning Exercise Sheet 6

# Optimization

**Group_369**

Fan Xue – fan98.xue@tum.de

Xing Zhou – xing.zhou@tum.de

Jianzhe Liu – jianzhe.liu@tum.de

December 2, 2021

---

## Problem 4

a) This statement is false, which can be proved by a counterexample. For that, we need to adopt an inference that the second derivation of a convex function is non-negative. We take $f(x) = x^2$ and $g(x) = -x$, both of which are convex funtions, since $\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = 2 \geq 0$ and $\frac{\mathrm{d}^2 g(x)}{\mathrm{d}x^2} = 0 \geq 0$. Then we take the composition of the two functions as $h(x) = g(f(x))$, and we have:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} h(x) = -2 \leq 0.$$

That means $h(x) = g(f(x))$ is non-convex.

b) This statement is true. We prove it with the definition of convex function. We take $h(x) = g(f(x))$, $\forall x_1, x_2 \in \mathbb{R}$ and $t \in (0, 1)$, since $f(x)$ is convex we have

$$f\left(tx_1 + (1-t)x_2\right) \leq tf\left(x_1\right) + (1-t)f\left(x_2\right).$$

Since $g(x)$ is non-decreasing and convex, we have

$$g\left(f\left(tx_1 + (1-t)x_2\right)\right) \leq g\left(tf\left(x_1\right) + (1-t)f\left(x_2\right)\right) \quad (1)$$

$$g\left(tf(x_1) + (1-t)f(x_2)\right) \leq tg\left(f(x_1)\right) + (1-t)g\left(f(x_2)\right). \quad (2)$$

Combine (1) and (2), we have

$$g\left(f\left(tx_1 + (1-t)x_2\right)\right) \leq tg\left(f(x_1)\right) + (1-t)g\left(f(x_2)\right)$$
$$\Updownarrow$$
$$h\left(tx_1 + (1-t)x_2\right) \leq th(x_1) + (1-t)h(x_2),$$

which proves $h(x) = g(f(x))$ is convex.

**Problem 5**

a) By observing the given function $\boldsymbol{f}$:

$$f(x_1, x_2) = 0.5{x_1}^2 + {x_2}^2 + 2x_1 + x_2 + \cos(\sin(\sqrt{\pi}))$$

we can find that it consists of several convex functions. Which means that when we separately find all the $x_n$ that minimize its own function, we find the $\boldsymbol{x}^\star$ that minimizes the function $\boldsymbol{f}$.
The gradient of f can be computed by followings:

$$\begin{pmatrix} \dfrac{\partial f(x_1, x_2)}{\partial x_1} \\ \dfrac{\partial f(x_1, x_2)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} x_1 + 2 \\ 2x_2 + 1 \end{pmatrix} = 0$$

So minimizer $\boldsymbol{x}^\star$ is:

$$\boldsymbol{x}^\star = \begin{pmatrix} -2 \\ -\dfrac{1}{2} \end{pmatrix}$$

b) 2 steps of gradient descent with leaning rate $\tau = 1$ can be performed by following procedure:

$$\text{step 1: } \begin{pmatrix} {x_1}^{(1)} \\ {x_2}^{(1)} \end{pmatrix} = \begin{pmatrix} {x_1}^{(0)} \\ {x_2}^{(0)} \end{pmatrix} - \tau \nabla_{\boldsymbol{x}} f\left(\boldsymbol{x}^{(0)}\right) = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

$$\text{step 2: } \begin{pmatrix} {x_1}^{(2)} \\ {x_2}^{(2)} \end{pmatrix} = \begin{pmatrix} {x_1}^{(1)} \\ {x_2}^{(1)} \end{pmatrix} - \tau \nabla_{\boldsymbol{x}} f\left(\boldsymbol{x}^{(1)}\right) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Here $\nabla f\left(\boldsymbol{x}\right)$ is obtained from **question a)** above, where

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = x_1 + 2$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = 2x_2 + 1$$

c) we can try computing two more iteration:

$$\begin{pmatrix} {x_1}^{(3)} \\ {x_2}^{(3)} \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} -2+2 \\ 0+1 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} {x_1}^{(4)} \\ {x_2}^{(4)} \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} - 1 \begin{pmatrix} -2+2 \\ -2+1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Obviously, it's an endless loop......Yet we already know that minimizer $\boldsymbol{x}^\star$ is: $\begin{pmatrix} -2 \\ \frac{1}{2} \end{pmatrix}$.

So the answer is, this gradient descent procedure can never converge to the true minimizer $\boldsymbol{x}^\star$.
To solve this problem, we must change the learning rate. In our case, it has to decrease to get to the minimizer.

## Problem 6

The result of the programming task is appended at the end of the file.

## Problem 7

a) This region S is not a convex region. Because we can find points whose line in between is outside region S. As shown in figure 1, red, green and blue lines are all in this situation.

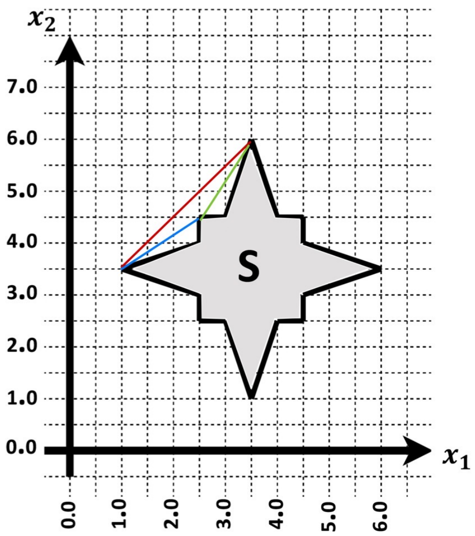


Figure 1: Figure 1

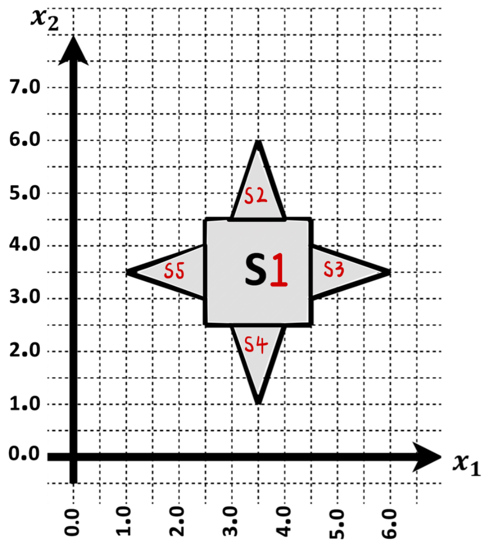


Figure 2: Figure 2

b) It's obvious that Function F is meaningful only if it's region is convex. So to give this **ConvOpt** algorithm a qualified input, we need to divide the original region S into 5 small regions, which are all convex regions(as shown in figure2).

This way, we can use this **ConvOpt** algorithm to get 5 different results, which are respectively the minimum in their region. Then we choose the minimum among this 5 values, it is the minimum of whole region S.