Exercise 06 of Machine Learning [IN 2064]

Name: Yiman Li
Matr-Nr: 03724352
cooperate with Kejia Chen(03729686)

Problem 1

a) We can not figure out clearly the convexity of the function $h(\boldsymbol{x})$ here.
Take $n = 1$ for instance. The second derivative of the composition function $h(\boldsymbol{x})$ is given by

$$h''(\boldsymbol{x}) = g_2''(g_1(\boldsymbol{x}))g_1'(\boldsymbol{x})^2 + g_2'(g_1(\boldsymbol{x}))g_1''(\boldsymbol{x}) \tag{1}$$

Since we don't know whether $g_2'(x)$ positive or negetive is, so the symbol of $h''(x)$ is ambigious, so we can not safely say that $h(\boldsymbol{x})$ here is convex.

b) We can prove that $h(\boldsymbol{x}) = g_2(g_1(\boldsymbol{x}))$ is convex in the given condition.
According to the book [1], firstly

- Since $g_1(\boldsymbol{x})$ is convex, so $g_1''(\boldsymbol{x}) \geq 0$;
- Since $g_2(\boldsymbol{x})$ is convex and non-decreasing, so $g_2''(\boldsymbol{x}) \geq 0$ and $g_2'(\boldsymbol{x}) \geq 0$.

The second derivative of the composition function $h(\boldsymbol{x})$ is given by

$$h''(\boldsymbol{x}) = g_2''(g_1(\boldsymbol{x}))g_1'(\boldsymbol{x})^2 + g_2'(g_1(\boldsymbol{x}))g_1''(\boldsymbol{x}) \geq 0 \tag{2}$$

i.e., $h(\boldsymbol{x}) = g_2(g_1(\boldsymbol{x}))$ is convex.

c) We can prove that $h(\boldsymbol{x})$ is convex if $\mathbf{dom}\ h = \mathbf{dom}\ g_1 \bigcap \mathbf{dom}\ g_2 \bigcap \cdots \bigcap \mathbf{dom}\ g_n$.
First we consider that $h_2(\boldsymbol{x}) = \max(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}))$, then we can derive that

$$\begin{aligned} h_2(\theta x + (1-\theta)y) &= \max\{g_1(\theta x + (1-\theta)y), g_2(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta g_1(x) + (1-\theta)g_1(y), \theta g_2(x) + (1-\theta)g_2(y)\} \\ &\leq \theta \max\{g_1(x), g_2(x)\} + (1-\theta)\max\{g_1(y), g_2(y)\} \\ &= \theta h_2(x) + (1-\theta)h_2(y) \end{aligned} \tag{3}$$

The equations above establish the convexity of $h_2(\boldsymbol{x})$. Then using the recursive method, we first assuming that $h_{n-1}(\boldsymbol{x}) = \max(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_{n-1}(\boldsymbol{x}))$ is a convex, then

$$\begin{aligned} h(\theta x + (1-\theta)y) &= \max\{h_{n-1}(\theta x + (1-\theta)y), g_n(\theta x + (1-\theta)y)\} \\ &\leq \max\{\theta h_{n-1}(x) + (1-\theta)h_{n-1}(y), \theta g_n(x) + (1-\theta)g_n(y)\} \\ &\leq \theta \max\{h_{n-1}(x), g_n(x)\} + (1-\theta)\max\{h_{n-1}(y), g_n(y)\} \\ &= \theta h(x) + (1-\theta)h(y) \end{aligned} \tag{4}$$

Combining equation 3 and equation 4, we can say that $h(\boldsymbol{x}) = \max(g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_n(\boldsymbol{x}))$ is also convex.

Problem 2

a) The objective function can be rewritten as

$$
\begin{aligned}
f(x_1, x_2) &= 0.5x_1^2 + x_2^2 + 2x_1 + x_2 + \cos(\sin(\sqrt{\pi})) \\
&= 0.5(x_1^2 + 4x_1 + 4) + (x_2^2 + x_2 + \frac{1}{4}) + \cos(\sin(\sqrt{\pi})) - \frac{9}{4} \\
&= 0.5(x_1 + 2)^2 + (x_2 + \frac{1}{2})^2 + \cos(\sin(\sqrt{\pi})) - \frac{9}{4} \\
&\geq \cos(\sin(\sqrt{\pi})) - \frac{9}{4}
\end{aligned} \tag{5}
$$

The equality is reached when $\boldsymbol{x}^* = (-2, -\frac{1}{2})^T$.

b) The derivative of the function can be written as

$$
f'(x_1, x_2) = (x_1 + 2, 2x_2 + 1) \tag{6}
$$

starting from $\boldsymbol{x}^{(0)}$ and $\tau = 1$, we have

$$
\begin{aligned}
\boldsymbol{x}^{(1)} &= \boldsymbol{x}^{(0)} - \tau \cdot f_0'(x_1, x_2) = (0, 0) - (2, 1) = (-2, -1) \\
\boldsymbol{x}^{(2)} &= \boldsymbol{x}^{(1)} - \tau \cdot f_1'(x_1, x_2) = (-2, -1) - (0, -1) = (-2, 0) \\
\boldsymbol{x}^{(3)} &= \boldsymbol{x}^{(2)} - \tau \cdot f_2'(x_1, x_2) = (-2, 0) - (0, 1) = (-2, -1)
\end{aligned} \tag{7}
$$

c) Oberseving from $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$ and $\boldsymbol{x}^{(3)}$, we can see that the function has run into an infinite loop and it can never converge to the true minumum $\boldsymbol{x}^*$ due to the oscillations resulted from too large learning rate. On this condition we should tune our learning rate a bit more smaller. But attention should be paid to other problems, for examply, we may find the minimum slowly, or end up in local minima or saddle points.

Problem 3

Show in the end.

Problem 4

a) No! the shaded region S in $\mathbb{R}^2$ is not convex. As shown in Figure 1, we draw a line between the

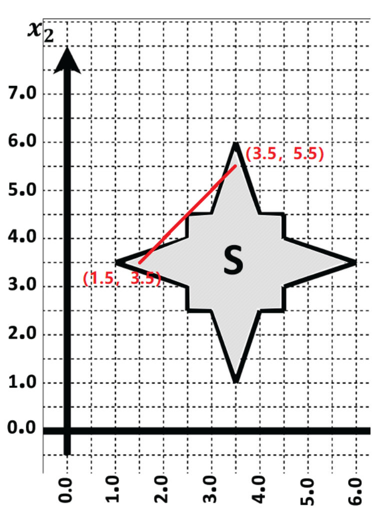

Figure 1: Not a convex

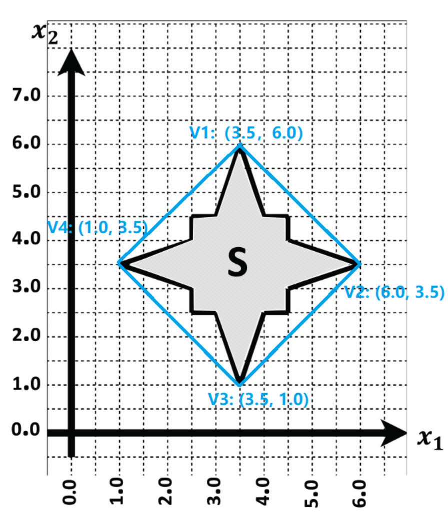

Figure 2: Convex after filling

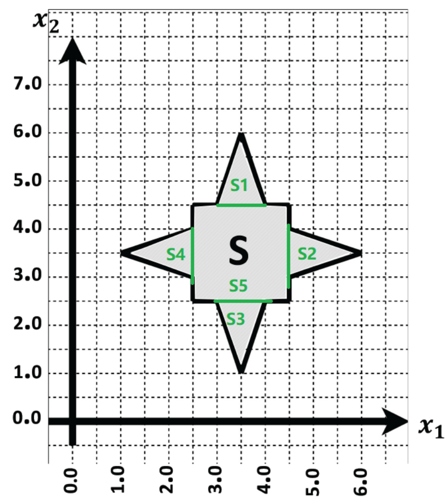

Figure 3: Five sub-region

two points $(1.5, 3.5)$ and $(3.5, 5.5)$, and there exist some points on the line distributed outside of the shaded region. That is to say, not all the point follows the law that

$$\lambda x + (1 - \lambda)y \in X \text{ for } \lambda \in [0, 1] \tag{8}$$

So the region here is not convex.

b) Since the maximum over a convex function on a convex set is obtained on a vertex, so as shown in Figure 2, we simply connecting four vertices, namely, $V_1(3.5, 6.0)$, $V_2(6.0, 3.5)$, $V_3(3.5, 1.0)$, $V_4(1.0, 3.5)$ of the shaded region, and now the augumented region becomes convex. So the maximum of convex function $f(x_1, x_2) = e^{x_1+x_2} - 5\log(x_2)$ must be one of these four vertices, so just compute:

$$\begin{aligned} f_{V_1}(x_1, x_2) &= f(3.5, 6.0) = e^{9.5} - 5\log(6.0) = 13346.80 \\ f_{V_2}(x_1, x_2) &= f(6.0, 3.5) = e^{9.5} - 5\log(3.5) = 13350.69 \\ f_{V_3}(x_1, x_2) &= f(3.5, 1.0) = e^{4.5} - 5\log(1.0) = 90.02 \\ f_{V_2}(x_1, x_2) &= f(1.0, 3.5) = e^{4.5} - 5\log(3.5) = 80.98 \end{aligned} \tag{9}$$

So the maximum of f over the shaded region S is obtained at the point $\boldsymbol{x}^* = (6.0, 3.5)$.

c) We can divide the shaded region into the combination of five sub-region, as is shown in Figure 3, namely, $S = S_1 \bigcap S_2 \bigcap S_3 \bigcap S_4 \bigcap S_5$, with every sub-region being convex region. Then we can use the algorith ConvOpt(f, D) to find the minimum on every sub-region. At last, we need to make contrast among these local minimums, and finally find the true global minimum.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.