Machine Learning homework 5 solution
Optimization

Wiktor Jurasz - M.Nr. 03709419

November 21, 2018

1 Problem 1

1.1

We know that sum of convex functions is also convex, so in this case it's enough to prove that each component is convex.

- x^2 - Convex because second derivative (2) is always non negative
- $2y$ - Convex because linear functions are convex.
- $cos(sin(\sqrt{\pi}))$ - Convex because constant are convex (it's also just linear function).
- $-\min\{-x^2, log(y)\} = \max(x^2, -log(y))$ - Convex because max of convex functions is convex (proven later). $-log(y)$ is convex because second derivative $\frac{1}{x^2}$ is always non negative (and logarithm is defined on $(1, 50)$.

$f(x, y)$ **is convex**

1.2

$$f'(x) = x^{-1} - 3x^2 \tag{1}$$
$$f''(x) = -x^{-2} - 6x \tag{2}$$

$$\begin{gathered} f''(x) \geq 0 \Rightarrow \\ -x^{-2} - 6x \geq 0 \Rightarrow \\ -1 - 6x^3 \geq 0 \Rightarrow \\ -1 \geq 6x^3 \Rightarrow \\ x \leq \sqrt[3]{\frac{-1}{6}} \Rightarrow \\ x \leq\sim -0.55 \end{gathered} \tag{3}$$

$(-\infty, -0.55) \cap (1, \infty) = \emptyset \Rightarrow f(x)$ **is not convex**

1.3

$$f(x) = -\min\{log(3x+1), -x^4 - 3x^2 + 8x - 42\} = \max\{-log(3x+1), x^4 + 3x^2 - 8x + 42\} \tag{4}$$

The first part of the max is convex because $-log$ is convex and non-decreasing and $3x + 1$ is convex and composition of convex non-decreasing function with convex function is convex.
The second part of max is convex because all elements of the polynomial are convex.
Since both functions are convex and max of convex functions is convex $f(x)$ **is convex**

1.4

$$
\begin{gathered}
\frac{\partial f}{\partial x} = 3yx^2 - 2yx \\
\frac{\partial f}{\partial y} = x^3 - x^2 + 2y + 1 \\
\frac{\partial f}{\partial x^2} = 6yx - 2y \\
\frac{\partial f}{\partial x \partial y} = 3x^2 - 2x \\
\frac{\partial f}{\partial y \partial x} = 3x^2 - 2x \\
\frac{\partial f}{\partial y^2} = 2 \\
M = \begin{bmatrix} 6yx - 2y & 3x^2 - 2x \\ 3x^2 - 2x & 2 \end{bmatrix}
\end{gathered} \tag{5}
$$

$$
\begin{bmatrix} a & b \end{bmatrix} M \begin{bmatrix} a \\ b \end{bmatrix} = a^2(6yx - 2y) + 2ab(3x^2 - 2x) + 2b^2 \tag{6}
$$

For function to be convex above scalar has to be non-negative for any natural a and b.
However, following values:

$$
\begin{gathered}
a = 1 \\
b = 0 \\
x = 0 \\
y = 1
\end{gathered} \tag{7}
$$

gives us $1 * -2 * 1 = -1 < 0 \Rightarrow f(x, y)$ **is not convex**

2 Problem 2

Let's take:

$$
\begin{gathered}
x, y \in \mathbb{R}^d \\
\lambda \in (0, 1)
\end{gathered}
$$

Then:

$$
\lambda x + (1 - \lambda) y
$$

Also belongs to convex set of f_1, f_2.

$$
\begin{aligned}
h(\lambda x + (1 - \lambda)y) &= \max\{f_1(\lambda x + (1 - \lambda)y), f_2(\lambda x + (1 - \lambda)y\} \\
&\leq \max\{\lambda f_1(x) + (1 - \lambda) f_1(y), \lambda f_2(x) + (1 - \lambda) f_2(y)\} \\
&\leq \lambda \max\{f_1(x), f_2(x)\} + (1 - \lambda) \max\{f_1(y), f_2(y)\} \\
&= \lambda h(x) + (1 - \lambda) h(y)
\end{aligned} \tag{8}
$$

Q.E.D

3 Problem 3

$$
g'(x) = f_1'(f_2(x)) f_2'(x) \tag{9}
$$

$$
g''(x) = f_1''(f_2(x))(f_2'(x))^2 + f_1'(f_2(x)) f_2''(x) \tag{10}
$$

From the fact that f_1, f_2 are convex and square property we can derive following:

$$
\begin{aligned}
f_1'' &\geq 0 \\
(f_2'(x))^2 &\geq 0 \\
f_2''(x) &\geq 0
\end{aligned}
\tag{11}
$$

For g to be convex $g''(x)$ has to be non-negative. However, $f_1'(f_2(x))$ can be both negative and positive. Now if we assume additional properties for f_1

- f_1 is decreasing on $\mathbb{R}$
- $f_1''(x) = 0$

(e.g $f_1(x) = -x$)
Then $g''(x) < 0 \Rightarrow f_1(f_2(x))$ **is not (always) convex**

4 Problem 4

Assumptions

- f is convex
- $x, y \in \mathbb{R}^n$
- $\nabla f(x) = 0, \nabla f(y) = 0$ are local minima
- $f(x) \geq f(y)$
- $\lambda \in (0, 1)$

From assumptions and convexity definition:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda f(x) + (1-\lambda)f(x) = f(x) \tag{12}$$

To write it more clearly:

$$f(\lambda x + (1-\lambda)y) \leq f(x) \tag{13}$$

Because of our assumptions regarding λ we know that $\lambda x + (1-\lambda)y$ lies somewhere "between" x and y. But we also said that $f(x)$ is a local minimum, which means it has to be smaller then all the values in it's neighbourhood. So the only way to satisfy equation 13 is to set $x = y$ which means that there is only one (global) minimum.