Machine Learning Homework Sheet 05

Optimization

1 Convexity

Problem 1: Prove or disprove whether the following functions are convex on the given set D:

i) $f(x, y) = x^2 + 2y + \cos(\sin(\sqrt{\pi})) - \min\{-x^2, \log(y)\}$ and $D = (-100, 100) \times (1, 50)$

ii) $f(x) = \log(x) - x^3$ and $D = (1, \infty)$

iii) $f(x) = -\min\{\log(3x + 1), -x^4 - 3x^2 + 8x - 42\}$ and $D = \mathbb{R}^+$

iv) $f(x, y) = y \cdot x^3 - y \cdot x^2 + y^2 + y + 4$ and $D = (-10, 10) \times (-10, 10)$

i) We will show that $f(x, y)$ is convex since it is a sum of convex functions:

- We can easily see that x^2 is convex, it's a parabola ($\frac{d^2}{dx^2}x^2 = 2 \geq 0$).
- $2y$ is convex since it is a linear function.
- $\cos(\sin(\sqrt{\pi}))$ is constant and therefore convex.
- $-\min\{-x^2, \log(y)\} = \max\{x^2, -\log(y)\}$. Thus, using the max rule this term is convex if x^2 and $-\log(y)$ are convex. x^2 is a parabola and therefore convex. Since $\frac{d^2}{dy^2}(-\log y) = \frac{d}{dy}(-\frac{1}{y}) = \frac{1}{y^2} \geq 0$ (on D) the second derivative is non-negative and $-\log(y)$ is also convex.

ii) The second derivative of this function is $\frac{d^2 f(x)}{dx^2} = \frac{d}{dx}\left(\frac{1}{x} - 3x^2\right) = -\frac{1}{x^2} - 6x$, which is negative on the given set D and therefore not convex.

iii)

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

is convex if both arguments of max are convex on $\mathbb{R}^+$.

$-\log(3x + 1)$ is convex: Second derivative test (on $\mathbb{R}^+$):

$$\frac{d^2}{dx^2}(-\log(3x + 1)) = \frac{d}{dx}\left(-\frac{3}{3x + 1}\right) = \frac{9}{(3x + 1)^2} \geq 0$$

$x^4 + 3x^2 - 8x + 42$ is convex: Second derivative test,

$$\frac{d^2}{dx^2}\left(x^4 + 3x^2 - 8x + 42\right) = \frac{d}{dx}\left(4x^3 + 6x - 8\right) = 12x^2 + 6 \geq 0$$

Thus $f(x)$ is convex.

iv) For the function $f(x, y)$ to be convex (on D) it has to hold:

$$\lambda f(x_1, y) + (1-\lambda) f(x_2, y) \geq f(\lambda x_1 + (1-\lambda) x_2, y) \quad \forall x_1, x_2, y \in D, \quad \forall \lambda \in [0,1]$$

Counterexample: Choose $y = 1, x_1 = -4, x_2 = 0, \lambda = 0.5$:

$$0.5 f(-4,1) + 0.5 f(0,1) \geq f(0.5 \cdot (-4) + 0.5 \cdot 0, 1) \quad (1)$$
$$\Leftrightarrow -0.5 \cdot 74 + 0.5 \cdot 6 \geq -6 \quad (2)$$
$$\Leftrightarrow -34 \geq -6 \quad ↯ \quad (3)$$

Thus $f(x, y)$ is not convex.

Problem 2: Prove the following statement: Let $f_1 : \mathbb{R}^d \to \mathbb{R}$ and $f_2 : \mathbb{R}^d \to \mathbb{R}$ be convex functions, then $h(\boldsymbol{x}) := \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is also a convex function.

Since f_1 and f_2 are convex (definiton of convexity): $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \; \forall \lambda \in [0,1]$:

$$\lambda f_1(\boldsymbol{x}) + (1-\lambda) f_1(\boldsymbol{y}) \geq f_1(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}),$$
$$\lambda f_2(\boldsymbol{x}) + (1-\lambda) f_2(\boldsymbol{y}) \geq f_2(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}).$$

Since

$$\lambda \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} + (1-\lambda) \max\{f_1(\boldsymbol{y}), f_2(\boldsymbol{y})\} \geq \lambda f_1(\boldsymbol{x}) + (1-\lambda) f_1(\boldsymbol{y}) \geq f_1(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}),$$
$$\lambda \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} + (1-\lambda) \max\{f_1(\boldsymbol{y}), f_2(\boldsymbol{y})\} \geq \lambda f_2(\boldsymbol{x}) + (1-\lambda) f_2(\boldsymbol{y}) \geq f_2(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})$$

it follows that

$$\lambda \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\} + (1-\lambda) \max\{f_1(\boldsymbol{y}), f_2(\boldsymbol{y})\} \geq \max\{f_1(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}), f_2(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y})\}$$

and therefore

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

Note that $h(\boldsymbol{x}) := \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$ is not necessarily differentiable.

Problem 3: Given two convex functions $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$, prove or disprove that the function $g(x) = f_1(f_2(x))$ is also convex.

Disprove via counterexample: Choose $f_1(x) = -x$ and $f_2(x) = x^2$. Both f_1 and f_2 are convex. However, $g(x) = -x^2$ is not, as evident by looking at the second derivative, $\frac{d^2 g(x)}{dx^2} = -2 < 0$.

2 Minimization of convex functions

Problem 4: Prove that for convex functions each local minimum is a global minimum. More specifically, given a convex function $f : \mathbb{R}^N \to \mathbb{R}$, prove that if $\nabla f(\boldsymbol{\theta}^*) = 0$ then $\boldsymbol{\theta}^*$ is a global minimum.

At a (local) minimum point $\boldsymbol{\theta}^*$ the gradient must be zero. Otherwise we could follow the gradient to get an even lower value.

$$f(\boldsymbol{\theta}^* - \epsilon \nabla f(\boldsymbol{\theta}^*)) = f(\boldsymbol{\theta}^*) - \epsilon ||\nabla f(\boldsymbol{\theta}^*))||_2^2 + O(\epsilon^2 ||\nabla f(\boldsymbol{\theta}^*)||_2^2) < f(\boldsymbol{\theta}^*)$$

for sufficiently small ϵ.

Now, using the first order criterion we have

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + (\boldsymbol{y} - \boldsymbol{x}) \nabla f(\boldsymbol{x})$$

If we replace $\boldsymbol{x}$ with $\boldsymbol{\theta}^*$ and plug in $\nabla f(\boldsymbol{\theta}^*) = 0$ we get: $f(\boldsymbol{y}) \geq f(\boldsymbol{\theta}^*)$ for all $\boldsymbol{y}$, meaning $\boldsymbol{\theta}^*$ is a global minimum.

3 Gradient Descent

Problem 5: Load the notebook `homework_05_notebook.ipynb` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to pdf and add it to the printout of your homework (instructions for this are provided within the notebook).

The solution notebook is uploaded on Piazza.