5. HW Optimization

Mittwoch, 28. November 2018 14:12

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 $D = (-100, 100) \times (1, 50)$

x y

$f: \mathbb{R}^2 \to \mathbb{R}$

sum of conv fn is conv

$-\min(-x^2, \log(y)) = \max(x^2, -\log(y))$

$\partial_y^2 - \log y = \partial_y\left(-\frac{1}{y}\right) = \frac{1}{y^2} \geq 0$ (on D)

ii) $f(x) = \log(x) - x^3$ and $D = (1, \infty)$ >0

$\partial_x^2 \left(\log(x) - x^3\right) = \partial_x \left(\frac{1}{x} - 3x^2\right) = -\frac{1}{x^2} - 6x < 0$

not convex

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

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

1) $\partial_x^2(-\log(3x+1)) = \partial_x \frac{-1}{3x+1} \cdot 3 = \frac{9}{(3x+1)^2} \geq 0$

2) $\partial_x^2(x^4 + 3x^2 - 8x + 42) = \partial_x(4x^3 + 6x + 8) = 12x^2 + 6 \quad \geq 0$

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

$\lambda f(x_1, y) + (1-\lambda) f(x_2, y) \geq f(\lambda x_1 + (1-\lambda) x_2, y)$

$y = 1, \; x_2 = 0, \; x_1 = -4, \; \lambda = 0{,}5$

$0{,}5 \, f(-4, 1) + 0{,}5 \, f(0, 1) \geq f(0{,}5(-4) + 0{,}5 \cdot 0, 1)$

$\Leftrightarrow \quad -0{,}5 \cdot 74 + 0{,}5 \cdot 6 \geq -6$

$\Leftrightarrow \quad -34 \geq -6$ ↯

$\Leftrightarrow \quad -34 \geq -6$

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.

$$\lambda f_i(\vec{x}) + (1-\lambda) f_i(\vec{y}) \geq f_i(\lambda\vec{x} + (1-\lambda)\vec{y})$$

$$\lambda h(\vec{x}) + (1-\lambda) h(\vec{y}) = \lambda \max\{f_1(\vec{x}), f_2(\vec{x})\} + (1-\lambda) \max\{f_1(\vec{y}), f_2(\vec{y})\}$$
$$\geq \lambda f_1(\vec{x}) + (1-\lambda) f_1(\vec{y}) \geq f_1(\lambda\vec{x} + (1-\lambda)\vec{y})$$
$$\geq \lambda f_2(\vec{x}) + (1-\lambda) f_2(\vec{y}) \geq f_2(\lambda\vec{x} + (1-\lambda)\vec{y})$$
$$\geq \max\{f_1(\lambda\vec{x} + (1-\lambda)\vec{y}), f_2(\lambda\vec{x} + (1-\lambda\vec{y}))\} =$$
$$= h(\lambda\vec{x} + (1-\lambda)\vec{y}) \quad \square$$

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.

$$f_1(x) = -x \,, \quad f_2(x) = x^2 \quad \cup$$

$$g(x) = -x^2 \quad \cap$$

$$\partial_x^2 g = -2 < 0$$

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 $\boldsymbol{\nabla} f(\boldsymbol{\theta}^*) = 0$ then $\boldsymbol{\theta}^*$ is a global minimum.

$$f(\vec{\theta}^* - \varepsilon \vec{\nabla} f(\vec{\theta}^*)) = f(\vec{\theta}^*) - \varepsilon \|\vec{\nabla} f(\vec{\theta}^*)\|_2^2 + \mathcal{O}(\varepsilon^2 \|\vec{\nabla} f(\vec{\theta}^*)\|_2^2) < f(\vec{\theta}^*)$$
$$\Rightarrow \vec{\nabla} f(\vec{\theta}^*) = 0$$

$$f(\vec{y}) \geq f(\vec{x}) + (\vec{y} - \vec{x}) \vec{\nabla} f(\vec{x})$$
$$\vec{x} = \vec{\theta}^*, \quad \vec{\nabla} f(\vec{\theta}^*) = 0$$
$$\Rightarrow f(\vec{y}) \geq f(\vec{\theta}^*)$$
$\Leftrightarrow$ 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).

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).

1) $$\frac{1}{1+e^{-t}}$$

2) $$E(\vec{w}) = -\sum_{i=1}^{N} \left[y_i \ln\left(\sigma(\vec{w}^T \vec{x}_i) + \varepsilon \right) + (1-y_i) \ln\left(1-\sigma(\vec{w}^T \vec{x}_i) + \varepsilon \right) \right]$$

3) $$s = \sigma(\vec{w}^T \vec{x}_i)$$

$$\partial_{\vec{w}} s = s(1-s)\vec{x}_i$$

$$\vec{\nabla}_{\vec{w}} E(\vec{w}) = -\sum_{i=1}^{N} \left[y_i \frac{\cancel{s}(1-s)}{\cancel{s}} \vec{x}_i + (1-y_i) \frac{-s\cancel{(1-s)}\vec{x}_i}{\cancel{1-s}} \right] =$$

$$= -\sum_{i=1}^{N} \vec{x}_i \left[y_i - s\cancel{y_i} - s + s\cancel{y_i} \right] = \sum_{i=1}^{N} \vec{x}_i (s - y_i)$$

$$\frac{1}{N} \frac{N}{b} \sum_{i=1}^{b} X_{i,:} \left(\sigma\left(\sum_{j=1}^{D} X_{i,j} \cdot w_j \right) - y_i \right) + \lambda \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \odot \vec{w}$$