Machine Learning Exercise Sheet 06

Optimization

Homework

1 Convexity of functions

Problem 1: Given n convex functions $g_i : \mathbb{R}^{d_i} \to \mathbb{R}$ for $i \in \{1, \dots, n\}$, prove or disprove that the function

a) $h(\boldsymbol{x}) = g_2(g_1(\boldsymbol{x}))$ is convex (here $d_1 \in \mathbb{N}$, $d_2 = 1$),

b) $h(\boldsymbol{x}) = g_2(g_1(\boldsymbol{x}))$ is convex if g_2 is non-decreasing (here $d_1 \in \mathbb{N}$, $d_2 = 1$),

c) $h(\boldsymbol{x}) = \max(g_1(\boldsymbol{x}), \dots, g_n(\boldsymbol{x}))$ is convex (here all $d_i \in \mathbb{N}$).

2 Optimization / Gradient descent

Problem 2: You are given the following objective function $f : \mathbb{R}^2 \to \mathbb{R}$

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

a) Compute the minimum $\boldsymbol{x}^*$ of f analytically.

b) Perform 2 steps of gradient descent on f starting from the point $\boldsymbol{x}^{(0)} = (0, 0)$ with a constant learning rate $\tau = 1$.

c) Will the gradient descent procedure from Problem b) ever converge to the true minimum $\boldsymbol{x}^*$? Why or why not? If the answer is no, how can we fix it?

Problem 3: Load the notebook `06_homework_optimization.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.

Note: We suggest that you use Anaconda for installing Python and Jupyter, as well as for managing packages. We recommend that you use Python 3.

For more information on Jupyter notebooks and how to convert them to other formats, consult the Jupyter documentation and nbconvert documentation.

Upload a single PDF file with your homework solution to Moodle by 24.11.2019, 23:59 CET. We recommend to typeset your solution (using LaTeX or Word), but handwritten solutions are also accepted. If your handwritten solution is illegible, it won't be graded and you waive your right to dispute that.

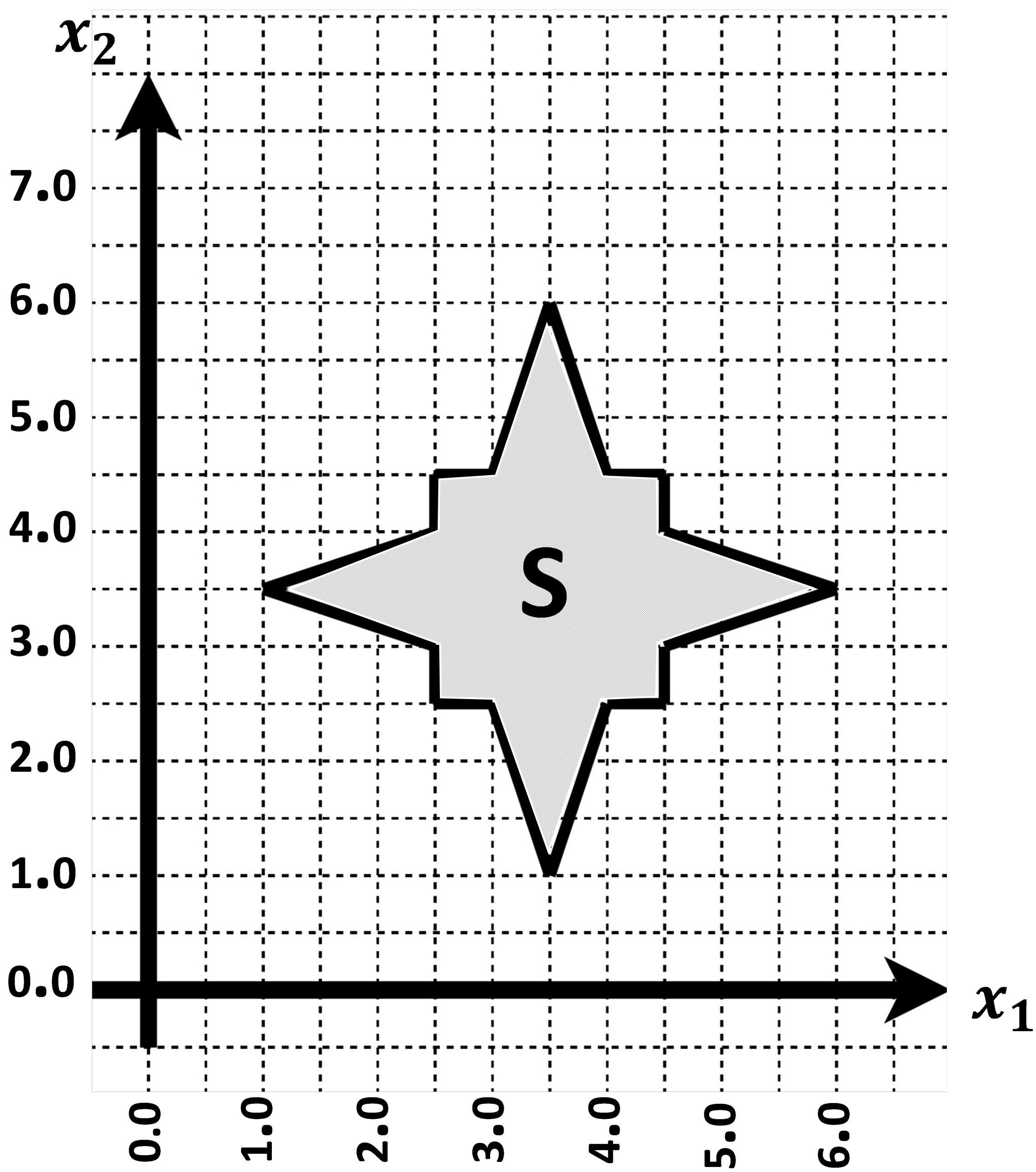
Problem 4: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the following convex function:

$$f(x_1, x_2) = e^{x_1+x_2} - 5 \cdot \log(x_2)$$

a) Consider the following shaded region S in $\mathbb{R}^2$. Is this region convex? Why?

b) Find the maximizer $\boldsymbol{x}^*$ of f over the shaded region S. Justify your answer.

c) Assume that we are given an algorithm `ConvOpt`(f, D) that takes as input a convex function f and convex region D, and returns the minimum of f over D. Using the `ConvOpt` algorithm, how would you find the global minimum of f over the shaded region S?



In-class Exercises

Problem 5: Prove or disprove whether the following functions $f : D \to \mathbb{R}$ are convex

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

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

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

Problem 6: Prove that the following functions $f : D \to \mathbb{R}$ are convex:

a) $D \subset \mathbb{R}^d$ bounded and closed set and $f : \mathbb{R}^d \to \mathbb{R}$ is defined by $f(\boldsymbol{x}) = \max_{\boldsymbol{w} \in D} \boldsymbol{x}^T \boldsymbol{w}$.

b) $D = \mathbb{R}^d$ and f is the objective function of logistic regression

$$E(\boldsymbol{w}) = -\ln p\big(\boldsymbol{y} \mid \boldsymbol{w}, \boldsymbol{X}\big) = -\sum_{i=1}^{N} \big(y_i \ln \sigma(\boldsymbol{w}^T \boldsymbol{x}_i) + (1 - y_i) \ln(1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_i))\big) .$$

Problem 7: Prove that for differentiable convex functions each local minimum is a global minimum. More specifically, given a differentiable convex function $f : \mathbb{R}^d \to \mathbb{R}$, prove that if there is a local minimum at $\boldsymbol{x}^*$ then $\nabla f(\boldsymbol{x}^*) = 0$ and if $\nabla f(\boldsymbol{x}^*) = 0$ then $\boldsymbol{x}^*$ is a global minimum.

Problem 8: Show that a twice differentiable function $f : D \to \mathbb{R}$ with a convex domain $D \subset \mathbb{R}^d$ is convex if and only if its Hessian is positive semi-definite on the whole domain D.