ML In-class exercise 6 - Optimization

Thursday, June 18, 2020 10:10 PM

In-class Exercises

Problem 1: 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 $f(x) = -\min(\log(3x+1), -x^4 - 3x^2 + 8x - 42)$, $\mathbb{R}^+ = (0, \infty)$

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

a) The second derivative of f 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 f is not convex.

b) Transform min to max

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

$\max(g_1(x), g_2(x))$ is convex if both g_1 and g_2 are convex on $D = \mathbb{R}^+$ (see Exercise Sheet 6, Problem 1c). $g_1(x) = -\log(3x+1)$ is convex since the second derivative is positive on $\mathbb{R}^+$:

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

$g_2(x) = x^4 + 3x^2 - 8x + 42$ is also convex:

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

Thus f is convex.

c) For the function $f(x, y)$ to be convex (on D) it has to hold for all $x_1, x_2, y \in D$ and $\lambda \in (0, 1)$ that

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

It does not hold in our case, consider $y = 1, x_1 = -4, x_2 = 0$ and $\lambda = 0.5$:

$$0.5 f(-4, 1) + 0.5 f(0, 1) = 0.5 \cdot (-74) + 0.5 \cdot 6 = -34$$
$$f(0.5 \cdot (-4) + 0.5 \cdot 0, 0.5 \cdot 1 + 0.5 \cdot 1)) = f(-2, 1) = -6 \;\; > -34$$

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

Proving convexity

1) $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \quad \forall x, y \in D \quad \forall \lambda \in (0,1)$

2) f is 1-diff.

$f(x) \geq f(y) + (x - y)^T \nabla f(y) \quad \forall x, y \in D$

!!! This was wrong in the recording !!!

3) is 2-diff.

$\underset{d \times d}{\nabla^2 f(x)} \succeq 0 \iff z^T \nabla^2 f(x) z \geq 0 \quad \forall x \in D \quad \forall z \in \mathbb{R}^d$

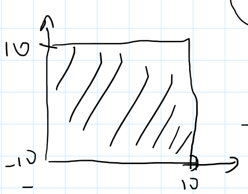
$f''(x) \geq 0$

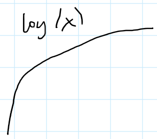
4) Rules

Disproving convexity

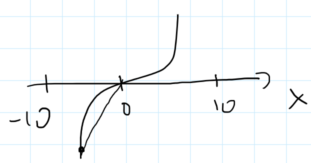
→ Counterexamples

e.g. find a single x: $\nabla^2 f(x)$ is not p.s.d $\iff \exists\, z \in \mathbb{R}^d : z^T \nabla^2 f(x) \cdot z < 0$





$-\min(a, b) = \max(-a, -b) \qquad h(x) = \max(f(x), g(x))$ – cvx. if $f(x)$ and $g(x)$ are cvx.



Problem 2: Prove that the following function (the loss function of logistic regression) $f : \mathbb{R}^d \to \mathbb{R}$ is convex:

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

N N×d z

First, let's simplify the above expression. For this we will need the following two facts

$$\sigma(z) = \frac{1}{1+e^{-z}} = \frac{e^z}{1+e^z} \quad \text{and} \quad 1 - \sigma(z) = \sigma(-z) = \frac{1}{1+e^z},$$

which implies that

$$\ln \sigma(z) = \ln\left(\frac{e^z}{1+e^z}\right) = z - \ln(1+e^z) \quad \text{and} \quad \ln(1-\sigma(z)) = -\ln(1+e^z).$$

Plugging this into the definition of the loss function we obtain

$$\begin{aligned} \boxed{f(\boldsymbol{w})} &= -\sum_{i=1}^{N} \left(y_i \ln \sigma(\boldsymbol{w}^T \boldsymbol{x}_i) + (1-y_i) \ln(1 - \sigma(\boldsymbol{w}^T \boldsymbol{x}_i)) \right) \\ &= -\sum_{i=1}^{N} \left(y_i \left(\boldsymbol{w}^T \boldsymbol{x}_i - \ln(1 + e^{\boldsymbol{w}^T \boldsymbol{x}_i}) \right) - (1-y_i) \ln(1 + e^{\boldsymbol{w}^T \boldsymbol{x}_i}) \right) \\ &= \sum_{i=1}^{N} \left(\underline{-y_i (\boldsymbol{w}^T \boldsymbol{x}_i)} + \underline{\ln(1 + e^{\boldsymbol{w}^T \boldsymbol{x}_i})} \right) \end{aligned}$$

We know that $\boldsymbol{w}^T \boldsymbol{x}_i$ is a convex (and concave) function of $\boldsymbol{w}$. Therefore, the first term $-y_i(\boldsymbol{w}^T \boldsymbol{x}_i)$ is also convex.

Now, if we show that $\ln(1 + e^z)$ is a nondecreasing and convex function of z on $\mathbb{R}$, we will be able to use the convexity preserving operations to prove that $f(\boldsymbol{w})$ is convex.

The first derivative of $\ln(1+e^z)$ is

$$\frac{d}{dz} \ln(1+e^z) = \frac{e^z}{1+e^z} = \sigma(z),$$

which is positive for all $z \in \mathbb{R}$, which means that $\ln(1+e^z)$ is an nondecreasing function.

The second derivative is

$$\frac{d^2}{dz^2} \ln(1+e^z) = \frac{d}{dz} \sigma(z) = \sigma(z)\sigma(-z), > 0$$

which is also positive for all $z \in \mathbb{R}$, which means that $\ln(1+e^z)$ is a convex function.

Using the following two facts

1. Sum of convex functions is convex
2. Composition of a convex function with a convex nondecreasing function is convex

we can verify that $f(\boldsymbol{w})$ is indeed convex in $\boldsymbol{w}$ on $\mathbb{R}^d$.

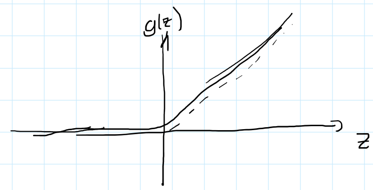
$$\frac{1}{(1+e^{-z}) \cdot e^z} \cdot e^z = \frac{e^z}{e^z + e^{-z} e^z} = \frac{e^z}{e^z + 1}$$

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b)$$

$a^T x$ – cvx. & concave in x

f - cvx.
g – cvx. & nondecreasing $\Rightarrow$ $g(f(x))$ – convex

$f(x) = w^T x_i$
$g(z) = \ln(1 + e^z)$ $\leftarrow$ convex & nondecreasing



$\sigma(z) \in (0,1) \to \sigma(z) > 0$

$g' \geq 0$

Problem 3: 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

a) if $\boldsymbol{x}^*$ is a local minimum, then $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$.

b) if $\boldsymbol{\nabla} f(\boldsymbol{x}^*) = \boldsymbol{0}$, then $\boldsymbol{x}^*$ is a global minimum.

We will show that if the gradient at some point $\boldsymbol{x}^*$ is not equal to zero, then this point cannot be a local optimum — we could simply follow the direction of the negative gradient and end up in a point with a lower value of the function.

More formally, suppose $\nabla f(\boldsymbol{x}^*) \neq \boldsymbol{0}$ for some $\boldsymbol{x}^*$. Then by Taylor's theorem for a sufficiently small $\varepsilon > 0$ we get

$\|y - x\|^2$

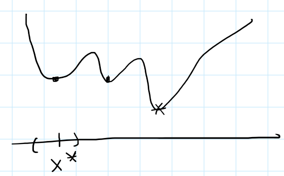
$$\begin{aligned} f(\underbrace{\boldsymbol{x}^* - \varepsilon \boldsymbol{\nabla} f(\boldsymbol{x}^*)}_{y}) &= f(\boldsymbol{x}^*) - (\varepsilon \boldsymbol{\nabla} f(\boldsymbol{x}^*))^T \boldsymbol{\nabla} f(\boldsymbol{x}^*) + O(\varepsilon^2 \|\boldsymbol{\nabla}^2 f(\boldsymbol{x}^*)\|_2^2) \\ &= f(\boldsymbol{x}^*) \underbrace{- \varepsilon \|\boldsymbol{\nabla} f(\boldsymbol{x}^*)\|_2^2 + O(\varepsilon^2 \|\boldsymbol{\nabla} f(\boldsymbol{x}^*)\|_2^2)}_{<0} \\ &< f(\boldsymbol{x}^*) \end{aligned}$$

Which means that $\boldsymbol{x}^*$ is not a local optimum. Therefore, the gradient must be equal to zero for any local optimum $\boldsymbol{x}^*$.

We will prove (b) using the first-order criterion for convexity:

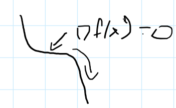
$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \cancel{(\boldsymbol{y} - \boldsymbol{x})^T \boldsymbol{\nabla} f(\boldsymbol{x})} = 0$$

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



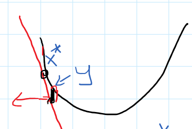
a) x^* local min. $\xRightarrow{\times} \nabla f(x^*) = 0$

$A \Rightarrow B$
$\neg B \Rightarrow \neg A$



$\forall y \in N_\varepsilon(x^*) = (x^* - \varepsilon, x^* + \varepsilon)$
$f(y) \geq f(x^*)$

If $\nabla f(x^*) \neq 0 \Rightarrow$ $\boxed{x^* \text{ is not a local min}}$



$y = x^* - \varepsilon \nabla f(x^*)$ for small enough $\varepsilon > 0$

$f(y) \leq f(x^*)$

$O(\varepsilon^2 \|\nabla f(x^*)\|^2) \leq C \cdot \varepsilon^2 \cdot \|\nabla f(x)\|^2$

for some $C > 0$

$-\varepsilon \|\nabla f(x)\|^2 < C \cdot \varepsilon^2 \|\nabla f(x)\|^2$

for small enough ε, $\varepsilon < \frac{1}{C}$

x^* is Global min.

$\forall y: \; f(y) \geq f(x^*)$