Introduction to Machine Learning Course

Dry 4 – Optimization

Submitted <u>individually</u> by Tuesday, 08.06, at 23:59. Each day of delay costs 5 points.

You may answer in Hebrew or English and write on a computer or by hand (but be clear).

Please submit a PDF file named like your ID number, e.g., *123456789.pdf.*

Bonus (maximal grade is 100): Writing on a computer (using LyX/LaTeX, Word + Equation tool, etc.) = 3 pts.

Notice: in most of the following questions, the answers should consist of 1-3 sentences and/or equations.

1. Recall the optimization formulations of Soft-SVM:

$$\boxed{\underset{w\in\mathbb{R}^d}{\operatorname{argmin}} \frac{1}{m}\sum_{i=1}^{m} \max\{0,1-y_i \cdot w^\top x_i\} + \lambda\|w\|_2^2}$$

1.1. Let $f, g: C \to \mathbb{R}$ be two convex functions. Prove that $q(z) \triangleq \max\{f(z),\ g(z)\}$ is convex w.r.t z.

1.2. Conclude that the hinge function $\ell_{hinge}(z) = \max\{0, 1 - z\}$ is convex w.r.t z.

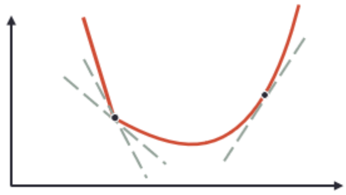
1.3. Using a rule from Lecture 07, conclude that $\max\{0, 1 - y_i \boldsymbol{w}^\top \boldsymbol{x}_i\}$ is convex w.r.t $\boldsymbol{w}$.

1.4. Let $f: C \to \mathbb{R}$ be a convex function and let $\alpha \in \mathbb{R}_{\geq 0}$. Prove that $u(z) = \alpha f(z)$ is convex w.r.t z.

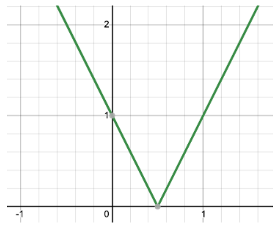
1.5. Using the above (and properties from Tutorial 07), conclude that the Soft-SVM optimization problem is convex w.r.t $\boldsymbol{w}$.

As we see in Tutorial 08, subgradients generalize gradients to convex functions which are not necessarily differentiable. Notice: you can solve this exercise even before watching Tutorial 08.

Definition: the set of subgradients of $f: V \to \mathbb{R}$ at point $\boldsymbol{u} \in V$ is:



$$\partial f(\boldsymbol{u}) \triangleq \{\boldsymbol{q} \in V | \forall \boldsymbol{v} \in V: f(\boldsymbol{v}) \geq f(\boldsymbol{u}) + \boldsymbol{q}^{\top}(\boldsymbol{v} - \boldsymbol{u})\}.$$



2. Let $f(x) = |2x - 1|$ and define its (sub) derivative $\frac{\partial}{\partial x} f(x) = \begin{cases} 2 & x > 0.5 \\ -2 & x < 0.5 \\ 0 & x = 0.5 \end{cases}$.

   2.1. Is f convex? No need to explain.

   2.2. Using the above definition, prove that 0 is indeed a sugbgradient at the point 0.5. That is, $0 \in \partial f(0.5)$.

   2.3. Set a learning rate of 0.25 and a starting point $x_0 = -1$.
   Running subgradient descent, will the algorithm converge to a minimum?
   Prove you answer by filling a table like we did in Tutorial 07 using as many rows as needed.

i	$\boldsymbol{x_i}$	$\boldsymbol{\frac{\partial}{\partial x} f(x_i)}$
0	-1	
1		
$\vdots$		

Notice: the magnitude of steps we take is always $\left|\eta \frac{\partial}{\partial x} \boldsymbol{f}(x_i)\right| = 2\eta$

   2.4. Let $\eta \in \mathbb{R}_{>0}$ be an unknown learning rate and set a starting point $x_0 = -1$.
   Write a necessary and sufficient condition (אמ"מ) for the algorithm to converge to the minimum.

   2.5. Based on your answers for the sub-questions above, does subgradient descent for <u>non-differentiable</u> convex functions converge (meaning $\lim_{i \to \infty} x_i = c$) for a small enough (fixed) η, like gradient descent for <u>differentiable</u> convex functions does (Lecture 07)? Explain briefly.