Home assignment № 4

The solutions to the following tasks can be submitted in the hand-written form scanned in PDF format. However, in this case the student is responsible for readability of the submitted text. The preferable way to prepare solutions is LaTeX or MS Word or any other tools for nice representation of equations. The following template `https://www.overleaf.com/read/vknkchxdwsmk` and tutorial `https://www.overleaf.com/learn/latex/Tutorials` can help in preparing solutions in LaTeX.

1. (2 pts) What claims from below list are correct and what are incorrect and why?
   - ☐ Any convex function is smooth
   - ☐ Any strongly convex function has unique global minimum
   - ☐ If convex function is bounded below, then if has a unique point of minimum $\mathbf{x}^*$
   - ☐ A strongly convex function is always differentiable
2. (7 pts) What functions below are convex or concave and why?
   - ☐ $f(\mathbf{x}) = \sup_{\mathbf{y} \in C} \langle \mathbf{y}, \mathbf{x} \rangle$, where C is some given set
   - ☐ $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|$, where $\|\cdot\|$ is an arbitrary norm
   - ☐ $f(\mathbf{x}) = \min\limits_{i=1,\dots,n} x_i$
   - ☐ $f(\mathbf{x}) = -\left(\prod_{i=1}^n x_i\right)^{1/n}$, $\mathrm{dom}\, f = \mathbb{R}^n_+$
   - ☐ $f(\mathbf{w}) = \sum\limits_{i=1}^m \log(1 + e^{-y_i \mathbf{w}^\top \mathbf{x}_i}) + \frac{1}{2}\|\mathbf{w}\|_2^2$, where $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}$. This function is basic loss for binary classification problem.
   - ☐ $f(\mathbf{X}, \mathbf{Y}) = \|\mathbf{A} - \mathbf{X}\mathbf{Y}\|_F^2$, where $\mathbf{X} \in \mathbb{R}^{m \times p}$, $\mathbf{Y} \in \mathbb{R}^{p \times n}$. The notation $\|\cdot\|_F$ means Frobenius norm that is computed as follows: $\|\mathbf{X}\|_F^2 = \sum_{i,j} x_{ij}^2$. The function f is the key ingredient of the loss in matrix factorization model used in the recommender systems. The matrix $\mathbf{A}$ is binary and represents the history of user-item interactions. Note that the convexity of f means that it is convex w.r.t. both arguments simultaneously.
   - ☐ $f(\mathbf{W}_1, \mathbf{W}_2) = \|\mathbf{W}_1 \max(\mathbf{W}_2\mathbf{x}, 0)\|_2$, where max is elementwise function here. Vector $\mathbf{x}$ is given. The function inside the 2-norm is the toy instance of the DeepReLU neural network. Note that the convexity of f means that it is convex w.r.t. $\mathbf{W}_1, \mathbf{W}_2$ simultaneously.
3. (2 pts) What claims from below list are correct, what are incorrect and why?
   - ☐ Lipschitz constant of gradient bounds from above the norm of hessian
   - ☐ Lipschitz constant of gradient bounds from above the absolute values of function
   - ☐ Lipschitz constant of function bounds from above the norm of hessian
   - ☐ Lipschitz constant of function bounds from above the norm of gradient