MATH3431 Machine Learning and Neural Networks III — Epiphany term 2023

Exercise sheet

Lecturer: Georgios P. Karagiannis — georgios.karagiannis@durham.ac.uk

Part 1. Stochastic learning

Exercise 1. ($\star$)Let $f:\mathbb{R}^d\rightarrow\mathbb{R}$ such that $f(w)=g(<w,x>+y)$ or some $x\in\mathbb{R}^d$, $y\in\mathbb{R}$. If $g$is convex function then $f$is convex function.

Solution. Let $u,v\in\mathbb{R}^d$ and $a\in[0,1]$. It is

$$\begin{aligned}
f(\alpha u+(1-\alpha)v)= & g(<\alpha u+(1-\alpha)v,x>+y)\\
= & g(<\alpha u,x>+<(1-\alpha)v,x>+y)\\
= & g(\alpha(<u,x>+y)+(1-\alpha)(<v,x>+y)) && y=\alpha y+(1-\alpha)y\\
\leq & \alpha g(<u,x>+y)+(1-\alpha)g(<v,x>+y) && (g\text{ is convex})\\
= & \alpha f(u)+(1-\alpha)f(v)
\end{aligned}$$

Exercise 2. ($\star$)Let functions g_1 be ρ_1-Lipschitz and g_2 be ρ_2-Lipschitz. Then f with $f(x)=g_1(g_2(x))$ is $\rho_1\rho_2$-Lipschitz.

Solution.

$$\begin{aligned}
|f(w_1)-f(w_2)|= & |g_1(g_2(w_1))-g_1(g_2(w_2))|\\
\leq & \rho_1|g_2(w_1)-g_2(w_2)|\\
\leq & \rho_1\rho_2|w_1-w_2|
\end{aligned}$$

Exercise 3. ($\star$)Let $f:\mathbb{R}^d\rightarrow\mathbb{R}$ with $f(w)=g(\langle w,x\rangle+y)$ $x\in\mathbb{R}^d$ and $y\in\mathbb{R}$. Let $g:\mathbb{R}\rightarrow\mathbb{R}$ be a β-smooth function. Then f is a $\left(\beta\|x\|^2\right)$-smooth.

Hint:: You may use Cauchy-Schwarz inequality $\langle y,x\rangle\leq\|y\|\|x\|$

$$
\begin{aligned}
f(v) =& g(\langle w, x\rangle + y) \\
\leq& g(\langle w, x\rangle + y) + g'(\langle w, x\rangle + y)\langle v-w, x\rangle + \frac{\beta}{2}(\langle v-w, x\rangle)^2 && (g \text{ is smooth}) \\
\leq& g(\langle w, x\rangle + y) + g'(\langle w, x\rangle + y)\langle v-w, x\rangle + \frac{\beta}{2}(\|v-w\|\,\|x\|)^2 && (\text{Cauchy-Schwatz inequality}) \\
=& f(w) + \langle \nabla f(w), v-w\rangle + \frac{\beta\|x\|^2}{2}\|v-w\|^2
\end{aligned}
$$

Exercise 4. $(\star) f: S \to \mathbb{R}$ is ρ-Lipschitz over an open convex set S if and only if for all $w \in S$ and $v \in \partial f(w)$ it is $\|v\| \leq \rho$.

Hint:: You may use Cauchy-Schwarz inequality $\langle y, x\rangle \leq \|y\|\,\|x\|$

Solution. $\Longrightarrow$ Let $f: S \to \mathbb{R}$ be ρ-Lipschitz over convex set S, $w \in S$ and $v \in \partial f(w)$.

- Since S is open we get that there exist $\epsilon > 0$ such as $u := w + \epsilon\frac{v}{\|v\|}$ where $u \in S$. So $\langle u-w, v\rangle = \epsilon\|v\|$ and $\|u-w\| = \epsilon$.
- From the subgradient definition we get

$$f(u) - f(w) \geq \langle u-w, v\rangle = \epsilon\|v\|$$

- From the Lipschitzness of $f(\cdot)$ we get

$$f(u) - f(w) \geq \rho\|u-w\| = \rho\epsilon$$

Therefore $\|v\| \leq \rho$.

Proof. $\Longleftarrow$ It is for all $w \in S$ and $v \in \partial f(w)$ it is $\|v\| \leq \rho$. □

- For any $u \in S$, it is

$$
\begin{aligned}
f(w) - f(u) \leq& \langle v, w-u\rangle && (\text{ because } v \in \partial f(w)\ ) \\
\leq& \|v\|\,\|w-u\| && \text{by Cauchy-Schwarz inequality} \\
\leq& \rho\|w-u\| && \text{because } \|v\| \leq \rho
\end{aligned}
\tag{1}
$$

- Similarly it results $u, w \in S$

$$f(w) - f(u) \leq \langle v, u-w\rangle\|v\| \leq \|v\|\,\|u-w\| \leq \rho\|u-w\|$$

from (**??**) because w, u can be swaped in (**??**) as they both are any values in S.

Exercise 5. $(\star)$Let $g_1(w), ..., g_r(w)$ be r convex functions, and let $g(\cdot) = \max_{\forall j}(g_j(\cdot))$. Show that for some w it is $\nabla g_k(w) \in \partial g(w)$ where $k = \arg\max_j(g_j(w))$ is the index of function $g_j(\cdot)$ presenting the greatest value at w.

Since g_j is convex, for all u

$$g_j(u) \geq g_j(w) + \langle u - w, \nabla g_j(w) \rangle$$

However $g(u) = \max_{\forall j}(g_j(u)) \geq g_j(u)$ for any j, and $g(w) = g_j(w)$ at w. Then

$$g(u) \geq g(w) + \langle u - w, \nabla g_j(w) \rangle$$

Then by the definition of the sub-gradient $\nabla g_j(w) \in \partial g(w)$
