MATH3431 Machine Learning and Neural Networks III — Epiphany term 2025

Exercise sheet

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

Part 1. Convex learning problems

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

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

Exercise 2. ($\star$)Let functions g_{1} be ρ_{1}-Lipschitz and g_{2} be ρ_{2}-Lipschitz. Then, show that, f with $f\left(x\right)=g_{1}\left(g_{2}\left(x\right)\right)$ is $\rho_{1}\rho_{2}$-Lipschitz.

Solution.

$$
\begin{aligned}
\left|f\left(w_{1}\right)-f\left(w_{2}\right)\right| =& \left|g_{1}\left(g_{2}\left(w_{1}\right)\right)-g_{1}\left(g_{2}\left(w_{2}\right)\right)\right| \\
\leq& \rho_{1}\left|g_{2}\left(w_{1}\right)-g_{2}\left(w_{2}\right)\right| \\
\leq& \rho_{1}\rho_{2}\left|w_{1}-w_{2}\right|
\end{aligned}
$$

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

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

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

Exercise 4. ($\star$)Show that $f : S \to \mathbb{R}$ 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

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

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

$\Longleftarrow$ 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

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

- 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 (1) because w, u can be swaped in (1) as they both are any values in S.

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

Solution. Since g_k is convex, for all u

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

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

$$\begin{aligned} f(u) \geq & g_k(u) \\ \geq & g_k(w) + \langle u - w, \nabla g_k(w) \rangle \\ = & f(w) + \langle u - w, \nabla g_k(w) \rangle \end{aligned}$$

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

Exercise 6. ($\star$)Consider the regression learning problem $(\mathcal{H}, \mathcal{Z}, \ell)$ with predictor rule $h(x) = \langle w, x \rangle$ labeled by some unknown parameter $w \in \mathcal{W}$, loss function $\ell(w, (x, y)) = (\langle w, x \rangle - y)^2$, feature $x \in \mathcal{X}$, and target $y \in \mathbb{R}$. Let $\mathcal{W} = \mathcal{X} = \left\{ \omega \in \mathbb{R}^d : |\omega| \leq \rho \right\}$ for some $\rho > 0$.

(1) Show that the resulting learning problem is Convex-Lipschitz-Bounded learning problem.
(2) Specify the parameters of Lipschitnzess.

Solution. According to the definitions given in the lecture:

- Convex-Lipschitz-Bounded Learning Problem $(\mathcal{H}, \mathcal{Z}, \ell)$ with parameters ρ, and B, is called the learning problem whose the hypothesis class $\mathcal{H}$ is a convex set, for all $w \in \mathcal{H}$ it is $\|w\| \leq B$, and the loss function $\ell(\cdot, z)$ is convex and ρ-Lipschitz function for all $z \in \mathcal{Z}$.

I have:

Convexity: The function $g : \mathbb{R} \to \mathbb{R}$, defined by $g(a) = a^2$ is convex convex. Eg. $\frac{\mathrm{d}^2}{\mathrm{d}a^2} g(a) = 1 \geq 0$ is non-negative. The convexity of $\ell(w, z = (x, y))$ for all z follows as a composition of g with a linear function.

Lipschitzness: The function $g(a) = a^2$ is 1-Lipschitz since It is

$$|g(a_2) - g(a_1)| = \left| a_2^2 - a_1^2 \right| = |(a_2 + a_1)(a_2 - a_1)| \leq 2\rho (a_2 - a_1) = 2\rho |a_2 - a_1|$$

Hence because $|x| \leq \rho$, $g(a)$ is $2\rho^2$-Lipschitz as a composition.

Boundness: The norm of each hypothesis w is bounded by ρ according to the assumptions.

Therefore,

(1) the learning problem under consideration is a Convex-Lipschitz-Bounded learning problem.
(2) the parameter of Lipschitzness is $2\rho^2$.

Exercise 7. ($\star$)If $f$is λ-strongly convex and u is a minimizer of f then for any w

$$f(w) - f(u) \geq \frac{\lambda}{2} \|w - u\|^2$$

Hint:: Use the definition, and set $\alpha \to 0$.

Solution.

Exercise 8. ($\star$) Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex and β-smooth function.

(1) Show that for $v, w \in \mathbb{R}^d$

$$f(v) - f(w) \in \left(\langle \nabla f(w), v - w \rangle, \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|^2 \right)$$

(2) Show that for $v, w \in \mathbb{R}^d$ such that $v = w - \frac{1}{\beta} \nabla f(w)$, it is

$$\frac{1}{2\beta} \|\nabla f(w)\|^2 \leq f(w) - f(v)$$

(3) Additionally assume that $f(x) > 0$ for all $x \in \mathbb{R}^d$. Show that for $w \in \mathbb{R}^d$,

$$\|\nabla f(w)\| \leq \sqrt{2\beta f(w)}$$

Solution.

(1) If $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth then it is

$$f(v) \leq f(w) + \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|^2$$
$$f(v) - f(w) \leq \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|^2$$

If it is convex then it is

$$f(v) \geq f(w) + \langle \nabla f(w), v - w \rangle$$
$$f(v) - f(w) \geq \langle \nabla f(w), v - w \rangle$$

Together these conditions imply upper and lower bounds

$$f(v) - f(w) \in \left(\langle \nabla f(w), v - w \rangle, \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|^2 \right)$$

(2) For $v, w \in \mathbb{R}^d$ such that $v = w - \frac{1}{\beta} \nabla f(w)$, it is

$$f(v) \leq f(w) + \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|_2^2 \quad \text{(due to smoothness)}$$
$$\iff f(w) - f(v) \leq f(w) - f(v)$$
$$\iff \langle \nabla f(w), v - w \rangle + \frac{\beta}{2} \|v - w\|_2^2 \leq f(w) - f(v)$$
$$\iff \left\langle \nabla f(w), \frac{1}{\beta} \nabla f(w) \right\rangle + \frac{\beta}{2} \left\| \frac{1}{\beta} \nabla f(w) \right\|_2^2 \leq f(w) - f(v)$$
$$\iff \frac{1}{2\beta} \|\nabla f(w)\|^2 \leq f(w) - f(v)$$
$$\|\nabla f(w)\|^2 \leq 2\beta (f(w) - f(v))$$

as $f(\cdot) \geq 0$

$$\|\nabla f(w)\|^2 \leq 2\beta f(w)$$

(3) From part 2, this is obvious because $f(x) > 0$ for all $x \in \mathbb{R}^d$, as

$$\|\nabla f(w)\|^2 \leq 2\beta f(w) \Leftrightarrow \|\nabla f(w)\| \leq \sqrt{2\beta f(w)}$$

Exercise 9. ($\star$)Let $f : \mathbb{R}^d \to \mathbb{R}$ be a λ-strongly convex function. Assume that w^* is a minimizer of f i.e.

$$w^* = \arg\min_w \{f(w)\}$$

Show that for any $w \in \mathbb{R}^d$ it holds

$$f(w) - f(w^*) \geq \frac{\lambda}{2} \|w - w^*\|^2$$

Hint: Use the definition of λ-strongly convex function, properly rearrange it, and ...

Solution. We use the definition of λ-strongly convex function; i.e. for all w, u, and $\alpha \in (0, 1)$ we have

$$f(aw + (1-\alpha)u) \leq af(w) + (1-\alpha)f(u) - \frac{\lambda}{2}\alpha(1-\alpha)\|w - u\|^2 \Leftrightarrow$$
$$\frac{f(aw + (1-\alpha)u) - f(u)}{\alpha} \leq f(w) + f(u) - \frac{\lambda}{2}(1-\alpha)\|w - u\|^2$$

For $u = w^*$ it is

$$\frac{f(aw + (1-\alpha)w^*) - f(w^*)}{\alpha} \leq f(w) + f(w^*) - \frac{\lambda}{2}(1-\alpha)\|w - w^*\|^2$$

When $a \to 0$

$$\frac{\lambda}{2}\alpha(1-\alpha)\|w - w^*\|^2 \to 0$$

I know that w^* is the minimizer of f. So 0 is the minimizer of g with $g(a) = f(aw + (1-\alpha)w^*)$ hencewhen $a \to 0$

$$\frac{f(aw + (1-\alpha)w^*) - f(w^*)}{\alpha} \to \frac{\mathrm{d}}{\mathrm{d}a} g(a)\bigg|_{a=0}$$

So

$$0 \leq f(w) + f(w^*) - \frac{\lambda}{2}\|w - w^*\|^2$$

which concludes the proof.

Exercise 10. ($\star$)Show that the function $J(x;\lambda) = \lambda\|x\|^2$ is 2λ-strongly convex

Solution. We just need to check that for all w, u, and $\alpha \in (0, 1)$ we have

$$J(aw + (1-\alpha)u;\lambda) \leq aJ(w;\lambda) + (1-\alpha)J(u;\lambda) - \frac{2\lambda}{2}\alpha(1-\alpha)\|w - u\|^2 \Longleftrightarrow$$
$$\|aw + (1-\alpha)u\|_2^2 \leq a\|w\|_2^2 + (1-\alpha)\|u\|_2^2 - a(1-\alpha)\|w - u\|_2^2 \Longleftrightarrow 0 \leq 0$$
