

Handout 1: Elements of convex learning problems

Lecturer & author: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

Aim. To introduce elements of convexity, Lipschitzness, and smoothness that can be used for the analysis of stochastic gradient related learning algorithms.

Reading list & references:

- Bishop, C. M. (2006). Pattern recognition and machine learning. New York: Springer.
- Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

1. MOTIVATIONS

Note 1. Introducing convexity and smoothness in the learning problems makes easier the (theoretical) analysis of the problem and its solution.

Note 2. Most of the ML problems discussed in the course (eg, Artificial neural networks, Gaussian process regression) are usually non-convex.

Note 3. Extensions of handle non-convex problems as below can be done via surrogates –to be discussed.

2. CONVEX LEARNING PROBLEM

Definition 4. Convex learning problem is a learning problem $(\mathcal{H}, \mathcal{Z}, \ell)$ is the learning problem that the hypothesis class \mathcal{H} is a convex set, and the loss function ℓ is a convex function for each example $z \in \mathcal{Z}$.

Example 5. Multiple linear regression $\langle w, x \rangle \rightarrow y$ with $y \in \mathbb{R}$, hypothesis class $\mathcal{H} = \{w \in \mathbb{R}^d\}$ and loss $\ell(w, (x, y)) = (\langle w, x \rangle - y)^2$ with

$$w^* = \arg \min_{\forall w} \mathbb{E} (\langle w, x \rangle - y)^2$$

or

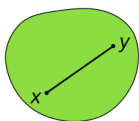
$$w^{**} = \arg \min_{\forall w} \frac{1}{m} \sum_{i=1}^m (\langle w, x_i \rangle - y_i)^2$$

is a convex learning problem for arguments discussed below.

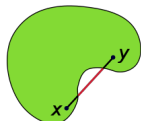
3. CONVEXITY

Definition 6. A set C is convex if for any $u, v \in C$, the line segment between u and v is contained in C . Namely,

- for any $u, v \in C$ and for any $\alpha \in [0, 1]$ we have that $\alpha u + (1 - \alpha) v \in C$.



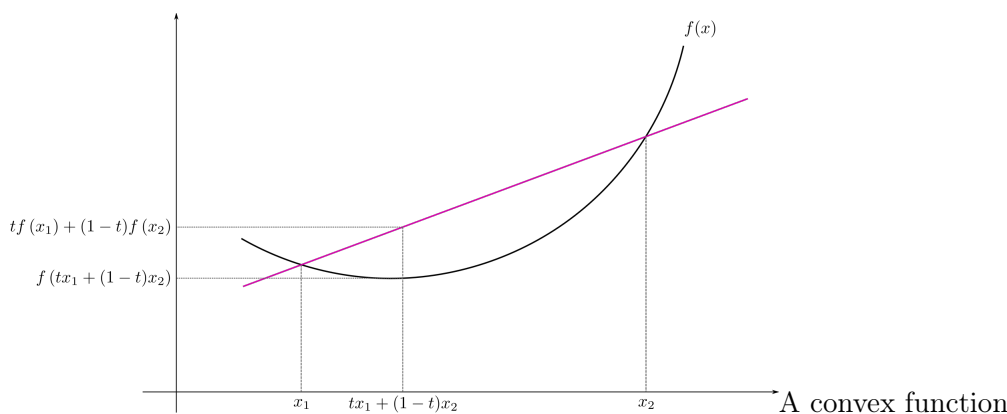
A convex set



A non-convex set

Definition 7. Let C be a convex set. A function $f : C \rightarrow \mathbb{R}$ is convex function if for any $u, v \in C$ and for any $\alpha \in [0, 1]$

$$f(\alpha u + (1 - \alpha)v) \leq \alpha f(u) + (1 - \alpha)f(v)$$



A convex function

Example 8. The function $f : \mathbb{R} \rightarrow \mathbb{R}_+$ with $f(x) = x^2$ is convex function. For any $u, v \in C$ and for any $\alpha \in [0, 1]$ it is

$$(\alpha u + (1 - \alpha)v)^2 \leq \alpha^2(u)^2 + (1 - \alpha)^2(v)^2 + 2\alpha u(1 - \alpha)v \leq \alpha(u)^2 + (1 - \alpha)(v)^2$$

Proposition 9. Every local minimum of a convex function is the global minimum.

Proposition 10. Let $f : C \rightarrow \mathbb{R}$ be convex function. The tangent of f at $w \in C$ is below f , namely

$$\forall u \in C \quad f(u) \geq f(w) + \langle \nabla f(w), u - w \rangle$$

Proposition 11. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $f(w) = g(\langle w, x \rangle + y)$ or some $x \in \mathbb{R}^d$, $y \in \mathbb{R}$. If g is convex function then f is convex function.

Proof. See Exercise 1 in the Exercise sheet. □

Example 12. Consider the regression problem $x \mapsto y$ with $x \in \mathbb{R}^d$, $y \in \mathbb{R}$ and predictor $h(x) = \langle w, x \rangle$. The risk $R(w) = (\langle w, x \rangle + y)^2$ because $g(a) = (a)^2$ is convex and Proposition 11.

Example 13. Let $f_j : \mathbb{R}^d \rightarrow \mathbb{R}$ convex functions for $j = 1, \dots, r$. Then:

- (1) $g(x) = \max_{\forall j} (f_j(x))$ is a convex function
- (2) $g(x) = \sum_{j=1}^r w_j f_j(x)$ is a convex function where $w_j > 0$

Solution.

(1) For any $u, v \in \mathbb{R}^d$ and for any $\alpha \in [0, 1]$

$$\begin{aligned}
g(\alpha u + (1 - \alpha)v) &= \max_{\forall j} (f_j(\alpha u + (1 - \alpha)v)) \\
&\leq \max_{\forall j} (\alpha f_j(u) + (1 - \alpha)f_j(v)) && (f_j \text{ is convex}) \\
&\leq \alpha \max_{\forall j} (f_j(u)) + (1 - \alpha) \max_{\forall j} (f_j(v)) && (\max(\cdot) \text{ is convex}) \\
&\leq \alpha g(u) + (1 - \alpha)g(v)
\end{aligned}$$

(2) For any $u, v \in \mathbb{R}^d$ and for any $\alpha \in [0, 1]$

$$\begin{aligned}
g(\alpha u + (1 - \alpha)v) &= \sum_{j=1}^r w_j f_j(\alpha u + (1 - \alpha)v) \\
&\leq \alpha \sum_{j=1}^r w_j f_j(u) + (1 - \alpha) \sum_{j=1}^r w_j f_j(v) && (f_j \text{ is convex}) \\
&\leq \alpha g(u) + (1 - \alpha)g(v)
\end{aligned}$$

Example 14. $g(x) = |x|$ is convex according to Example 13, as $g(x) = |x| = \max(-x, x)$.

4. LIPSCHITZBNESS