# Homework 2

2021-31645-T1 Machine Learning
Emmanouil Palaiologos

November 1, 2021

## Convexity

Assuming that the two d-ary functions $f$ and $g$ are **convex**, then for a pair of $d \times 1$ vectors $x_1, x_2 \in \mathbb{R}^d$ and a constant $\lambda \in [0, 1]$ we write:

$$\begin{cases} f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \\ g\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda g(x_1) + (1-\lambda) g(x_2) \end{cases} \tag{1}$$

### Property 1

**For any constant $\alpha > 0$: $h(x) = a \cdot f(x)$ is also convex**

*Proof.* According to the respective definition, in order for $h(x)$ to be convex, it's necessary to satisfy the following inequality for any $\lambda \in [0, 1]$:

$$h\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda h(x_1) + (1-\lambda) h(x_2) \tag{2}$$

$$\implies \alpha f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda \alpha f(x_1) + (1-\lambda) \alpha f(x_2) \tag{3}$$

$$\implies f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \quad \text{(since } \alpha > 0) \tag{4}$$

Which is already confirmed by (1) q.e.d.

### Property 2

**The sum of $f$ and $g$: $h(x) = f(x) + g(x)$ is also a convex function.**

*Proof.* For $h(x)$ and $\forall \lambda \in [0, 1]$ it holds that :

$$h\big(\lambda x_1 + (1-\lambda)x_2\big) = f\big(\lambda x_1 + (1-\lambda)x_2\big) + g\big(\lambda x_1 + (1-\lambda)x_2\big) \tag{5}$$

$$\leq \lambda f(x_1) + (1-\lambda) f(x_2) + \lambda g(x_1) + (1-\lambda) g(x_2) \quad \text{(using 1)} \tag{6}$$

$$= \lambda[f(x_1) + g(x_1)] + (1-\lambda)[f(x_2) + g(x_2)] \tag{7}$$

$$= \lambda h(x_1) + (1-\lambda) h(x_2) \tag{8}$$

q.e.d.

## Property 3

**The pointwise maximum of $f$ and $g$: $h(x) = \max\{f(x), g(x)\}$ is also a convex function.**

*Proof.* Let $\phi_i$ $(i \in [1,2])$ denote any of the two convex functions inside the argument of *"max"* ($f$ or $g$). Then using (1) we could claim $\forall i$ that:

$$\phi_i\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda\phi_i(x_1) + (1-\lambda)\phi_i(x_2) \tag{9}$$

$$\leq \sup_{i\in[1,2]} \left(\lambda\phi_i(x_1) + (1-\lambda)\phi_i(x_2)\right) \tag{10}$$

$$= \max\left\{\lambda\phi_i(x_1) + (1-\lambda)\phi_i(x_2)\right\} \tag{11}$$

$$= \lambda \max\left\{\phi_i(x_1)\right\} + (1-\lambda)\max\left\{\phi_i(x_2)\right\} \tag{12}$$

$$= \lambda h(x_1) + (1-\lambda)h(x_2) \tag{13}$$

Since all of the above apply to any of the $f$ or $g$, it follows that the upper bound $\lambda h(x_1) + (1-\lambda)h(x_2)$ also applies to the maximum function $h$.

q.e.d.

## Property 4

**The composite function $h(x) = f(Ax + b)$ of the convex function $f$, where $A$ is a $d\times d$ matrix and $b$ a $d\times 1$ vector, is also a convex function.**

*Proof.*

- The *affine* function $g(x) = A\cdot x + b$ is both *convex* and *concave* ($\forall\lambda \in [0,1]$ it satisfies the exact equality ):

$$g\big(\lambda x_1 + (1-\lambda)x_2\big) = A\cdot[\lambda x_1 + (1-\lambda)x_2] + b \tag{14}$$

$$= \lambda A x_1 + A x_2 - \lambda A x_2 + b \tag{15}$$

$$= \underbrace{\lambda A x_1} + \underbrace{A x_2 - \lambda A x_2 + b - \lambda b} + \underbrace{\lambda b} \tag{16}$$

$$= \lambda\cdot(A x_1 + b) + b\cdot(1-\lambda) + A x_2\cdot(1-\lambda) \tag{17}$$

$$= \lambda(A x_1 + b) + (1-\lambda)(A x_2 + b) \tag{18}$$

$$= \lambda g(x_1) + (1-\lambda)g(x_2) \tag{19}$$

- For $h$ we apply the above equality(19), using the convexity of $f$ (eq.1):

$$h\big(\lambda x_1 + (1-\lambda)x_2\big) = f\Big(g(\lambda x_1 + (1-\lambda)x_2)\Big) \tag{20}$$

$$= f\Big(\lambda g(x_1) + (1-\lambda)g(x_2)\Big) \tag{21}$$

$$\leq \lambda f\Big(g(x_1)\Big) + (1-\lambda)f\Big(g(x_2)\Big) \tag{22}$$

$$= \lambda h(x_1) + (1-\lambda)h(x_2) \tag{23}$$

q.e.d.