# AI3001: Advanced Topics in Machine Learning
# Classroom Quiz 2
# September 27, 2022

*Instructions:*

- *All the problems are compulsory.*
- *Total duration of the quiz is 30 minutes.*

**Problem 1** (1+2+2= 5 marks)**.** Compute the Bregman Divergence of

1. $R(x) = \frac{1}{2}||x||_2^2$ for $x \in \mathbb{R}^d$
2. $R(x) = -2\sum_{i=1}^{d} \sqrt{x_i}$ for $x \in (0, +\infty)^d$
3. $R(x) = \sum_{i=1}^{d} x_i(\log(x_i) - 1)$ for $x \in \Delta_d$. Use $0\log(0) = 0$.

**Problem 2** (1+1+1+2 marks)**.** Prove or disprove the following statements.

1. Bregman divergence $B_R(x||y)$ is convex in $x$ even if $R$ is not convex.
2. If $f$ is $\alpha$-strongly convex and $g$ is convex then $f + g$ is $\alpha$-strongly convex.
3. Dual norm of 1-norm is $\infty$-norm and vice-versa. Here $||x||_1 = \sum_{i=1}^{d} |x_i|$ and $||x||_\infty = \max\{|x_1|, |x_2|, \cdots, |x_d|\}$
4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Then $f$ is strongly convex (with respect to norm $||.||$) if and only if $g$ defined as $g(x) = f(x) - \frac{1}{2}||x||^2$ is convex.

**Solution to Problem 1.**

1.

$$\begin{aligned}B_R(x||y) &= \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle y, x-y\rangle \\ &= \frac{1}{2}||x||_2^2 + \frac{1}{2}||y||_2^2 - \langle y, x\rangle = \frac{1}{2}||x-y||_2^2\end{aligned}$$

2.

$$\begin{aligned}B_R(x||y) &= -2\sum_{i=1}^{d}\sqrt{x_i} + 2\sum_{i=1}^{d}\sqrt{y_i} + \sum_{i=1}^{d}\frac{x_i - y_i}{\sqrt{y_i}} \\ &= \sum_{i=1}^{d}\frac{x_i + y_i - 2\sqrt{x_i y_i}}{\sqrt{y_i}} = \sum_{i=1}^{d}\frac{(\sqrt{x_i} - \sqrt{y_i})^2}{\sqrt{y_i}}\end{aligned}$$

3.

$$\begin{aligned}B_R(x||y) &= \sum_{i=1}^{d} x_i \log(x_i) - \sum_{i=1}^{d} x_i - y_i \log(y_i) + \sum_{i=1}^{d} y_i - \sum_{i=1}^{d}\log(y_i)(x_i - y_i) \\ &= \sum_{i=1}^{d} x_i \log(x_i) - \sum_{i=1}^{d} x_i \log(y_i) = \sum_{i=1}^{d} x_i \log(\frac{x_i}{y_i}) = KL(x||y).\end{aligned}$$

■

**Solution of Question 2.**

1. Take $R(x) = \log(x)$ for $x > 0$. We have $B_R(x||y) = \log(x) - \log(y) - \langle 1/y, x-y\rangle$. Clearly $B_R(x||y)$ is not convex since we have $\nabla_x^2 B_R(x||y) = -1/x^2 < 0$.

2. NOte that $\nabla(f+g)(x) = \nabla f(x) + \nabla f(x)$. We have from strong convexity of $f$ and convexity of $g$ that

$$(\nabla f(y) - \nabla f(x))^T(y-x) \geq \alpha \tag{1}$$

$$(\nabla g(y) - \nabla g(x))^T(y-x) \geq 0 \tag{2}$$

adding (1) and (2),

$$\langle \nabla(f+g)(y) - \nabla(f+g)(x), y-x \rangle \geq \alpha$$

.

3. Dual norm is given by $||y||^* = \sup_{||x|| \leq 1} \langle x, y \rangle$. Let $i^* = \arg\max_i y_i$. If $y_{i^*} \geq 0$ we take $x'_{i^*} = 1$ and $x'_i = 0$ for $i \neq i^*$. We now show that $x'$ is the supremum. For contradiction let $x \neq x'$ be any other vector such that $x_i \geq 0, \sum_i x_i \leq 1$ and $\sum_{i=1}^n x_i y_i > y_{i^*}$. We have

$$\sum_{i \neq i^*} x_i y_i > (1 - x_{i^*}) y_{i^*}.$$

However, since $\sum_i x_i \leq 1$ we have $(1 - x_{i^*}) \geq \sum_{i \neq i^*} x_i$. We get

$$\sum_{i \neq i^*} x_i y_i > (1 - x_{i^*}) y_{i^*} \geq \sum_{i \neq i^*} x_i y_{i^*} \geq \sum_{i \neq i^*} x_i y_i$$

. A contradiction. The last inequality follows from the fact that $y_{i^*}$ is the maximal element. On the other hand if $y_{i^*} < 0$ multiply all $y_i$ and corresponding $x_i$'s by $-1$ and we land in the previous case. (Essentially, in this case we take $x'_{i^*} = -1$ and $x_i = 0$ for all $i \neq i^*$).

4. Not true in general. Let $f(x) = \frac{1}{\eta} \sum_{i=1}^d x_i \log(x_i)$ for some $\eta > 0$. We have that $f$ is $1/\eta$-strongly convex. However for function $g(x) = f(x) - 1/2||x||^2$ we have $\nabla^2 g(x) = \begin{pmatrix} 1/\eta x_1 - 1 & 0 & \cdots 0 \\ 0 & 1/\eta x_2 - 1 \cdots 0 \\ 0 & 0 & \cdots \; 1/\eta x_d - 1 \end{pmatrix}$ is convex only when $x_i \leq 1/\eta$ for all $i$. For instance, take $\eta \geq d+1$ and $x \in \Delta_d$ we have $1 = \sum_i x_i \leq d/\eta \leq d/(d+1) < 1$ a contradiction. However, it is true for 1-strongly convex functions as given in the lecture notes.

   The other side implication also does not hold. Let $g$ be twice differentiable convex function on $\mathbb{R}$. We have $g''(x) = f''(x) - 1 \geq 0$. This implies $f''(x) \geq 1$. HOwever, this does not imply strong convexity for $\alpha < 1$.

■