AI3001: Advanced Topics in Machine Learning
Midterm 1
September 28, 2022

Instructions:

- *The total number of marks is 25.*
- *The total duration of exam is 48 hours. Deadline is 10 am Friday, 30th Sept.*
- *Late submissions will not be considered under any circumstances.*

Problem 1. Prove the following

1. If the loss function is α-exp-concave for some $\alpha > 0$ then it is also α'-exp-concave for any $\alpha' \in (0, \alpha]$. [2 marks]
2. $e^x \leq 1 + x + x^2$ for all $x \in [0, 1]$. [2 marks]
3. Let $h : \mathbb{R} \to \mathbb{R}$ be a concave non-decreasing function and let $g : \mathcal{K} \to \mathbb{R}$ be concave. Prove that $f(x) = h(g(x))$ is concave. [2 marks]
4. $$\arg\min_{x \in \mathcal{K}} \left\{ \nabla_t^T (x - x_t) + \frac{1}{2\eta_t} ||x - x_t||_2^2 \right\} = \arg\min_{x \in \mathcal{K}} \left\{ ||x - (x_t - \eta_t \nabla_t)||_2^2 \right\} \quad \text{[2 marks]}$$

Problem 2. Consider that the loss function is bounded in range $[0, 1]$. Show that the regret of FTL algorithm is upper bounded by the number of times the *leader*[1] is changed during the sequence of plays. [2 marks]

Problem 3. A set $\mathcal{K} \in \mathbb{R}$ is called symmetric if for any $x \in \mathcal{K}$ we have $-x \in \mathcal{K}$. Define the function $||.||_\# : \mathbb{R}^d \to \mathbb{R}$ as

$$||x||_\# = \arg\min_{\alpha > 0} \left\{ \frac{1}{\alpha} x \in \mathcal{K} \right\}.$$

Show that $||.||_\#$ is a norm if and only if $\mathcal{K}$ is a convex set. [3 marks]

Problem 4 (Dual Function)**.** In class we saw that the OMD algorithm[2] performs the update in the dual space defined by the gradients. This can be interpreted (slightly more elaborately) as follows. OMD first maps the point x_t from primal space (i.e. x-space) to the dual space defined by $\nabla R(.)$ i.e. computes $\theta_t = \nabla R(x_t)$, performs the (gradient) update in gradient space i.e. computes $\theta_{t+1} = \theta_t - \eta_t \nabla_t$, maps the updated point using $(\nabla R)^{-1}$ back to the original space i.e. computes y_{t+1} such

[1] The leader at time t is a choice x_{t+1} i.e. optimal choice of the next round.

[2] Here, we will consider all assumptions we made for OMD.

that $\nabla R(y_{t+1}) = \theta_{t+1}$ i.e. $y_{t+1} = (\nabla R)^{-1}(\theta_{t+1})$ and then take the Bregman projection onto the convex set i.e. compute $x_{t+1} = \arg\min_{x \in \mathcal{K}} B_R(x||y_{t+1})$.

Define a function $R^*(\theta) = \sup_{x \in \mathbb{R}^d}(\langle x, \theta \rangle - R(x))$. The function $R^*(.)$ is called as the Fenchel dual/conjugate of R. In this question we will show that $\nabla R^* = (\nabla R)^{-1}$.

1. Write the Fenchel conjugates of below functions. [3+3 = 6 marks]
   (a) $R(x) = \log(\sum_{i=1}^{d} e^{x_i})$
   (b) $R(x) = \frac{1}{2}x^T Q x$ where $Q \in \mathbb{R}^{n \times n}$ is a positive definite matrix.
2. Show that the following two conditions are equivalent. [2 marks]
   (a) $R(u) + R^*(v) = \langle u, v \rangle$
   (b) $v = \nabla R(u)$
3. Finally, using the above result and also using the fact that $R^{**} = R$, show that $u = \nabla R^*(\nabla R(u))$ and $u = \nabla R(\nabla R^*(u))$. [4 marks]