E1 245 - Online Prediction and Learning, Aug-Dec 2019
Homework #3

1. *Strong convexity of the entropy function (Pinsker's inequality)*

   (a) Show that the negative entropy function $R(x) = \sum_{i=1}^{2} x_i \log x_i$, is 1-strongly convex with respect to the $||\cdot||_1$ norm over the 2 dimensional simplex $\Delta_2 := \{(x_1, x_2) : x_1, x_2 \geq 0, x_1 + x_2 = 1\}$.

   (b) Prove the same result in any finite number of dimensions $d \geq 2$.
   Hint: One way is to find a reduction to the $d = 2$ case. Let x and y be two vectors in Δ_d. Let $A := \{i : x_i \geq y_i\}$ be the coordinates where x dominates y. Can you somehow find two new vectors x_A and y_A in Δ_2 so that $||x - y||_1 = ||x_A - y_A||_1$ and proceed?

2. *Fenchel duality and Bregman divergence*
   Compute the Fenchel dual functions and Bregman divergences for the following functions defined over $\mathbb{R}^d$, unless stated otherwise. State the domain clearly in each case.

   (a) $F(x) = e^{x_1} + \cdots + e^{x_d}$.

   (b) $F(x) = \log(e^{x_1} + \cdots + e^{x_d})$.

   (c) $F(x) = \frac{1}{2}||x||_p^2 := \frac{1}{2}\left(\sum_{1=1}^{d} |x_i|^p\right)^{1/p}$, $p \in [1, \infty]$.

   (d) $F(x) := \sum_{i=1}^{d} x_i \log x_i - \sum_{i=1}^{d} x_i$, $x \in (0, +\infty)^d$.

3. Three point equality for Bregman divergences
   Show the following ('law of cosines') for the Bregman divergence $D_R(x, y)$ induced by a differentiable convex function $R : \mathbb{R}^d \to \mathbb{R}$:

$$\forall u, v, w \in \mathbb{R}^d : \quad D_R(u, v) + D_R(v, w) = D_R(u, w) + \langle u - v, \nabla R(w) - \nabla R(v) \rangle.$$

4. *Exponential Weights as FTRL*
   Show that executing Follow The Regularized Leader (FTRL) on the simplex $\Delta_N := \left\{(x_1, \ldots, x_N) : \sum_{i=1}^{N} x_i = 1, \forall i\, x_i \geq 0\right\}$ with the entropic regularizer[1] $R_\eta(x) := \frac{1}{\eta} \sum_{i=1}^{N} x_i \log x_i$, and linear loss functions $f_t(x) = \langle z_t, x \rangle$, is equivalent to running the Exponential Weights algorithm on N experts with loss vectors $\{z_t\}_{t \geq 1}$ and parameter η.
   Hint: There are two ways to do this. (1) You can derive this directly from first principles and the definition of the FTRL rule. (2) An alternative way is by using (a) the equivalence between FTRL and (unconstrained minimization + Bregman projection) proven in class, and (b) observing that Bregman projection wrt the regularizer R onto Δ_N is equivalent to scaling by the $||\cdot||_1$ norm.

[1] $0 \log 0$ is interpreted as 0 in the definition.