MLT Homework set 1

Due 20 September 2017 before 14:00
either via `elo.mastermath.nl` or on paper

1 Notation and Definitions

Definition 1. Fix a differentiable convex function $\phi : \mathbb{R}^k \to \mathbb{R}$. The *Bregman divergence* from $x \in \mathbb{R}^k$ to $y \in \mathbb{R}^k$ generated by ϕ is

$$B_\phi(x, y) \;=\; \phi(x) - \phi(y) - \langle x - y, \nabla\phi(y)\rangle$$

where $\langle x, y\rangle$ denotes the dot product $\sum_{i=1}^k x_i y_i$, and $\nabla\phi(y)$ is the gradient (vector of partial derivatives) of ϕ at y.

Definition 2. Let us write $[k] = \{1, \ldots, k\}$, and let us denote the probability simplex by $\triangle_k = \{x \in \mathbb{R}^k \mid \sum_{i=1}^k x_i = 1 \text{ and } \forall i\, x_i \geq 0\}$. Fix a loss function $\ell : [k] \times \triangle_k \to \mathbb{R}$, and let

$$L(p, q) \;=\; \sum_{i=1}^k p_i \ell(i, q)$$

be its associated risk. A loss function is called *proper* if for all $p, q \in \triangle_k$

$$L(p, p) \;\leq\; L(p, q).$$

2 Exercises

1. [4 pt, one each] *In this question we are following the common special-case notation for* 2 *outcomes, where outcomes are* $\{0, 1\}$ *and distributions on these* 2 *outcomes are parametrised by the probability* $q \in [0, 1]$ *of observing the outcome* 1.

   Determine (prove or disprove) whether the following losses are proper.

   (a)
   $$\ell(0, q) \;=\; q \qquad \ell(1, q) \;=\; 1 - q$$

   (b)
   $$\ell(0, q) \;=\; q^2 \qquad \ell(1, q) \;=\; (1 - q)^2$$

(c)
$$\ell(0,q) \;=\; \sqrt{\frac{q}{1-q}} \qquad \ell(1,q) \;=\; \sqrt{\frac{1-q}{q}}$$

(d)
$$\ell(0,q) \;=\; \sqrt{q} \qquad \ell(1,q) \;=\; \sqrt{1-q}$$

2. [3 pt, one each] Let ϕ be any differentiable convex function.
   (a) Show that $B_\phi(x,x) = 0$.
   (b) Show that $B_\phi(x,y) \geq 0$.
   (c) Show that $B_\phi(x,y)$ is convex in x.

3. [4 pt] Consider a differentiable convex function $\phi : \triangle_k \to \mathbb{R}$. Let δ^i be the i^{th} standard basis vector (i.e. $\delta^i_i = 1$ and $\delta^i_j = 0$ for $j \neq i$). Define the loss function $\ell_\phi : [k] \times \triangle_k \to \mathbb{R}$ by
$$\ell_\phi(i,q) \;=\; B_\phi(\delta^i, q)$$
   Show that ℓ_ϕ is proper.

4. [3 pt] Consider a random variable $X \in \mathbb{R}^k$ with mean $\mu = \mathbb{E}[X]$. Show
$$\mu \;\in\; \arg\min_{v \in \mathbb{R}^k} \mathbb{E}\left[B_\phi(X,v)\right].$$

5. [4 pt, 2 each] A final grade between 1 and 10 is composed of 40% homework and 60% exams. There are 13 homework sets of 5 questions each. For each homework set, one question is selected uniformly at random and is graded. There are 2 exams with 5 questions each, which are all graded. Every graded question receives a binary grade: full points or no points.

   Imagine a student that hands in all homeworks and exams, and that answers each question correctly independently at random with probability $p = 0.8$.

   (a) What is the probability that the student gets at least grade 7.5. Calculate 3 significant digits.
   (b) In an ideal world all questions would be graded. Calculate the probability that the same student gets at least grade 7.5. Calculate 3 significant digits.