CSE 203B W21 Homework 2 Sample Solution

Due Time : 11:50pm, Monday Feb. 1, 2021 Submit to Gradescope
Gradescope: `https://gradescope.com/`

In this homework, we work on exercies from text book including level sets of convex, concave, quasi-convex, quasi-concave functions (3.2), second-order conditions for convexity on affine sets (3.9), Kullback-Leibler divergence (3.13), determination of convex, concave, quasi-convex, quasi-concave functions (3.16), conjugate functions (3.36), and gradient and Hessian of conjugate functions (3.40). Extra assignments are given on entropy, Kullback-Leibler divergence, and conjugate function.

Total points: 30. Exercises are graded by completion, assignments are graded by content.

I. Exercises from textbook chapter 3 (6 pts, 1pt for each problem)

3.2, 3.9, 3.13, 3.16, 3.36, 3.40.

II. Assignments (24 pts)

II. 1. Entropy. (8 pts)

In information theory, entropy is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i,$$

where $P(X = x_i) = p_i$ for $i \in \{1, 2, ..., n\}$, $p_i > 0\, \forall i$ and $\sum_{i=1}^{n} p_i = 1$.

(1) Show that the entropy $H(X)$ is concave.

We can show that entropy $H(x)$ is concave with $\nabla^2 H(x) \preccurlyeq 0$.

(2) Derive the conjugate function of the negative entropy $(-H(X))$.

By taking the gradient of the conjugate function and enforce the equality, we have $e^{y_i-1} = p_i$. Then, we plug in the derived relation into the conjugate function and arrive at $f^*(y) = \sum_{i=1}^{n} e^{y_i-1}$. (You get full credit if you reach this step.) If we plug in the constraint of p, we will have $\sum_{i=1}^{n} e^{y_i-1} = 1$.

II. 2. Properties and application of Kullback-Leibler divergence. (8 pts): It is often useful to measure the difference between two probability distributions over the same random variable. Here, we define the Kullback-Leibler (KL) divergence as $KL(p,q) = \sum_i p_i \log(\frac{p_i}{q_i})$, where $p_i > 0,\ q_i > 0,\ \forall i \in \{1,2,...,n\}$, and $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i = 1$.

(1) Assume we have two six-sided dice with the probability distributions as shown in the given table. Please calculate the KL divergences $KL(P,Q)$ and $KL(Q,P)$ of the given example.

Event	P	Q
1	1/6	1/12
2	1/6	1/6
3	1/6	1/4
4	1/6	1/4
5	1/6	1/6
6	1/6	1/12

Plug in the given numbers into the definition of KL-divergence, we have $KL(P,Q) = \frac{1}{6}ln(2) + \frac{1}{6}ln(1) + \frac{1}{6}ln(\frac{2}{3}) + \frac{1}{6}ln(\frac{2}{3}) + \frac{1}{6}ln(1) + \frac{1}{6}ln(2) \approxeq 0.09589$, and $KL(Q,P) = \frac{1}{12}ln(\frac{1}{2}) + \frac{1}{6}ln(1) + \frac{1}{4}ln(\frac{3}{2}) + \frac{1}{4}ln(\frac{3}{2}) + \frac{1}{6}ln(1) + \frac{1}{12}ln(\frac{1}{2}) \approxeq 0.08721$.

(2) Use the example in (2) to illustrate that the Hellinger distance can serve as a lower bound:

$$KL(P,Q) \geq \sum_i (\sqrt{p_i} - \sqrt{q_i})^2.$$

Plug in the numbers and we get the lower bound $KL(P,Q) \geq (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{12}})^2 + (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{6}})^2 + (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{4}})^2 + (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{4}})^2 + (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{6}})^2 + (\sqrt{\frac{1}{6}} - \sqrt{\frac{1}{12}})^2 \approxeq 0.04543$

(3) Create a simple but nontrivial numerical example to demonstrate that $KL(P,Q) \geq 0$ for all possible P,Q and the equality holds only when $P = Q$.

Any reasonable examples are accepted as correct answers.

II. 3. Further practice on conjugate functions. (4 pts)
Consider the function

$$f(x) = \begin{cases} \frac{1}{2}x^2, & |x| < 1, \\ |x| - \frac{1}{2}, & |x| > 1. \end{cases}$$

Find the conjugate function f^*.

$$f^*(y) = \begin{cases} \frac{1}{2}y^2, & |y| < 1, \\ \infty, & |y| > 1. \\ undefined, & |y| = 1 \end{cases}$$

No points will be deducted if the undefined case is missing.

II. 4. Show that the dual of the l_p norm is the l_q norm, where $\frac{1}{p} + \frac{1}{q} = 1$. (4 pts)

Solve $\nabla_y \frac{x^T y}{||y||_p} = 0$. Using the fractions differentiation, we can simplify the equation and arrive at the expression $x_i ||y||_p = (x^T y) \frac{y_i^{p-1} ||y||_p}{\sum_{k=1}^n y_k^p}$. Next, we raise each x_i to the exponential of $\frac{p}{p-1}$ and sum up the n elements. This will lead us to $\sum x_i^{\frac{p}{p-1}} = (x^T y)^{\frac{p}{p-1}}$. Last, we let $q = \frac{p}{p-1}$, and we can express $x^T y$ as $(\sum x_i^q)^{\frac{1}{q}}$, which is equivalent to $||x||_q$.