CS7643: Deep Learning
Fall 2019
Problem Set 0

Instructor: Dhruv Batra
TAs: Harsh Agrawal, Neha Jain, Ashwin Kalyan, Harish Kamath,
Anishi Mehta, Nirbhay Modhe, Michael Piseno, Viraj Prabhu, Sarah Wiegreffe
Discussions: https://piazza.com/gatech/fall2019/cs48037643

Due: Thursday, August 22, 11:00am

Instructions

1. We will be using Gradescope to collect your assignments. Please read the following instructions for submitting to Gradescope carefully! Failure to follow these instructions may result in parts of your assignment not being graded. We will not entertain regrading requests for failure to follow instructions.

   - For Section 1: Multiple Choice Questions, it is mandatory to use the LaTeX template provided on the class webpage (https://www.cc.gatech.edu/classes/AY2020/cs7643_fall/assets/ps0.zip). For every question, there is only one correct answer. To mark the correct answer, change `\choice` to `\CorrectChoice`
   - For Section 2: Proofs, each problem/sub-problem is in its own page. This section has 5 total problems/sub-problems, so you should have 5 pages corresponding to this section. Your answer to each sub-problem should fit in its corresponding page.
   - For Section 2, LaTeX'd solutions are strongly encouraged (solution template available at https://www.cc.gatech.edu/classes/AY2020/cs7643_fall/assets/ps0.zip), but scanned handwritten copies are acceptable. If you scan handwritten copies, please make sure to append them to the pdf generated by LaTeX for Section 1.

2. Hard copies are **not** accepted.

3. We generally encourage you to collaborate with other students. You may talk to a friend, discuss the questions and potential directions for solving them. However, you need to write your own solutions and code separately, and *not* as a group activity. Please list the students you collaborated with.
   Exception: PS0 is meant to serve as a background preparation test. You must NOT collaborate on PS0.

1 Multiple Choice Questions

1. (1 point) true/false We are machine learners with a slight gambling problem (very different from gamblers with a machine learning problem!). Our friend, Bob, is proposing the following payout on the roll of a dice:

$$\text{payout} = \begin{cases} \$1 & x = 1 \\ -\$1/4 & x \neq 1 \end{cases} \tag{1}$$

where $x \in \{1, 2, 3, 4, 5, 6\}$ is the outcome of the roll, (+) means payout to us and (−) means payout to Bob. Is this a good bet i.e are we expected to make money?

○ True ● **False**

2. (1 point) X is a continuous random variable with the probability density function:

$$p(x) = \begin{cases} 4x & 0 \leq x \leq 1/2 \\ -4x + 4 & 1/2 \leq x \leq 1 \end{cases} \tag{2}$$

Which of the following statements are true about equation for the corresponding cumulative density function (cdf) $C(x)$?
[*Hint:* Recall that CDF is defined as $C(x) = Pr(X \leq x)$.]

- ● $C(x) = 2x^2$ **for** $0 \leq x \leq 1/2$
- ○ $C(x) = -2x^2 + 4x - 3/2$ for $1/2 \leq x \leq 1$
- ○ All of the above
- ○ None of the above

3. (2 point) A random variable x in standard normal distribution has following probability density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

Evaluate following integral

$$\int_{-\infty}^{\infty} p(x)(ax^2 + bx + c)dx \tag{4}$$

[*Hint:* We are not sadistic (okay, we're a little sadistic, but not for this question). This is not a calculus question.]

○ a + b + c ○ c ● **a + c** ○ b + c

4. (2 points) Consider the following function of $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$:

$$f(\mathbf{x}) = \sigma\left(\log\left(5\left(\max\{x_1, x_2\} \cdot \frac{x_3}{x_4} - (x_5 + x_6)\right)\right) + \frac{1}{2}\right) \tag{5}$$

where σ is the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

Compute the gradient $\nabla_{\mathbf{x}} f(\cdot)$ and evaluate it at at $\hat{\mathbf{x}} = (5, -1, 6, 12, 7, -5)$.

○ $\begin{bmatrix} 0.157 \\ 0.0 \\ 0.131 \\ -0.065 \\ -0.846 \\ -0.846 \end{bmatrix}$ ● $\begin{bmatrix} 0.157 \\ 0 \\ 0.131 \\ -0.065 \\ -0.314 \\ -0.314 \end{bmatrix}$ ○ $\begin{bmatrix} 0.031 \\ 0 \\ 0.026 \\ -0.013 \\ -0.062 \\ -0.062 \end{bmatrix}$ ○ $\begin{bmatrix} -0.468 \\ 0 \\ -0.390 \\ 0.195 \\ 0.937 \\ 0.937 \end{bmatrix}$

5. (2 points) Which of the following functions are convex?

   ○ $||\mathbf{x}||_{\frac{1}{2}}$
   ○ $\min_i \mathbf{a}_i^T \mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$
   ○ $\log(1 + \exp(\mathbf{w}^T \mathbf{x}_i))$ for $\mathbf{w} \in \mathbb{R}^d$
   ● **All of the above**

6. (2 points) Suppose you want to predict an unknown value $Y \in \mathbb{R}$, but you are only given a sequence of noisy observations $x_1 ... x_n$ of Y with i.i.d. noise ($x_i = Y + \epsilon_i$).. If we assume the noise is I.I.D. Gaussian ($\epsilon_i \sim N(0, \sigma^2)$), the maximum likelihood estimate ($\hat{y}$) for Y can be given by:

   ○ A: $\hat{y} = \text{argmin}_y \sum_{i=1}^n (y - x_i)^2$
   ○ B: $\hat{y} = \text{argmin}_y \sum_{i=1}^n |y - x_i|$
   ○ C: $\hat{y} = \frac{1}{n} \sum_{i=1}^n x_i$
   ● **Both A & C**
   ○ Both B & C

Q7:

Consider $f(x) = \ln(x) - (x-1)$

$$f'(x) = \frac{1}{x} - 1.$$

For $x \in (0,1]$, $f'(x) > 0 \Rightarrow f(x)$ is increasing.

$x \in (1, \infty)$ $f'(x) < 0 \Rightarrow f(x)$ is decreasing

Function attains local maxima

at $f'(x) = 0 \Rightarrow x = 1.$

The value at $x=1 \Rightarrow 0.$

So, $\ln x \leq x-1$ is true $\forall x > 0$

and equal only when $x = 1.$

Q8: a. $KL(p,q) = \sum_{i=1}^{k} p_i \log\left(\frac{p_i}{q_i}\right)$.

Consider $-KL$. $\Rightarrow \sum_{i=1}^{p} p_i \log \frac{q_i}{p_i}$

$$\leq \sum p_i \left(\frac{q_i}{p_i} - 1\right)$$

$$\leq \sum q_i - \sum p_i$$

$$= 0.$$

$$\Rightarrow -KL \leq 0 \Rightarrow \boxed{KL \geq 0}$$

b. From prev question (Q7) inequality is equal only when $x=1$.

so $p_i = q_i \; \forall i$.

c. Consider 2 prob dist $p = (0.3, 0.7)$
$q = (0.4, 0.6)$

$KL(p,q) = 0.3 \ln(3/4) + 0.7 \ln(7/6) = 0.021$

$KL(q,p) = 0.4 \ln(4/3) + 0.6 \ln(6/7) = 0.022$

Scanned with CamScanner

Q9. A function f is convex if

$$f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$\forall \alpha \in [0$

Given fun, $L(\omega) = -\log\left(\frac{e^{z_y}}{\sum_k e^{z_k}}\right)$

$$= -\log\left(\frac{e^{\omega^y x}}{\sum_k e^{\omega^k x}}\right)$$, where ω^y is y^{th} row in

RHS $= \alpha L(\omega_1) + (1-\alpha) L(\omega_2)$

$$= -\alpha \log\left(\frac{e^{\omega_1^y x}}{\sum e^{\omega_1^k x}}\right) + (1-\alpha)\left(-\log\left(\frac{e^{\omega_2^y x}}{\sum e^{\omega_k^y x}}\right)\right)$$

$$= -\log\left(\frac{e^{\alpha \omega_1^y x + (1-\alpha) e^{\omega_2^y x}}}{\left(\sum e^{\omega_1^k x}\right)^{\alpha}\left(\sum e^{\omega_2^k x}\right)^{(1-\alpha)}}\right)$$

LHS $= L(\alpha \omega_1 + (1-\alpha)\omega_2)$

$$= -\log\left(\frac{e^{\alpha w_1^y x + (1-\alpha) w_2^y x}}{\sum e^{\alpha w_1^k x + (1-\alpha) w_2^k x}} \right).$$

Let $e^{w_1^k x} = a_k$, $e^{w_2^k x} = b_k$.

we can rewrite denominators as.

$$RHS_{den} = \left(\sum_k a_k\right)^\alpha \left(\sum_k b_k\right)^{1-\alpha}.$$

$$LHS_{den} = \sum_k a_k^\alpha b_k^{1-\alpha}.$$

when $\alpha \in (0,1)$, weknow.

$a_k, b_k > 0$

$$\left(\sum a_k\right)^\alpha \left(\sum b_k\right)^{1-\alpha} \geq \sum a_k^\alpha b_k^{1-\alpha}.$$

As numerators are same in LHS & RHS.

$$-\log\left(\frac{c}{RHS_{den}}\right) \geq -\log\left(\frac{c}{LHS_{den}}\right),$$

when $RHS_{den} \geq LHS_{den}$

Hence $LHS \leq RHS$