IE534/CS547: Deep Learning **(Due: Feb-14-2021)**

Homework #2

Instructor: Richard B. Sowers

Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- **This is a group homework, every group only submit ONE solution on Compass .** Please include the names of all the group members.
- **Due time is at 11:59pm** at the due date. **No late submission!**
- All students are expected to abide by the Honor Code
- All date-times will be in Champaign-Urbana
- Please put your typed solution in a PDF format. For code, you can either submit it in .py file or jupyter notebook file also with its google colab shared link in your solution PDF file.

Problem 1: EntropyConvex (10 points)

With

$$H(p', p) \stackrel{\text{def}}{=} p' \ln \frac{p'}{p} + (1-p') \ln \frac{1-p'}{1-p}$$

for p and p' in $(0,1)$, show that $p' \mapsto H(p', p)$ is convex for each $p \in (0,1)$.

Problem 2: FenchelB (Extra credit 10 points)

Let's use *entropy* as a pretext for understanding Euler's equations of optimality. For p and p' in $(0,1)$, relative entropy is (as usual)

$$H(p', p) \stackrel{\text{def}}{=} p' \ln \frac{p'}{p} + (1-p') \ln \frac{1-p'}{1-p}$$

Entropy is a fundamental concept in both statistical mechanics and information theory, as it is the *Legendre-Fenchel* transform (an object of interest in tail behavior due to the Ellis-Gärtner theorem) of the logarithm of the moment generating function of a Bernoulli random variable. Fixing p and p' in $(0,1)$, compute

$$\max_{\theta \in (0,1)} \left\{ \theta p - \ln \left\{ p' e^{\theta} + (1-p') \right\} \right\}$$

Note that $p' e^{\theta} + (1-p')$ is indeed the moment generating function of a Bernoulli(p') random variable.

Problem 3: Coding question (10 points)

Take $N = 200$ Gaussian points on the line centered at 0 and with variance 1. Assign label 1 to the ones to the right of the origin and assign label 0 to the ones on the left of the origin. Carry out the following with PyTorch

(1) Carry out logistic regression, and note the width of the transition layer. If your logit function $\log(\frac{P}{1-P}) = mx + b$, then the width is defined as $\frac{1}{m}$.
(2) Flip 5 points on each side of the origin to the 'wrong' label and note the width of the transition layer
(3) Repeat this for 15, 20, 25, 30, 35 points and plot the width of the transition layer as a function of the number of points with the 'wrong' label.