CS7643: Deep Learning
Fall 2017
Homework 0

Instructor: Dhruv Batra
TAs: Harsh Agrawal, Michael Cogswell, Nirbhay Modhe, Erik Wijmans
Discussions: https://piazza.com/gatech/fall2018/cs48037643

Due: Wednesday, Sep 5, 11:55pm

Instructions

1. We will be using Gradescope to collect your assignments. Please read the following instructions for submitting to Gradescope carefully!
   - Each problem/sub-problem should be on one or more pages. This assignment has 11 total problems/sub-problems, so you should have at least 11 pages in your submission.
   - When submitting to Gradescope, make sure to mark which page corresponds to each problem/sub-problem.
   - For the coding problem (problem 8), please use the provided `collect_submission.sh` script and upload `hw0.zip` to the HW0 Code assignment on Gradescope. While we will not be explicitly grading your code, you are still required to submit it.

     Please make sure you have saved the most recent version of your jupyter notebook before running this script.
   - Note: This is a large class and Gradescope's assignment segmentation features are essential. 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.

     Please read https://stats200.stanford.edu/gradescope_tips.pdf for additional information on submitting to Gradescope.
2. LaTeX'd solutions are strongly encouraged (solution template available at cc.gatech.edu/classes/AY2019/cs7643_fall/assets/sol0.tex), but scanned handwritten copies are acceptable. 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.

1 Probability and Statistics

1. (1 point) 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? Are we expected to make money?

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

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

What is the equation for the corresponding cumulative density function (cdf) $C(x)$?

[*Hint:* Recall that CDF is defined as $C(x) = Pr(X \leq x)$.]

3. (1 point) Recall that the variance of a random variable is defined as $Var[X] = E[(X - \mu)^2]$, where $\mu = E[X]$. Use the properties of expectation to show that we can rewrite the variance of a random variable X as

$$Var[X] = E[X^2] - (E[X])^2 \tag{3}$$

4. (1 point) A random variable x in standard normal distribution has following probability density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{4}$$

Evaluate following integral

$$\int_{-\infty}^{\infty} p(x)(ax^2 + bx + c)dx \tag{5}$$

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

2 Proving Stuff

5. (2 points) Prove that

$$\log_e x \leq x - 1, \qquad \forall x > 0 \tag{6}$$

with equality if and only if $x = 1$.

[*Hint:* Consider differentiation of $\log(x) - (x - 1)$ and think about concavity/convexity and second derivatives.]

6. (3 points) Consider two discrete probability distributions p and q over k outcomes:

$$\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} q_i = 1 \tag{7a}$$

$$p_i > 0, q_i > 0, \quad \forall i \in \{1, \dots, k\} \tag{7b}$$

The Kullback-Leibler (KL) divergence (also known as the *relative entropy*) between these distributions is given by:

$$KL(p, q) = \sum_{i=1}^{k} p_i \log \left(\frac{p_i}{q_i} \right) \tag{8}$$

It is common to refer to $KL(p, q)$ as a measure of distance (even though it is not a proper metric). Many algorithms in machine learning are based on minimizing KL divergence between two probability distributions. In this question, we will show why this might be a sensible thing to do.

(a) Using the results from Q5, show that $KL(p, q)$ is always positive.

(b) When is $KL(p, q) = 0$?

(c) Provide a counterexample to show that the KL divergence is not a symmetric function of its arguments: $KL(p, q) \neq KL(q, p)$

[*Hint:* This question doesn't require you to know anything more than the definition of $KL(p, q)$ and the identity in Q5]

3 Calculus

7. (3 points) Consider the following function of $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)$:

$$f(\mathbf{x}) = \sigma \left(\log \left(5 \left(\max\{x_1, x_2\} \cdot \frac{x_3}{x_4} - (x_5 + x_6) \right) \right) + \frac{1}{2} \right) \tag{9}$$

where σ is the sigmoid function

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

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

4 Softmax Classifier

8. (5 points) Implement a Softmax classifier (from scratch, no ML libraries allowed), and train it (via SGD) on CIFAR-10:
`cc.gatech.edu/classes/AY2019/cs7643_fall/hw0-q8/`.

In your solutions, please include the output of cell 3 in the jupyter notebook (the cell with `grad_check_sparse`), the plot of the training loss, and, the weight visualizations with a brief comment on how well the weight visualizations correspond with their respective classes as the answer to this problem.

9. (3 points) In this question, you will prove that cross-entropy loss for a softmax classifier is convex in the model parameters, thus gradient descent is guaranteed to find the optimal parameters. Formally, consider a single training example $(\mathbf{x}, y)$. Simplifying the notation slightly from the implementation writeup, let

$$\mathbf{z} = W\mathbf{x} + \mathbf{b}, \tag{11}$$

$$p_j = \frac{e^{z_j}}{\sum_k e^{z_k}}, \tag{12}$$

$$L(W) = -\log(p_y) \tag{13}$$

Prove that $L(\cdot)$ is convex in W.

[*Hint:* One way of solving this problem is "brute force" with first principles and Hessians. There are more elegant solutions.]