HW 0 (“Earn-your-Grace-Days” HW)

CS5480: Deep Learning
IIT-Hyderabad
Jan-Apr 2018

Max Points: No points. You will be credited with your grace days for the semester, if you submit this HW on time.
Due: 21st Jan 2018 11:59 pm

This homework is intended to cover the following topics:

- Machine learning and math background you are expected to know for this course
- Basics of neural networks and backpropagation

Instructions

- Please use Google Classroom to upload your submission by the deadline mentioned above. Your submission should comprise of a single file (PDF/ZIP), named `<Your_Roll_No>_HW0`, with all your solutions.
- Late submission of this HW will result in forfeit of your grace days for the semester.
- Incomplete submissions will receive proportionally prorated number of grace days for the semester.
- Please read the department plagiarism policy. Do not engage in any form of cheating or plagiarism - if we find such behavior in your submission, you will not be credited any grace days for the semester (for both receiver and giver), as well as be imposed with other suitable penalties. Please talk to instructor or TAs if you have concerns.

1 Questions

1. **Probability:** The organizers of ELAN decide to have a stall this year (this is made up of course, don’t suggest this to them!) that has‘proposed the following payout on the roll of an unbiased dice:

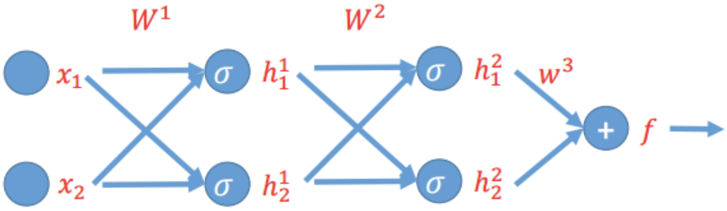
$$\text{Payout} = \begin{cases} +\text{Rs } 100, & x = 1 \\ -\text{Rs } 25, & x \neq 1 \end{cases}$$

where $x \in \{1, 2, 3, 4, 5, 6\}$ is the outcome of a dice roll, $(+)$ means payout to ELAN, and $(-)$ means payout to a visitor. Is this a good plan? Will ELAN organizers make money?

2. **Backpropagation:** Consider a 3-layer network:

$$h^1 = \sigma(W^1 x), h^2 = \sigma(W^2 h^1), f(x) =< w^3, h^2 >$$

. Compute $\frac{\partial f}{\partial W^1_{i,j}}$.



3. **More Backprop:** You are training a 3-layer neural network and would like to use backprop to compute the gradient of the cost function. In the backprop algorithm, one of the steps is to update $\Delta^{(2)}_{ij} := \Delta^{(2)}_{ij} + \delta^{(3)}_i * (a^{(2)})_j$ for every i, j. Can you rewrite the above equation for all the weights in layer 2 in vector form? *(HINT: $\Delta^{(2)} := \Delta^{(2)} + \cdots ??$)*

4. **Perceptron:** Consider a perceptron that has 2 input units (no bias) and 1 output unit and has learned a function f using a given training set. Is it always possible to construct another perceptron, also with 2 input units and 1 output unit that also learns f but has a fixed bias value = 1? Briefly explain how or why not.

5. **Matrix Calculus:** Recall that the gradient of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$ is the vector whose components are the partial derivatives of f.

$$\nabla f(\mathbf{x}) = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}) \in \mathbb{R}^n$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$. Consider the function $f(\mathbf{x}) = \log\left(\sum_{i=1}^n e^{x_i}\right)$:

   (a) Compute the partial derivatives $\frac{\partial f}{\partial x_i}, i = 1, \cdots, n$.

   (b) What is the final gradient, $\nabla f(\mathbf{x})$?

   (c) Based on the solution to the above problem, do you think there is a matrix calculus equivalent of the chain rule from calculus? Formulate this rule.

6. **MLE and Neural Networks:** Show mathematically how backpropagation in a neural network w.r.t. a loss function is equivalent to Maximum Likelihood Estimation. (You can choose a error function of your choice to show this.)
(HINT: Reading the first few chapters of `http://neuralnetworksanddeeplearning.com/` *may help you with this!)*

7. (OPTIONAL) **Machine Learning Methods:** Continuing with our discussion in class, state two comparisons between logistic regression and a perceptron: (i) one comparison between the solutions obtained using the two methods, and (ii) another comparison in terms of the inference obtained using the two methods.

8. **Calibration:** How long did you spend on each problem in this homework, and on the homework in total?