DS-GA 1008: Deep Learning, Spring 2019
Homework Assignment 1

Due: 6pm on Friday, Feb 15, 2019

```
He who learns but does not think is lost.
He who thinks but does not learn is in great danger.
Confucius (551 - 479 BC)
```

1. Backprop

Backpropagation or "backward propagation through errors" is a method which calculates the gradient of the loss function of a neural network with respect to its weights.

1.1 Warm-up

The chain rule is at the heart of backpropagation. Assume you are given input $\boldsymbol{x}$ and output $\boldsymbol{y}$, both in $\mathbb{R}^2$, and the error backpropagated to the output is $\frac{\partial L}{\partial \boldsymbol{y}}$. In particular, let

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b},$$

where $\boldsymbol{W} \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^2$. Give an expression for $\frac{\partial L}{\partial \boldsymbol{W}}$ and $\frac{\partial L}{\partial \boldsymbol{b}}$ in terms of $\frac{\partial L}{\partial \boldsymbol{y}}$ and $\boldsymbol{x}$ using the chain rule.

1.2. Softmax

Multinomial logistic regression is a generalization of logistic regression into multiple classes. The softmax expression is at the crux of this technique. After receiving n unconstrained values, the softmax function normalizes these values to n values that all sum to 1. This can then be perceived as probabilities attributed to the various classes by a classifier. Your task here is to back-propagate error through this module. The softmax expression which indicates the probability of the j-th class is as follows:

$$\mathbb{P}(z = j \mid \boldsymbol{x}) = y_j = \frac{\exp(\beta x_j)}{\sum_i \exp(\beta x_i)} \tag{1}$$

What is the expression for $\frac{\partial y_j}{\partial x_i}$? (Hint: Answer differs when $i = j$ and $i \neq j$).

Note that the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ aren't scalars but vectors. While $\boldsymbol{x}$ represents the n values input to the system, $\boldsymbol{y}$ represents the n probabilities output from the system. Therefore, the expression y_j represents the j-th element of $\boldsymbol{y}$.

DS-GA 1008: Deep Learning, Spring 2019
Homework Assignment 1

Due: 6pm on Friday, Feb 15, 2019

---

2. PyTorch

If you haven't already, install most recent versions of Python (3.6 or higher), PyTorch 1.0 (we recommend using conda for the installation), and Jupyter.

Complete the exercises provided in the course Google Drive folder (link):

`Homework/Assignment_1/DS-GA-1008-HW_assignment_1.ipynb`

3. Evaluation

Homework is worth a total of 100 points.

- Part 1 - 50 points
- Part 2 - 50 points

4. Submission

You are required to write up your solutions to Part 1 using LaTeX.

Send an email to `sp19dl@gmail.com` by the deadline (6pm on Friday, Feb 15, 2019) with subject line

`[First-name Last-name netID] Assignment 1`

and attachments:

- `First-name_Last-name_netID_A1.tex` file for Part 1
- `First-name_Last-name_netID_A1.pdf` file for Part 1
- `First-name_Last-name_netID_A1.ipynb` file for Part 2

5. Disclaimers

You are allowed to discuss problems with other students in the class but have to write up your solutions on your own.

As feedback might be provided during the first days, the current homework assignment might be undergoing some minor changes. We'll notify you if this happens.