DS-GA.1008 Deep Learning February 5, 2016

Solutions to Assignment 1

Team: R2DEEP2 (Ankit Vani, Srivas Venkatesh) *Due: Thursday, February 11*

Problem 1 - Backpropagation

1. Warmup: Logistic regression is a pretty popular technique in machine learning to classify data into two categories. This technique builds over linear regression by using the same linear model but this is followed by the sigmoid function which converts the output of the linear model to a value between 0 and 1. This value can then be interpreted as a probability. This is usually represented as:

$$P(y=1|x_{in}) = x_{out} = \sigma(x_{in}) = \frac{1}{1+\mathrm{e}^{-x_{in}}} \tag{1}$$

where x_{in} as the name would suggest is the input scalar (which is also the output of linear model) and x_{out} is the output scalar.
If the error backpropagated to x_{out} is $\frac{\partial E}{\partial x_{out}}$, write the expression for $\frac{\partial E}{\partial x_{in}}$ in terms of $\frac{\partial E}{\partial x_{out}}$.

Solution:
We know that using the chain rule we can write $\frac{\partial E}{\partial x_{in}} = \frac{\partial E}{\partial x_{out}} \frac{\partial F(x_{in})}{\partial x_{in}}$ where $F(x_{in}) = P(y = 1|x_{in})$. Applying this to the above function we get:

$$\begin{aligned} \frac{\partial F(x_{in})}{\partial x_{in}} &= \frac{\mathrm{e}^2}{(1+\mathrm{e}^2)} \\ \implies \frac{\partial E}{\partial x_{in}} &= \frac{\partial E}{\partial x_{out}} \frac{\mathrm{e}^2}{(1+\mathrm{e}^2)} \end{aligned} \tag{2}$$

□

2. 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 expression 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 backpropagate error through this module. The softmax expression which indicates the probability of the i-th class is as follows:

$$P(y=i|X_{in}) = (X_{out})_i = \frac{\mathrm{e}^{-\beta(X_{in})_i}}{\sum_k \mathrm{e}^{-\beta(X_{in})_k}} \tag{3}$$

What is the expression for $\frac{\partial (X_{out})_i}{\partial (X_{in})_j}$? (Hint: Answer differs when $i = j$ and $i \neq j$).
The variables X_{in} and X_{out} arent scalars but vectors. While X_{in} represents the n values input to the system, X_{out} represents the n probabilities output from the system. Therefore, the expression $(X_{out})_i$ represents the i-th element of X_{out}.

Solution:
For this we will consider that components of X_{in} are independent of one another. Then using simple product rule of differentiation we get:

$$\frac{\partial (X_{out})_i}{\partial (X_{in})_j} = \begin{cases} -\beta \frac{\mathrm{e}^{-\beta (X_{in})_i}}{\sum_k \mathrm{e}^{-\beta (X_{in})_k}} + \beta \left(\frac{\mathrm{e}^{-\beta (X_{in})_i}}{\sum_k \mathrm{e}^{-\beta (X_{in})_k}} \right)^2 = \beta (X_{out})_i ((X_{out})_i - 1) & \text{if } i = j \\ \beta \frac{\mathrm{e}^{-\beta (X_{in})_i} \cdot \mathrm{e}^{-\beta (X_{in})_j}}{\left(\sum_k \mathrm{e}^{-\beta (X_{in})_k} \right)^2} = \beta (X_{out})_i (X_{out})_j & \text{if } i \neq j \end{cases} \tag{4}$$

□

Solutions to Assignment 1

Team: R2DEEP2 (Ankit Vani, Srivas Venkatesh) *Due: Thursday, February 11*

Problem 2 - Torch (MNIST Handwritten Digit Recognition)

Solution: TODO □