TTIC 31230 Problem Set 1
Win 2017

Hao Jiang

January 9, 2017

Problem 1

a

Using chain rule, we have

$$\frac{\partial \ell}{\partial x} = \frac{\partial \ell}{\partial g}\frac{\partial g}{\partial x} + \frac{\partial \ell}{\partial h}\frac{\partial h}{\partial x}$$

Using backpropagation, we first compute the result for the hidden layer (the layer of y and z), having the value of $\delta_y = \frac{\partial \ell}{\partial y}$ and $\delta_z = \frac{\partial \ell}{\partial z}$. and for the input layer,

$$\begin{aligned}\frac{\partial \ell}{\partial x} &= g'(y)\delta_y + h'(z)\delta_z \\ &= \frac{\partial \ell}{\partial g}\frac{\partial g}{\partial x} + \frac{\partial \ell}{\partial h}\frac{\partial h}{\partial x}\end{aligned}$$

This result is equal to the result from chain rule and shows that backpropagation correctly compute the derivative.

b

For second derivative

$$\begin{aligned}\frac{\partial^2 \ell}{\partial x^2} &= \frac{\partial}{\partial x}(\frac{\partial \ell}{\partial g}\frac{\partial g}{\partial x} + \frac{\partial \ell}{\partial h}\frac{\partial h}{\partial x}) \\ &= \frac{\partial^2 \ell}{\partial g^2}\left(\frac{\partial g}{\partial x}\right)^2 + \frac{\partial \ell}{\partial g}\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 \ell}{\partial h^2}\left(\frac{\partial h}{\partial x}\right)^2 + \frac{\partial \ell}{\partial h}\frac{\partial^2 h}{\partial x^2}\end{aligned}$$

This is not equivalent to the backpropagation result

$$\frac{\partial^2 \ell}{\partial g^2}\frac{\partial^2 g}{\partial x^2} + \frac{\partial^2 \ell}{\partial h^2}\frac{\partial^2 h}{\partial x^2}$$

Problem 2

Softmax function

$$\begin{aligned} Y = [y_1, y_2, \dots, y_n] &= S([x_1, x_2, \dots, x_n]) = S(X) \\ &= \frac{1}{\sum_{i=1}^n \exp(x_i)} [\exp(x_1), \exp(x_2), \dots, \exp(x_n)] \end{aligned}$$

Let $Z = \sum_{i=1}^n \exp(x_i)$, the derivative

$$\frac{\partial y_j}{\partial x_i} = \begin{cases} \frac{\exp(x_j)}{Z} - \frac{(\exp^2(x_j)}{Z^2} = y_i(1 - y_i) & i = j \\ -\frac{\exp(x_i)\exp(x_j)}{Z^2} = -y_i y_j & i \neq j \end{cases}$$

When using backpropagation to compute ∇x_i, we know $\nabla x_i = \sum_{i=k}^n \nabla y_k \frac{\partial y_k}{\partial x_i}$, thus

$$\nabla X = J \nabla Y$$

where $J_{ij} = \frac{\partial y_j}{\partial x_i}$.

$$\begin{aligned} J\nabla Y &= \left(\begin{bmatrix} y_1 & & & \\ & y_2 & & \\ & & \ddots & \\ & & & y_n \end{bmatrix} - YY^T \right) \nabla Y \\ &= \begin{bmatrix} y_1 \nabla y_1 \\ y_2 \nabla y_2 \\ \vdots \\ y_n \nabla y_n \end{bmatrix} - YY^T \nabla Y \\ &= \begin{bmatrix} y_1 \nabla y_1 \\ y_2 \nabla y_2 \\ \vdots \\ y_n \nabla y_n \end{bmatrix} - \texttt{gvdot} * Y \\ &= \begin{bmatrix} y_1(\nabla y_1 - \texttt{gvdot}) \\ y_2(\nabla y_2 - \texttt{gvdot}) \\ \vdots \\ y_n(\nabla y_n - \texttt{gvdot}) \end{bmatrix} \end{aligned}$$

This is an element-wise multiplication of Y and $(\nabla Y - \texttt{gvdot})$, and demonstrate the correctness of Softmax in edf.

Problem 3