Team Hodge Theaters
David Kasofsky – dk2353@nyu.edu
Jonay Trénous – jgt275@nyu.edu

Assignment 1

1 Backpropagation

1.1

By the Chain Rule, we can write

$$\begin{aligned}\frac{\partial E}{\partial x_{in}} &= \frac{\partial E}{\partial x_{out}} \cdot \frac{\partial x_{out}}{\partial x_{in}} \\ &= \frac{\partial E}{\partial x_{out}} \cdot \frac{\partial (1+\exp(-x_{in}))^{-1}}{\partial x_{in}} \\ &= \frac{\partial E}{\partial x_{out}} \cdot (1+\exp(-x_{in}))^{-2} \exp(-x_{in}) \\ &= \frac{\partial E}{\partial x_{out}} \cdot \exp(-x_{in}) \cdot x_{out}^2\end{aligned}$$

1.2

First, consider the case $i \neq j$:

$$\begin{aligned}\frac{\partial (X_{\text{out}})_i}{\partial (X_{\text{in}})_j} &\overset{\text{Prod. Rule}}{=} \frac{\partial \exp(-\beta (X_{\text{in}})_i)}{\partial (X_{\text{in}})_j} \left(\sum_k \exp(-\beta (X_{\text{in}})_k) \right)^{-1} + \frac{\partial \left(\sum_k \exp(-\beta (X_{\text{in}})_k)\right)^{-1}}{\partial (X_{\text{in}})_j} \exp(-\beta (X_{\text{in}})_i) \\ &= 0 + \beta \frac{\exp(-\beta (X_{\text{in}})_j) \exp(-\beta (X_{\text{in}})_i)}{\left(\sum_k \exp(-\beta (X_{\text{in}})_k)\right)^2} \\ &= \beta (X_{\text{out}})_j (X_{\text{out}})_i\end{aligned}$$

In the opposite case $i = j$, the first term will not be zero:

$$\frac{\partial (X_{\text{out}})_i}{\partial (X_{\text{in}})_i} \overset{\text{Prod. Rule}}{=} \frac{\partial \exp(-\beta (X_{\text{in}})_i)}{\partial (X_{\text{in}})_i} \left(\sum_k \exp(-\beta (X_{\text{in}})_k) \right)^{-1} + \frac{\partial \left(\sum_k \exp(-\beta (X_{\text{in}})_k) \right)^{-1}}{\partial (X_{\text{in}})_i} \exp(-\beta (X_{\text{in}})_i)$$

$$= \beta \frac{\exp(-\beta (X_{\text{in}})_i) \exp(-\beta (X_{\text{in}})_i)}{\left(\sum_k \exp(-\beta (X_{\text{in}})_k) \right)^2} - \beta \frac{\exp(-\beta (X_{\text{in}})_i)}{\sum_k \exp(-\beta (X_{\text{in}})_k)}$$

$$= \beta \left((X_{\text{out}})_i^2 - (X_{\text{out}})_i \right)$$