Assignment 6: Introduction to Neural Networks and Backpropagation

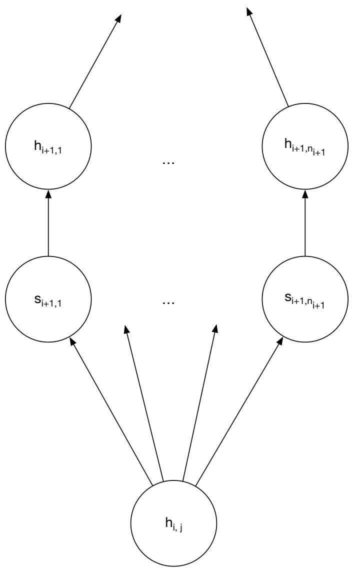
Machine Learning

Fall 2019

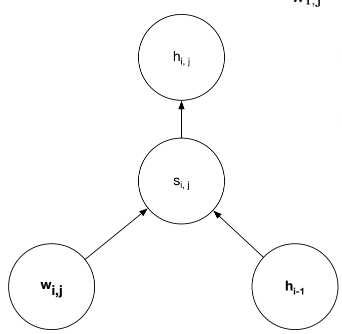
Learning Objectives

- Gain some familiarity with some of the key ideas in machine learning.
- Review of mathematical concepts we will be using in the beginning part of this course.
- Familiarize yourself with computational tools for machine learning.
- Learn linear regression using a “top-down” approach.

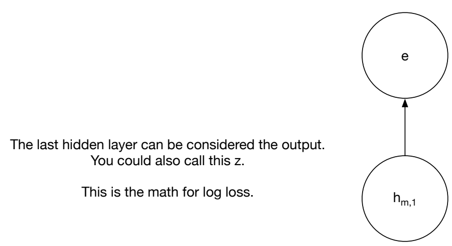
HMC Multivariable Chain Rule Page



$$
\begin{aligned}
\frac{\partial e}{\partial h_{i,j}} &= \sum_{k=1}^{n_{i+1}} \frac{\partial s_{i+1,k}}{\partial h_{i,j}} \frac{\partial h_{i+1,k}}{\partial s_{i+1,k}} \\
&= \sum_{k=1}^{n_{i+1}} w_{j,k}^{i+1} \sigma(s_{i+1,k})(1 - \sigma_{i+1,k})
\end{aligned}
$$



$$
\begin{aligned}
\nabla_{\mathbf{w_{i,j}}} e &= \frac{\partial e}{s_{i,j}} \nabla_{\mathbf{w_{i,j}}} s_{i,j} \\
&= \frac{\partial e}{h_{i,j}} \frac{\partial h_{i,j}}{s_{i,j}} \nabla_{\mathbf{w_{i,j}}} s_{i,j} \\
&= \frac{\partial e}{h_{i,j}} \sigma(s_{i,j})(1 - \sigma(s_{i,j}))\mathbf{h}_{i-1}
\end{aligned}
$$



$$
\frac{\partial e}{\partial h_{m,1}} = -y \frac{1}{h_{m,1}} - (1 - y) \frac{1}{1 - h_{m,1}}
$$

Todo: use a variable for layer instead of just hidden (this would simplify things).

Hidden unit to error:

$$
\frac{\partial e}{\partial h_{i,j}} = \sum_{k=1}^{n_{i+1}} \frac{\partial s_{i+1,k}}{\partial h_{i,j}} \frac{\partial h_{i+1,k}}{\partial s_{i+1,k}} \frac{\partial e}{\partial h_{i+1,k}} \tag{1}
$$

$$
= \sum_{k=1}^{n_{i+1}} w_{j,k}^{i+1} \sigma(s_{i+1,k})(1 - \sigma_{i+1,k}) \frac{\partial e}{\partial h_{i+1,k}} \tag{2}
$$

Weights to error:

$$\begin{align}
\nabla_{\mathbf{w_{i,j}}} e &= \frac{\partial e}{s_{i,j}} \nabla_{\mathbf{w_{i,j}}} s_{i,j} \tag{3} \\
&= \frac{\partial e}{h_{i,j}} \frac{\partial h_{i,j}}{s_{i,j}} \nabla_{\mathbf{w_{i,j}}} s_{i,j} \tag{4} \\
&= \frac{\partial e}{h_{i,j}} \sigma(s_{i,j})(1 - \sigma(s_{i,j}))\mathbf{h}_{i-1} \tag{5}
\end{align}$$

Output to error (serves as a base case. For simplicity we use $h_{m,1}$ to refer to the single node in the mth layer (which is the output layer).

$$\frac{\partial e}{\partial h_{m,1}} = -y \frac{1}{h_{m,1}} - (1-y) \frac{1}{1 - h_{m,1}} \tag{6}$$