# CS/CS 316/365 Deep Learning

# Activity 7

October 10, 2024

## Backpropagation

Activity needs to be handwritten. Submission will be online on canvas only.

- Calculate the derivative $\partial l_i / \partial f[x_i, \theta]$ for the least squares loss function:

$$l_i = (y_i - f[x_i, \theta])^2.$$

solution:

$$\frac{\partial \ell_i}{\partial f\left[\mathrm{x}_i, \phi\right]} = -2\left(y_i - \mathrm{f}\left[\mathrm{x}_i, \phi\right]\right)$$

- Calculate the derivative $\partial l_i / \partial f[x_i, \theta]$ for the binary classification loss function:

$$\ell_i = -(1-y_i)\log\Big[1 - \mathrm{sig}\big[\mathrm{f}[\mathbf{x}_i, \boldsymbol{\phi}]\big]\Big] - y_i \log\Big[\mathrm{sig}\big[\mathrm{f}[\mathbf{x}_i, \boldsymbol{\phi}]\big]\Big]$$

solution:

First, we note that:

$$\frac{\partial\, \mathrm{sig}[z]}{\partial z} = \frac{\exp[-z]}{(1+\exp[-z])^2}$$

Using this formula, the derivative becomes:

$$\begin{aligned}\frac{\partial \ell_i}{\partial \mathrm{f}\left[\mathbf{x}_i, \boldsymbol{\phi}\right]} = (1-y_i) & \frac{1}{1-\mathrm{sig}\left[\mathrm{f}\left[\mathrm{x}_i, \boldsymbol{\phi}\right]\right]} \frac{\exp\left[-\mathrm{f}\left[\mathrm{x}_i, \boldsymbol{\phi}\right]\right]}{\left(1-\exp\left[-\mathrm{f}\left[\mathrm{x}_i, \boldsymbol{\phi}\right]\right]\right)^2} \\ -y_i & \frac{1}{\mathrm{sig}\left[\mathrm{f}\left[\mathrm{x}_i, \boldsymbol{\phi}\right]\right]} \frac{\exp\left[-\mathrm{f}\left[\mathbf{x}_i, \boldsymbol{\phi}\right]\right]}{\left(1+\exp\left[-\mathrm{f}\left[\mathrm{x}_i, \boldsymbol{\phi}\right]\right]\right)^2}\end{aligned}$$

- Show that for $\mathrm{z} = \beta + \Omega h$:

$$\tfrac{\partial z}{\partial h} = \Omega^T$$

where $\frac{\partial z}{\partial h}$ is a matrix containing the term $\frac{\partial z_i}{\partial h_j}$ in its ith column and jth row. To do this, first find an expression for the constituent elements $\frac{\partial z_i}{\partial h_j}$ , and then consider the form that the matrix $\frac{\partial z}{\partial h}$ must take.

solution:

$$z_i = \beta_i + \sum_j \omega_{ij} h_j$$

and so when we take the derivative, we get:

$$\frac{\partial z_i}{\partial h_j} = \omega_{ij}$$

This will be at the $i^{th}$ column and $j^{th}$ row (i.e., at position j,i) of the matrix $\partial \mathbf{z} / \partial \mathbf{h}$ which is hence $\mathbf{\Omega}^T$.