Mid-Term: Question 4, Part 2a

Ian Dover

March 2023

The loss function in this section can be written as:

$$L(w, \alpha, \beta) = \sum_{i=1}^{n} (y_i - \sigma(w^T z_i))^2$$

1 Part A

Derive the gradient of the loss function $L(w, \alpha, \beta)$. Simplify the gradients by assuming that $u_i = w^T z_i$. Use Chain rule and set $\sigma'(x) = \sigma(x)[1 - \sigma(x)]$ The gradient requires that we take the derivative with respect to all variables in the domain. The gradient (∇) of our function can be written as:

$$\nabla L(w, \alpha, \beta) = \left[\frac{\partial L(w, \alpha, \beta)}{\partial w}, \frac{\partial L(w, \alpha, \beta)}{\partial \alpha}, \frac{\partial L(w, \alpha, \beta)}{\partial \sigma} \right]$$

We must now get each element of the gradient vector using the Chain rule:

$$\frac{\partial L(w, \alpha, \beta)}{\partial w} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(w^T z_i)) * \sigma(w^T z_i)[1 - \sigma(w^T z_i)] * z_i$$

$$\frac{\partial L(w, \alpha, \beta)}{\partial \alpha} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(w^T z_i)) * \sigma(w^T z_i)[1 - \sigma(w^T z_i)] * w_1 \sigma(\alpha^T x_i)[1 - \sigma(\alpha^T x_i)] * x_i$$

$$\frac{\partial L(w, \alpha, \beta)}{\partial \beta} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(w^T z_i)) * \sigma(w^T z_i)[1 - \sigma(w^T z_i)] * w_2 \sigma(\beta^T x_i)[1 - \sigma(\beta^T x_i)] * x_i$$

These derivatives can be simplified using $u_i = w^T z_i$:

$$\frac{\partial L(w, \alpha, \beta)}{\partial w} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(u_i)) * \sigma(u_i)[1 - \sigma(u_i)] * \frac{\partial u_i}{\partial w}$$

$$\frac{\partial L(w, \alpha, \beta)}{\partial \alpha} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(u_i)) * \sigma(u_i)[1 - \sigma(u_i)] * \frac{\partial u_i}{\partial \alpha}$$

$$\frac{\partial L(w, \alpha, \beta)}{\partial \beta} = -\sum_{i=1}^{n} 2 * (y_i - \sigma(u_i)) * \sigma(u_i)[1 - \sigma(u_i)] * \frac{\partial u_i}{\partial \beta}$$