# Softmax function

Kristian Wichmann

February 8, 2017

## 1 Definition

Given an $n$-dimensional input vector $z$, the *softmax function* (also known as the *normalized exponential function*), has the output:

$$\sigma_i(z) = \frac{e^{z_i}}{\sum_{k=1}^{n} e^{z_k}} \tag{1.1}$$

This means, that the outputs can be interpreted as an discrete probability distribution, since they will always sum to 1.

It will be convenient to give a shorthand for the normalization "constant", so we set:

$$N(z) = \sum_{k=1}^{n} e^{z_k} \tag{1.2}$$

### 1.1 Example

Figure 1 shows the softmax function applied to the set $\{1, 2, 3, \cdots, 8\}$. As is evident, comparatively small values are given much less overall weight than higher ones.

## 2 Derivative

We might now want to differentiate with respect to the component $z_j$, which is done by applying the quotient rule:

$$\frac{\partial \sigma_i(z)}{\partial z_j} = \frac{\partial}{\partial z_j} \frac{e^{z_i}}{N(z)} = \frac{\left(\frac{\partial}{\partial z_j} e^{z_i}\right) N(z) - e^{z_i} \left(\frac{\partial}{\partial z_j} N(z)\right)}{(N(z))^2} \tag{2.1}$$

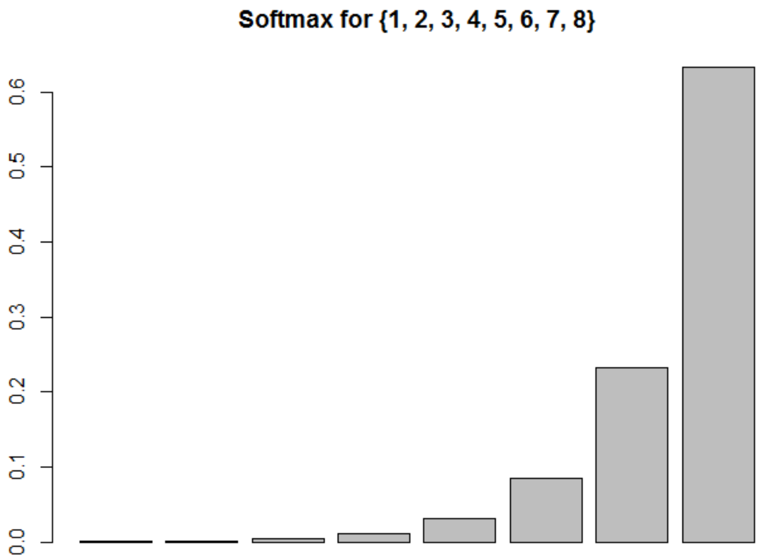


Figure 1: Softmax example

The two derivatives are:

$$\frac{\partial}{\partial z_j} e^{z_i} = \delta_{ij} e^{z_i}, \quad \frac{\partial}{\partial z_j} N(z) = \sum_{k=1}^{n} \delta_{jk} e^{z_k} = e^{z_j} \tag{2.2}$$

Inserting into equation 2.1 this yields:

$$\frac{\delta_{ij} e^{z_j} N(z) - e^{z_i + z_j}}{(N(z))^2} \tag{2.3}$$

The numerator can be rewritten:

$$e^{z_i} \left( \delta_{ij} N(z) - e^{z_j} \right) \tag{2.4}$$

Now divide by $N(z)$ twice, once "outside the parenthesis" and once "inside" to get:

$$\frac{\partial \sigma_i(z)}{\partial z_j} = \frac{e^{z_i}}{N(z)} \left( \delta_{ij} - \frac{e^{z_j}}{N(z)} \right) = \sigma_i(z) \left( \delta_{ij} - \sigma_j(z) \right) \tag{2.5}$$

The likeness to the derivative of the logistic function should be clear.