Loss Functions and Activation Functions

Exercises

1. The logistic function is defined as
$$\sigma(z) = \frac{1}{1+e^{-z}} \; .$$
Prove that
$$\frac{d\sigma(z)}{dz} = \sigma(z)\big(1-\sigma(z)\big) \; .$$

2. Consider the classification problem in which you have 2 classes, C_0 and C_1. Suppose we have a labelled dataset,
$$\left\{\left(\mathbf{x}^{(1)}, t^{(1)}\right), \ldots, \left(\mathbf{x}^{(N)}, t^{(N)}\right)\right\}_i$$
where $\mathbf{x}^{(i)} \in \mathbb{R}^X$ and
$$t^{(i)} = \begin{cases} 0 & \text{if } \mathbf{x}^{(i)} \in C_0 \\ 1 & \text{if } \mathbf{x}^{(i)} \in C_1 \; . \end{cases}$$
Suppose you have a neural network whose operation is represented by the function f, such that for a given input $\mathbf{x}^{(i)}$, the output is
$$y^{(i)} = f\left(\mathbf{x}^{(i)}\,;\,\theta\right)$$
where θ represents all the connection weights and biases in the network. Let us also assume that the output of the network, $y^{(i)} \in (0,1)$, is $P(\mathbf{x}^{(i)} \in C_1)$, the probability that the $\mathbf{x}^{(i)}$ belongs to class C_1.

   Prove that the negative log-likelihood of observing the entire dataset is
$$-\sum_{i=1}^{N}\left[t^{(i)} \ln y^{(i)} + \left(1-t^{(i)}\right)\ln\left(1-y^{(i)}\right)\right] \; .$$

3. Download the jupyter notebook `ex02.ipynb`. It contains two classes: `Identity`, and `MSE`. You can probably guess what they are supposed to do. If not, you can also read their doc strings. However, both classes are missing part of their implementations.

   The notebook contains prompts for the following exercises.

   (a) Complete their `__call__` and `derivative` functions according to the usage specified in the documentation.

   (b) Add some code to the notebook that creates a 2D array `z`, and applies the identity function to it, yielding `y`. Compute the derivative of the identity function for that data. Does it make sense?

   (c) Create another 2D array `t` by adding Gaussian noise to `y` (see `numpy.random.normal`). Then compute the MSE between `y` and `t`. Also, evaluate the derivative of the MSE (with respect to `y`) at `(y, t)`. Does it make sense?