Tutorial Business Analytics

Homework 12

Exercise 12.3 Backpropagation II

Continuing with exercise 12.2, consider again the following feed-forward neural network that consists of

- An input layer ($l = 0$) representing two-dimensional points

$$a^{[0]} = \left(a_1^{[0]}, a_2^{[0]}\right)^T \in \mathbb{R}^2$$

- A hidden layer $l = 1$ with 2 hidden nodes and sigmoid activation function $g^{[1]}$
- An output layer $l = 2$ with one node and sigmoid activation function $g^{[2]}$.

a) Given the following dataset of $n = 4$ observations (in columns) and initial parameters, perform one forward pass (calculate the *empirical risk* $\mathcal{L}$).

$$X = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & 1 & 1 & 1 \end{pmatrix}$$

$$W^{[1]} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad b^{[1]} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad W^{[2]} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \quad b^{[2]} = (0)$$

b) Perform a partial backward pass and update the parameter $W^{[2]}$ using a gradient descent with learning rate $\alpha = 1$. Perform another forward pass on the full data. Did the risk decrease?

Exercise 12.4 [Programming Exercise]

Perform one backward pass through the network of the previous exercise and update the parameters using gradient descent. Use learning rate $\alpha = 1$.

Hint: *For n observations, the full parameter derivatives of the risk function are given by:*

$$dW^{[1]} = \frac{1}{n} dZ^{[1]} X^T, \qquad db^1 = \frac{1}{n} \mathrm{rowSums}(dZ^{[1]}),$$
$$dW^{[2]} = \frac{1}{n} dZ^{[2]} A^{[1]^T}, \qquad \mathrm{d}b^2 = \frac{1}{n} \mathrm{rowSums}(dZ^{[2]})$$

with

$$dZ^{[1]} = W^{[2]^T} dZ^{[2]} * A^{[1]} * (1 - A^{[1]}), \qquad dZ^{[2]} = A^{[2]} - Y$$

where $*$ is element-wise multiplication and for a given expression F we use the shorthand notation $dF = {}^{\partial\mathcal{L}}/_{\partial F}$.