Study Report

Shuo Xu

July 29, 2018

Abstract

This report is mainly about the derivation process of the formulas backpropagation and calculation method of matrix form.

1 Derication pross

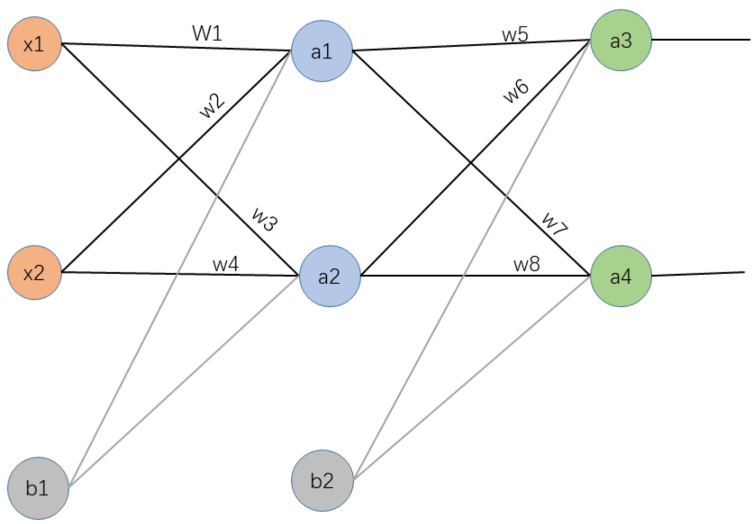

Figure 1: neural network

My derivation is about the picture above, and the following are the formulas I will use.

$$Z = WA + b. \tag{1}$$

$$A = f(Z) = \frac{1}{1+e^{-Z}}. \tag{2}$$

$$L = \frac{1}{2}\sum(Y-A)^2. \tag{3}$$

$$dZ = f'(Z) = f(Z)(1-f(Z)) = A(1-A). \tag{4}$$

The upper case letters in the formula are all matrix forms, I will first complete the computation one by one, and then simplify the expression in matrix form.

Forward propagation

$z_1 = w_1x_1 + w_2x_2 + b_1 \quad z_2 = w_3x_1 + w_4x_2 + b_1 \quad a_1 = \frac{1}{1+e^{-z_1}} \quad a_2 = \frac{1}{1+e^{-z_2}}$

$z_3 = w_5a_1 + w_6a_2 + b_2 \quad z_4 = w_7a_1 + w_8a_2 + b_2 \quad a_3 = \frac{1}{1+e^{-z_3}} \quad a_4 = \frac{1}{1+e^{-z_4}}$

$L = L_1 + L_2 = \frac{1}{2}(y_1 - a_3)^2 + \frac{1}{2}(y_2 - a_4)^2$

If we use the matrix form to simplify the upper expression:

$$A_0 = X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad Z_1 = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad A_1 = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad Z_2 = \begin{bmatrix} z_3 \\ z_4 \end{bmatrix} \quad A_2 = \begin{bmatrix} a_3 \\ a_4 \end{bmatrix}$$

We have known the unmpy package in Python, according to its broadcast, b does not need to be processed. If we want to use matrix to simplify, the W we need are:

$$W_1 = \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix} \quad W_2 = \begin{bmatrix} w_5 & w_6 \\ w_7 & w_8 \end{bmatrix}$$

In this case, we can use the formula 1 and 2, and if we use many examples to make the matrix X, these formulas are also effective. In order to facilitate my expression, I do not list this situation here.

Backward propagation

Because the backward propagation is a little complicated, I will only talk about the top path. According to the chain rule we have:

$da_3 = \frac{\partial L}{\partial a_3} = \frac{\partial L_1}{\partial a_3} = a_3 - y_1 \qquad dz_3 = \frac{\partial L}{\partial z_3} = da_3 \frac{\partial a_3}{\partial z_3} = da_3 \times a_3(1 - a_3) \qquad dw_5 = \frac{\partial L}{\partial w_5} = dz_3 \frac{\partial z_3}{\partial w_5} = dz_3 \times a_1$

In the figure 1 we can see w_5 and w_7 are connected to a_1, if we want to calculate da_1, we need use two routes and that's the same to a_2, b_2 and b_1. Fortunately we don't need to calculate dx, and if we use the matrix, all the thing will be simple.

$da_1 = \frac{\partial L}{\partial a_1} = \frac{\partial L1}{\partial a_1} + \frac{\partial L2}{\partial a_1} = dz_3 \frac{\partial z_3}{\partial a_1} + dz_4 \frac{\partial z_4}{\partial a_1}$