Neural Network Example

January 1, 2017

1 Introduction

This is an example for a simple NN network with backpropagation.

There is a single input, x, with a bias $b_i = -1$.

There are two hidden neurons, h_1, h_2 with a bias $b_h = -1$, with a sigmoid activation function $f_h(v) = 1/(1+e^{-v})$

There is a single output o, with a sigmoid activation function $f_h(v) = 1/(1+e^{-v})$

Hence, these are the weights: w_{i_h} has a size of 2x2 (2 for input+bias and 2 for hidden). w_{h_o} has a size of 3x1 (3 for hidden+bias and 1 for output).

Then:

$h = f(v_h)$, where $v_h = xw_{i_h}$, and
$o = f(v_o)$, where $v_o = hw_{h_o}$.

2 The weights

We will set the weights to have the following values:

$$w_{i_h} = \begin{pmatrix} 0.1 & 0.2 \\ -0.1 & -0.2 \end{pmatrix}$$

$$w_{h_o} = \begin{pmatrix} 0.5 \\ -0.6 \\ 0.7 \end{pmatrix}$$

3 Feed forward example

Let us set an input $x = 1.0$

Then:

$$v_h = \begin{pmatrix} 1.0 & -1 \end{pmatrix} \begin{pmatrix} 0.1 & 0.2 \\ -0.1 & -0.2 \end{pmatrix} = \begin{pmatrix} 1.0 \times 0.1 + (-1) \times (-0.1) \\ 1.0 \times 0.2 + (-1) \times (-0.2) \end{pmatrix} = \begin{pmatrix} 0.2 \\ 0.4 \end{pmatrix}$$

$$h = f(v_h) = \begin{pmatrix} 1/(1+e^{-0.2}) \\ 1/(1+e^{-0.4}) \end{pmatrix} = \begin{pmatrix} 0.55 \\ 0.60 \end{pmatrix}$$

Next, ...

$$v_o = \begin{pmatrix} 0.55 & 0.60 & -1 \end{pmatrix} \begin{pmatrix} 0.5 \\ -0.6 \\ 0.7 \end{pmatrix} = 0.55 \times 0.5 - 0.60 \times 0.6 - 1 \times 0.7 = -0.78$$

$o = f(v_o) = 1/(1 + e^{0.78}) = 0.31$

4 Backpropagation

Let the real output be y.

Then the error $e_o = y - o$ and the energy is $E = 0.5 e_o^2$

The derivative of the logistic function is $df(v) = f(v)(1 - f(v))$

Thus, the derivative of the energy with respect to w_{h_o}, is:
$\frac{dE}{dw_{h_o}} = e_o \frac{de_o}{dw_{h_o}} = e_o df(o) \frac{dv_o}{dw_{h_o}} = e_o df(o) h$.

The dimensions are: $|e_o| = 1, |df(o)| = 1, |h| = (2+1) \times 1$, thus $|\frac{dE}{dw_{h_o}}| = 3 \times 1$, which is the same dimension as dw_{h_o}, as it should be.

Also, the derivative of the energy with respect to w_{i_h}, is:
$\frac{dE}{dw_{i_h}} = e_o \frac{de_o}{dw_{i_h}} = e_o df(o) \frac{dv_o}{dw_{i_h}} = e_o df(o) w_{h_o} \frac{dh}{dw_{i_h}} = e_o df(o) w_{h_o} df(h) \frac{dv_h}{dw_{i_h}} = e_o df(o) w_{h_o} df(h) x$.

The dimesnions are: $|e_o| = 1, |df(o)| = 1, |w_{h_o}| = 1 \times 2 (!!! no bias), |df(h)| = 1 \times 2, |x| = 2 \times 1$. Above we can define $e_h = e_o df(o) w_{h_o}$, which has the dimension $|e_h| = 1 \times 2$. Thus, we can see the backpropagation: $\frac{dE}{dw_{i_h}} = e_h df(h) x$, where $e_h df(h)$ is an *element-wise multiplication*.

5 Backpropagation example

Let us set $y = 0.81$. Then:

$e_o = 0.81 - 0.31 = 0.5$

$df(o) = f(v_o)(1 - f(v_o)) = 0.31 \times 0.69 = 0.21$

$h = \begin{pmatrix} 0.55 \\ 0.60 \\ -1 \end{pmatrix}$.

Hence,

$$\frac{dE}{dw_{h_o}} = 0.5 \times 0.21 \times \begin{pmatrix} 0.55 \\ 0.60 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.05 \\ 0.06 \\ 0.10 \end{pmatrix}$$

We also have:

$$e_h = 0.5 \times 0.21 \times \begin{pmatrix} 0.5 \\ -0.6 \end{pmatrix} = \begin{pmatrix} 0.05 \\ -0.06 \end{pmatrix}$$

$$df(h) = f(v_h)(1 - f(v_h)) = \begin{pmatrix} 0.55 \\ 0.60 \end{pmatrix} \times^{\text{element-wise}} \begin{pmatrix} 1 - 0.55 \\ 1 - 0.60 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.24 \end{pmatrix}$$

$$\frac{dE}{dw_{i_h}} = e_h df(h) x = \left[\begin{pmatrix} 0.05 \\ -0.06 \end{pmatrix} \times^{\text{element-wise}} \begin{pmatrix} 0.25 \\ 0.24 \end{pmatrix} \right] \begin{pmatrix} 1.0 & -1 \end{pmatrix} = \begin{pmatrix} 0.01 \\ -0.01 \end{pmatrix} \begin{pmatrix} 1.0 & -1 \end{pmatrix} = \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.01 \end{pmatrix}$$

6 Changing the weights

Now that we have all the derivatives, we set the learning rate $\eta = 0.9$ (it is large just so we will see the change!!!) and get the new weights:

$$w_{h_o} = \begin{pmatrix} 0.5 \\ -0.6 \\ 0.7 \end{pmatrix} + 0.9 \begin{pmatrix} 0.05 \\ 0.06 \\ 0.10 \end{pmatrix} = \begin{pmatrix} 0.54 \\ -0.03 \\ 0.79 \end{pmatrix}$$

$$w_{i_h} = \begin{pmatrix} 0.1 & 0.2 \\ -0.1 & -0.2 \end{pmatrix} + 0.9 \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.01 \end{pmatrix} = \begin{pmatrix} 0.11 & 0.19 \\ -0.11 & -0.19 \end{pmatrix}$$