The Maths behind Neural Networks

Alex Punnen

Contents

The Maths behind Neural Networks

Alex Punnen
© All Rights Reserved

Contents

Chapter 5

Back propagation - Pass 2 (Scalar Calculus)

Let's consider the following two layer neural network

```
x is the Input
y is the Output.
l is the number of layers of the Neural Network.
a is the activation function ,(we use sigmoid here)
```

$$x \to a^{l-1} \to a^l \to y$$

Where the activation a^l is sigmoid here

$$a^l = \sigma(w^l a^{l-1} + b^l).$$

We will try to derive out via back propogation the effect of the weights in an inner layer to the final loss.

There is one caveat here;these equations are just illustrative with respect to scalar calculus and not accounting for the matrix calculus we will need when modelling a practical neural network. The maths is a bit complex with matrix calculus and we can start with the simplified discourse with Chain rule.

We can write a^l as

$$ a^{l} = (z^l) where

z^l =w^l a^{l-1} +b^l $$

We can also easily calculate

$$\frac{\mathbf{a^l}}{\mathbf{z^l}} = \frac{(\mathbf{z^l})}{\mathbf{z^l}} = {}'(\mathbf{a^l}) \quad \rightarrow (\mathbf{a})$$

Where σ' = derivative of Sigmoid with respect to Weight

Note x can also be written as a^0

Our two layer neural network can be written as

$$\mathbf{a^0} \rightarrow \mathbf{a^1} \rightarrow \mathbf{a^2} \rightarrow \mathbf{y}$$

Note that a^2 does not denote the exponent but just that it is of layer 2.

Lets write down the Chain rule first.

$$\frac{\mathbf{C}}{\mathbf{w^l}} = \frac{\mathbf{z^l}}{\mathbf{w^l}}.\frac{\mathbf{a^l}}{\mathbf{z^l}}.\frac{\mathbf{C}}{\mathbf{a^l}} = \frac{\mathbf{a^l}}{\mathbf{w^l}}.\frac{\mathbf{C}}{\mathbf{a^l}}$$

We will use the above equation as the basis for the rest of the chapter.

Gradient Vector of Loss function In Output Layer

Lets substitute l and get the gradient of the Cost with respect to weights in layer 2 and layer 1.

For the last layer - Layer 2

$$\frac{\mathbf{C}}{\mathbf{w^2}} = \frac{\mathbf{z^2}}{\mathbf{w^2}}.\frac{\mathbf{a^2}}{\mathbf{z^2}}.\frac{\mathbf{C}}{\mathbf{a^2}}$$

The first term is

$$\frac{\partial z^2}{\partial w^2} = \frac{\partial(a^1.w^2)}{\partial w^2} = a^1 \quad \to (1.1)$$

The second term is

$$\frac{\partial a^2}{\partial z^2} = \frac{\partial \sigma(z^2)}{\partial z^2} = \sigma'(z^2) \quad \to (1.2)$$

The third term is

$$\frac{\mathbf{C}}{(\mathbf{a^2})} = \frac{(\frac{\mathbf{1}}{\mathbf{2}}\|\mathbf{y}-\mathbf{a^2}\|^{\mathbf{2}})}{(\mathbf{a^2})} = \frac{\mathbf{1}}{\mathbf{2}} * \mathbf{2} * (\mathbf{a^2}-\mathbf{y}) = (\mathbf{a^2}-\mathbf{y}) \to (\mathbf{1.3})$$

Putting 1.1,2.1 & 3.1 together we get the final equation for the second layer. This is the output layer.

$$\boxed{\frac{\mathbf{C}}{\mathbf{w^2}} = \mathbf{a^1} * \ '(\mathbf{z^2}) * (\mathbf{a^2}-\mathbf{y}) \quad \to (\mathbf{A})}$$

Gradient Vector of Loss function in Inner Layer

Now let's do the same for the inner layer.

$$\frac{\partial C}{\partial w^1} = \frac{\partial z^1}{\partial w^1}.\frac{\partial a^1}{\partial z^1}.\frac{\partial C}{\partial a^1}$$

The first term is similar to (1.1)

$$\frac{\partial z^1}{\partial w^1} = \frac{\partial a^0.w^1}{\partial w^1} = a^0 \quad \to (2.1)$$

The second term is also similar to (1.2)

$$\frac{\partial a^1}{\partial z^1} = \frac{\partial \sigma(z^1)}{\partial z^1} = \sigma'(z^1) \quad \to (2.2)$$

For the third part, we use Chain Rule to split like below, the first part of which we calculated in the earlier step. This is where Chain Rule helps.

$$\frac{\partial C}{\partial (a^1)} = \frac{\partial C}{\partial (a^2)}.\frac{\partial (a^2)}{\partial (a^1)}$$

Note that in the previous section we had calculated

$$\frac{\partial C}{\partial (a^2)} = (a^2 - y) \rightarrow (2.3.1)$$

$$\text{Now to calculate} \quad \frac{\partial (a^2)}{\partial (a^1)}$$

$$\frac{\partial (a^2)}{\partial (a^1)} = \frac{\partial (a^2)}{\partial (z^2)} . \frac{\partial (z2)}{\partial (a^1)}$$

$$= \frac{\partial \sigma (z^2)}{\partial (z^2)} . \frac{\partial (w^2 . a^1)}{\partial (a^1)}$$

$$= \sigma' (z^2) . w^2$$

$$\frac{\partial (a^2)}{\partial (a^1)} = \sigma' (z^2) . w^2 \quad \rightarrow (2.3.2)$$

Putting (2.1),(2.2),(2.3.1)and (2.3.2) together, we get

$$\frac{\mathbf{C}}{\mathbf{w^1}} = \mathbf{a^0} * \ '(\mathbf{z^1}) * (\mathbf{a^2} - \mathbf{y}). \ '(\mathbf{z^2}).\mathbf{w^2} \quad \rightarrow (\mathbf{B})$$

Repeating here the previous equation (A) as well

$$\frac{\mathbf{C}}{\mathbf{w^2}} = \mathbf{a^1} * \ '(\mathbf{z^2}) * (\mathbf{a^2} - \mathbf{y}) \quad \rightarrow (\mathbf{A})$$

- Next A Simple NeuralNet with Back Propagation