1 Neural network simplification

Simplification The network can be simplified to some degree using some maths.

Definition of output neuron The output neuron is is defined as the following.

$$O = sign(B(W * S)) \tag{1}$$

Where B is the batchnorm function, W is the weight matrix, and S is the synapses (synapses are the output of the previous layer). "sign" is the sign function, giving -1 if the number is negative and 1 if it is positive. The batchnorm function is defined as follows.

$$B(x) = \gamma * \sigma^{-1} * (x - \bar{x}) + \beta \tag{2}$$

γ, σ^{-1}, $\bar{x}$ and β is provided with the network. This function will be applied on each element in the $W * S$ vector. The following equations will show a simplification of this function.

$$B(x) = \gamma * \sigma^{-1} * (x - \bar{x}) + \beta \tag{3}$$

$$B(x) = \gamma * \sigma^{-1} * x - \gamma * \sigma^{-1} * \bar{x} + \beta \tag{4}$$

This can be rewritten as

$$B(x) = a * x + b \tag{5}$$

Where

$$a = \gamma * \sigma^{-1} \tag{6}$$

and

$$b = -\gamma * \sigma^{-1} * \bar{x} + \beta \tag{7}$$

Since we are only interested in the sign of $B(x)$, it is possible to simplify even more. Multiplying by $\frac{1}{a}$, will not change the sign of the expression. The multiplication might change the sign of $a * x$, but will also change the sign of b by the same amount.

$$sign(B(x)) = sign(a * x + b) = sign((a * x + b) * \frac{1}{a}) = sign(x + \frac{b}{a}) \tag{8}$$

In the case of pre trained neural networks, the values of a and b will not change. Then the $\frac{b}{a}$ can be precomputed, saving time when feeding forward inputs. The sign operation of a subtractions is really just a comparison. So in this case, it is sufficient to compare the values of x and $\frac{b}{a}$ and give an output accordingly.

Summary Each layer in the feed forward network can be simplified by the binary matrix multiplication with the previous output $W * S$, and then comparing with a single vector to determine if the output should be 1 or -1 (Compare to $\frac{b}{a}$).