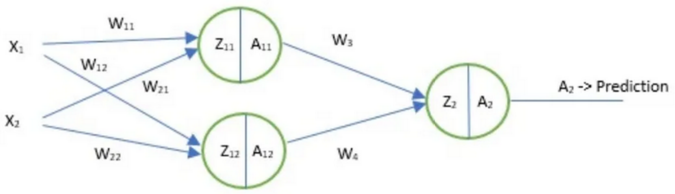
XOR single-layer model theory

Max Cotton

1 Setup

X_1	X_2	$Y = X_1$ XOR X_2
0	0	0
0	1	1
1	0	1
1	1	0



- Where the weights W11, W12, W21 and W22 are together in an array W1, as the hidden weights, initially given random values
- Weights W3 and W4 are together in array W2, as the output weights
- Z11 and Z12 are together in an array Z1, and is the dot product of the W1 array and the input array (X)
- A11 and A12 are together in an array A1, and is sigmoid(Z1)
  - Where $sigmoid(Z) = \frac{1}{1+e^{-Z}}$
- Z2 is the dot product of the W2 array and A1
- A2 is sigmoid(Z2), which is the prediction

2 Forward Propagation

For each epoch the input array consisting of a combination of a 0 and/or 1, is fed through the network to obtain a prediction A2

3 Back Propagation

- Once a prediction is obtained, you then move back through the network adjusting the weights
- The "Loss" (L) (how wrong the prediction is) can be calculated with the Loss function, which calculates the average difference between the prediction (A2) and the actual value (Y), the average is found by summing the result for each prediction then dividing by the number of predictions (m). This shows how well the network is performing.

- Where $L = -(\frac{1}{m}) * \sum(Y * log(A2) + (1-Y) * log(1-A2))$ and "log()" is the natural logarithm (e is the base)

- Weights are adjusted via Gradient Descent, which aims to get the minimum Loss value, with the following formula:
  - $W = W - learningRate * \frac{\partial L}{\partial W}$
  - Where $\frac{\partial L}{\partial W2} = (A2 - Y) * A1$
  - And $\frac{\partial L}{\partial W1} = X * A1 * (1 - A1) * W2 * (A2 - Y)$

4 Derivations for $\frac{\partial L}{\partial W2}$ and $\frac{\partial L}{\partial W1}$

- Functions used so far:
  1. $Z1 = W1.X$
  2. $A1 = \frac{1}{1+e^{-Z1}}$
  3. $Z2 = W2.A1$
  4. $A2 = \frac{1}{1+e^{-Z2}}$
  5. $L = -(\frac{1}{m}) * \sum(Y * log(A2) + (1 - Y) * log(1 - A2))$

- $\frac{\partial L}{\partial W2}$ derivation:
  - By the Chain rule, $\frac{\partial L}{\partial W2} = \frac{\partial L}{\partial A2} * \frac{\partial A2}{\partial Z2} * \frac{\partial Z2}{\partial W2}$
  - Using function 5, $\frac{\partial L}{\partial A2} = (-\frac{1}{m})(\frac{Y-A2}{A2(1-A2)})$
  - Using function 4, $\frac{\partial A2}{\partial Z2} = A2 * (1 - A2)$
  - Using function 3, $\frac{\partial Z2}{\partial W2} = A1$
  - => $\frac{\partial L}{\partial W2} = (-\frac{1}{m})(\frac{Y-A2}{A2(1-A2)}) * A2 * (1 - A2) * A1$
    $= (A2 - Y) * A1$

- $\frac{\partial L}{\partial W1}$ derivation:
  - By the chain rule, $\frac{\partial L}{\partial W1} = \frac{\partial A1}{\partial Z1} * \frac{\partial Z1}{\partial W1} * \frac{\partial L}{\partial Z2} * \frac{\partial Z2}{\partial A1}$
  - Using function 2, $\frac{\partial A1}{\partial Z1} = A1 * (1 - A1)$
  - Using function 1, $\frac{\partial Z1}{\partial W1} = X$
  - Using function 5, $\frac{\partial L}{\partial Z2} = A2 - Y$
  - Using function 3, $\frac{\partial Z2}{\partial A1} = W2$
  - => $\frac{\partial L}{\partial W1} = X * A1 * (1 - A1) * W2 * (A2 - Y)$