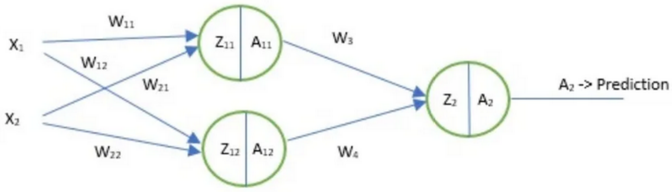
XOR single-layer model theory

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 Y

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 Y

3 Back Propagation

- Once a prediction is obtained, you then move back through the network adjusting the weights
- The "Cost" (how wrong the prediction is) can be calculated with the cost function, which calculates the average squared difference between the prediction and the actual values. This shows how well the network is performing.

- Where $Cost = -(\frac{1}{nInputs}) * \sum(Y * log(A2) + (1 - Y) * log(1 - A2))$

- Weights are adjusted via Gradient Descent, which aims to get the minimum cost 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 $\frac{\partial L}{\partial W1}$