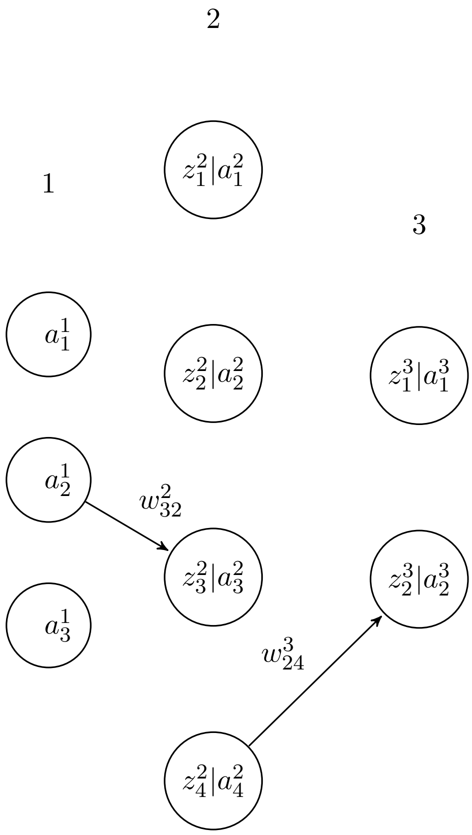
BACKPROPAGATION

Backpropagation is a computer algorithm used for supervised training of Artificial Neural Networks. It is based on two ideas - the Chain Rule of derivatives in Calculus and Dynamic Programming.

1 Notation



The layers are of the Neural Network are denoted by positive integers with input layer denoted by 1, the first hidden layer by 2 and so on till the output layer. The inputs in the input layer are denoted by a_i^1 as shown above. For all the other neurons in other layers, z_i^l denotes the weighted sum of the neuron activations in the previous layer and a_i^l denotes that neuron's activation. The weights are superscripted by the index of the layer of the neuron at the end, and subscripted first by the layer index of the end neuron and then by that of the start neuron.

2 Forward Pass

For any layer l which is apart from the input layer, it can be written,

$$z_i^l = \left(\sum_j w_{ij}^l a_j^{l-1} \right) + b_i^l$$

Taking examples from the figure,

$$\begin{aligned} z_1^2 &= (w_{11}^2 a_1^1 + w_{12}^2 a_2^1 + w_{13}^2 a_3^1) + b_1^2 \\ z_2^2 &= (w_{21}^2 a_1^1 + w_{22}^2 a_2^1 + w_{23}^2 a_3^1) + b_2^2 \\ z_3^2 &= (w_{31}^2 a_1^1 + w_{32}^2 a_2^1 + w_{33}^2 a_3^1) + b_3^2 \\ z_4^2 &= (w_{41}^2 a_1^1 + w_{42}^2 a_2^1 + w_{43}^2 a_3^1) + b_4^2 \end{aligned}$$

These equations can be rewritten in matrix form as follows,

$$\begin{pmatrix} z_1^2 \\ z_2^2 \\ z_3^2 \\ z_4^2 \end{pmatrix} = \begin{pmatrix} w_{11}^2 & w_{12}^2 & w_{13}^2 \\ w_{21}^2 & w_{22}^2 & w_{23}^2 \\ w_{31}^2 & w_{32}^2 & w_{33}^2 \\ w_{41}^2 & w_{42}^2 & w_{43}^2 \end{pmatrix} \begin{pmatrix} a_1^1 \\ a_2^1 \\ a_3^1 \end{pmatrix} + \begin{pmatrix} b_1^2 \\ b_2^2 \\ b_3^2 \\ b_4^2 \end{pmatrix}$$

or

$$\boldsymbol{Z}^2 = \boldsymbol{W}^2 \boldsymbol{A}^1 + \boldsymbol{B}^2$$

Generalizing for all layers,

$$\boldsymbol{Z}^l = \boldsymbol{W}^l \boldsymbol{A}^{l-1} + \boldsymbol{B}^l \tag{1}$$

Activation function σ is applied to z_i^l to yield a_i^l.

$$\begin{aligned} a_1^2 &= \sigma(z_1^2) \\ a_2^2 &= \sigma(z_2^2) \\ a_3^2 &= \sigma(z_3^2) \\ a_4^2 &= \sigma(z_4^2) \end{aligned}$$

which can be written in matrix form as,

$$\boldsymbol{Z}^2 = \sigma(\boldsymbol{A}^2)$$

or more generally as,

$$\boldsymbol{Z}^l = \sigma(\boldsymbol{A}^l) \tag{2}$$

Given $\boldsymbol{A}^1$, any $\boldsymbol{Z}^l$ or $\boldsymbol{A}^l$ can be calculated using the two equations (1) and (2). This completes the analysis of the forward pass of backpropagation algorithm.

3 Backward Pass