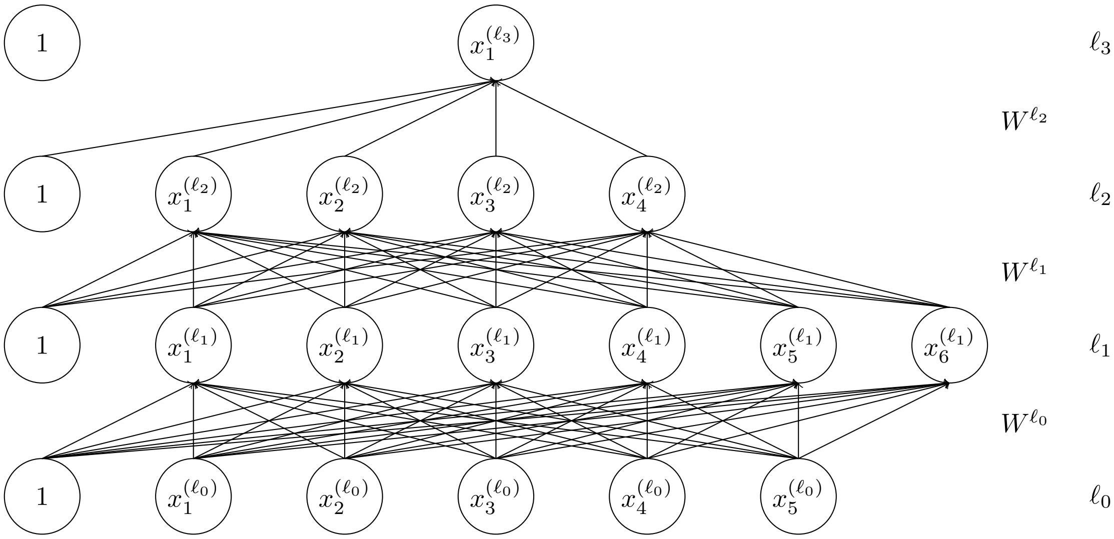

ℓ_k is the k^{th} layer of the network. There are $n+1$ total layers. ($n=3$ in the depicted network)
$\vec{x}^{(\ell_k)}$ is the vector of inputs at layer k, the 0^{th} entry of which is always 1 (for the bias/intercept).
$\dim(\vec{x}^{(\ell_k)}) = d_k$. In the depicted network, $\vec{d} = \langle 6, 7, 5, 2 \rangle$.
$W^{(\ell_k)}$ is the weight matrix connecting ℓ_k to ℓ_{k+1}.
$g_k = \lambda \vec{x}. \left\langle 1 | \theta(W^{(\ell_k)T}\vec{x}) \right\rangle$ = Propagation function at layer k
$|$ denotes vector concatenation, and θ is the network's activation function (e.g. $\theta = \tanh$).
$\vec{x}^{(\ell_{k+1})} = g_k(\vec{x}^{(\ell_k)}) = \left\langle 1 | \theta(W^{(\ell_k)T}\vec{x}^{(\ell_k)}) \right\rangle$
$\dim(W^{(\ell_k)}) = d_k \times (d_{k+1} - 1)$
$f(\vec{x})$ = The function that the neural network approximates
$\tilde{f}(\vec{x})$ = The output of the neural network = $\vec{x}^{(\ell_n)}$
$\tilde{f} = \bigcirc_{k=0}^{n-1} g_k$, where $\bigcirc$ represents iterated function composition.
$E(\vec{x}) = \frac{1}{2} ||f(\vec{x}) - \tilde{f}(\vec{x})||^2$

$$\frac{\partial E}{\partial W^{(\ell_{n-1})}}(\vec{x}) = ||f(\vec{x}) - \tilde{f}(\vec{x})||$$

$$\frac{\partial E}{\partial W_{ij}^{(\ell_k)}} = \frac{\partial E}{\partial \theta(W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j} \frac{\partial \theta(W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j}{\partial (W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j} \frac{\partial (W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j}{\partial W_{ij}^{(\ell_k)}}$$

$$\frac{\partial E}{\partial \theta(W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j} = \Sigma_{h=0}^{d_{k+1}} \left(\frac{\partial E}{\partial \theta(W^{(\ell_{k+1})T}\vec{x}^{(\ell_{k+1})})_h} \frac{\partial \theta(W^{(\ell_{k+1})T}\vec{x}^{(\ell_{k+1})})_h}{\partial (W^{(\ell_k)T}\vec{x}^{(\ell_k)})_h} W_{jh}^{(\ell_{k+1})} \right)$$

$$\frac{\partial \theta(W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j}{\partial (W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j} = \theta'(W^{(\ell_k)T}\vec{x}^{(\ell_k)})$$

$$\frac{\partial (W^{(\ell_k)T}\vec{x}^{(\ell_k)})_j}{\partial W_{ij}^{(\ell_k)}} = \frac{\partial}{\partial W_{ij}^{(\ell_k)}} \Sigma_{h=0}^{d_k} W_{hj}^{(\ell_k)} x_h^{(\ell_k)} = x_i^{(\ell_k)}$$