Deriving Backpropogation:-

Hidden layer - Output layer.

The weight update eqn we are trying to derive is.

$$\hat{w}_1^{(l+1)} = \hat{w}_1^{(l)} - \eta \cdot \nabla E(\hat{w}_1^{(l)})$$

$$E = \frac{1}{2} \sum_{i=1}^{m} \left(y^{(i)} - t^{(i)}\right)^2.$$

$$\nabla E = \frac{d}{dw} \frac{1}{2} \sum_{i=1}^{m} \left(y^{(i)} - t^{(i)}\right)^2.$$

Using chain rule.

$$\nabla E(w) = \frac{1}{2} \sum_{i=1}^{m} 2\left(y^{(i)} - t^{(i)}\right) \frac{d}{d\hat{w}_1} \left(y^{(i)} - t^{(i)}\right)$$

$$\nabla E(w) = \sum_{i=1}^{m} \left(y^{(i)} - t^{(i)} \frac{d}{d\hat{w}_1} \left(y^{(i)} - t^{(i)}\right)\right).$$

$$\nabla E(w) = \sum_{i=1}^{m} \left(y^{(i)} - t^{(i)}\right) \frac{d}{dw} y^{(i)})$$

$$= \sum_{i=1}^{m} \left(y^{(i)} - t^{(i)}\right) \frac{d}{d\hat{w}_1} g(\hat{r}^{(i)}).$$

where $\hat{r}$ is the input to the neuron in output layer.

$$\nabla E(\hat{w}_1) = \sum_{i=1}^{n} (y^{(i)} - t^{(i)}) \frac{\partial}{\partial w_1} g\left(\left(\sum_{j=1}^{n'} g_j(r_j^{(i)}) \cdot \hat{w}_1 \right) + \hat{b} \right).$$

$\hat{b}$ is the bias of output neuron. Again using chain rule,

$$\nabla E(\hat{w}) = \sum_{i=1}^{n} (y^{(i)} - t^{(i)}) g'(\hat{r}^{(i)}) \cdot \frac{\partial}{\partial w} \left(\left(\cdots \right. \right.$$
$$\left. \left. \cdots \sum_{j=1}^{n_1} g_j(r_j^{(i)} \cdot \hat{w}_i) + \hat{b} \right) \right)$$

$$\Rightarrow \nabla E(w) = \sum_{i=1}^{n} (y^{(i)} - t^{(i)} g'(\hat{r}^{(i)}) \cdot \frac{\partial}{\partial w} \left(g(r^{(i)}) \cdot \hat{w}_1 \right) + \hat{b})$$

$$= \sum_{i=1}^{n} (y^{(i)} - t^{(i)}) g'(\hat{r}^{(i)}) \cdot g(r_1^{(i)}).$$

Formula for weight update:

$$w^{(k+1)} = w_1^{(k)} - h \cdot \sum_{i=1}^{n} (y^{(i)} - t^{(i)}) g'(\hat{r}^{(i)}) \cdot g(r^{(i)}).$$

Input layer – Hidden layer.

$$w^{(l+1)} = w^{(l)} - \eta \cdot \nabla E(w^{(l)})$$

$$E = \frac{1}{2} \sum_{i=1}^{n} \left(y^{(i)} - t^{(i)} \right)^2 .$$

$$\frac{\partial}{\partial w} \nabla E = \frac{\partial}{\partial w^{(l)}} \frac{1}{2} \sum_{i=1}^{n} \left(y^{(i)} - t^{(i)} \right)^2 .$$

$$\Rightarrow \frac{1}{2} \sum_{i=1}^{n} 2 \left(y^{(i)} - t^{(i)} \right) \frac{\partial}{\partial w} \left(y^{(i)} - t^{(i)} \right)$$

$$= \sum_{i=1}^{n} \left(y^{(i)} - t^{(i)} \right) \frac{\partial}{\partial w} y^{(i)} =$$

$$\sum_{i=1}^{n} \left(y^{(i)} - t^{(i)} \right) \frac{\partial}{\partial w} g(\hat{r}^{(i)}) .$$