output = act(w3 * act(w2 * act(w1 * input)))

$$\frac{\partial \text{output}}{\partial w_1} = \frac{\partial \text{output}}{\partial \text{hidden}_2} * \frac{\partial \text{hidden}_2}{\partial \text{hidden}_1} * \frac{\partial \text{hidden}_1}{\partial w_1}$$

$$\frac{\partial \text{error}}{\partial w_1} = \frac{\partial \text{error}}{\partial \text{output}} * \frac{\partial \text{output}}{\partial \text{hidden}_2} * \frac{\partial \text{hidden}_2}{\partial \text{hidden}_1} * \frac{\partial \text{hidden}_1}{\partial w_1}$$

actual output - a, predicted - ϕ.

$$\text{error} = J = \frac{1}{2}(\vec{\phi} - \vec{a})^2$$

$$\frac{d\,\text{error}}{d\vec{\phi}} = 2 \cdot \frac{1}{2}(\vec{\phi} - \vec{a}) = \vec{\phi} - \vec{a}$$

$$\frac{\partial J}{\partial w^1_{ij}} = (\phi_j - \vec{a}) \frac{\partial \phi_j}{\partial w^1_{ij}}$$

$$\frac{\partial \phi_j}{\partial w'_{ij}} = \frac{\partial \phi_j}{\partial \phi_i} \frac{\partial \phi_i}{\partial w'_{ij}}$$

$$\boxed{\frac{\partial \phi_j}{\partial \phi_i} = \phi_j(1-\phi_j)}$$

similar to derivative of sigmoid function

$$\frac{\partial \phi_j}{\partial w'_{ij}} = \phi_j(1-\phi_j) \frac{\partial \phi_i}{\partial w'_{ij}}$$

$$\frac{\partial z}{\partial w'_{ij}} = (\phi_j - a^T) * \phi_j(1-\phi_j) \frac{\partial \phi_i}{\partial w'_{ij}}$$

$$\frac{\partial \phi_i}{\partial w'_{ij}} = \sum_j \frac{\partial \phi_i}{\partial z_j} \frac{\partial z_j}{\partial w'_{ij}}$$

$$\frac{\partial \phi_i}{\partial w'_{ij}} = w^3_{ij} * \frac{\partial z_j}{\partial w'_{ij}}$$

$$\frac{\partial z_j}{\partial w'_{ij}} = \frac{\partial z_j}{\partial z_i} \frac{\partial z_i}{\partial w'_{ij}} = z_j(1-z_j) \frac{\partial z_i}{\partial w'_{ij}}$$

$$z\_j = \text{sigmoid}(z\_i)$$

$$\frac{\partial z_i}{\partial w^1_{ij}} = \sum \frac{\partial z_i}{\partial y_j} \frac{\partial y_j}{\partial w^1_{ij}} = \sum w^2_{ij} \frac{\partial y_j}{\partial w^1_{ij}}$$

$$\frac{\partial y_j}{\partial w^1_{ij}} = \frac{\partial y_j}{\partial y_i} \frac{\partial y_i}{\partial w^1_{ij}} = y_j(1-y_j) \frac{\partial y_i}{\partial w^1_{ij}}$$

$$\frac{\partial y_i}{\partial w^1_{ij}} = \sum x_j$$

$$\frac{\partial J}{\partial w^1_{ij}} = (y_j - a^2) * \phi_j (1-\phi_j) * \sum_j [w^3_{ij} * z_j (1-z_j) * \sum_j [w^2_{ij} * y_j (1-y_j) * \sum x_j]]$$

gradient descent :

repeat until $\frac{\partial J}{\partial w^1_{ij}} \to 0$

$$w^1_{ij} = w^1_{ij} - \alpha \frac{\partial J}{\partial w^1_{ij}}$$