

Derivation of Back Propagation Algorithm

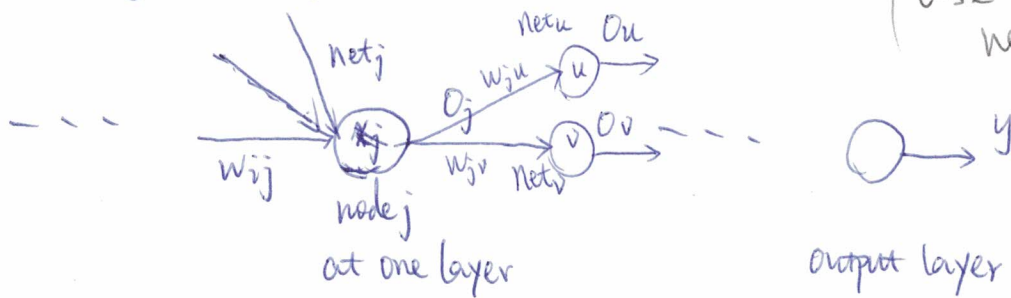
Objective: Minimize network prediction error on training examples

Algorithm: Gradient descent.

Let network output be y , target value be t , define error as $E = \frac{1}{2}(t-y)^2$

Look at gradient of error w.r.t. a weight

(Use the figure on the next page.)



net_j : input that node j receives, $net_j = \sum_{i \in \text{previous layer}} o_i w_{ij}$

o_j : output of node j

Sigmoid function: $o_j = \sigma(net_j) = \frac{1}{1 + e^{-net_j}}$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ij}}$$

$$w_{ij} = w_{ij} - \alpha \frac{\partial E}{\partial w_{ij}}$$

$$\begin{aligned} 1) \quad \frac{\partial o_j}{\partial net_j} &= - \frac{1}{(1 + e^{-net_j})^2} e^{-net_j} (-1) = \frac{1}{1 + e^{-net_j}} \cdot \frac{e^{-net_j}}{1 + e^{-net_j}} \\ &= \frac{1}{1 + e^{-net_j}} \cdot \left(1 - \frac{1}{1 + e^{-net_j}} \right) = o_j \cdot (1 - o_j) \end{aligned}$$

$$2) \quad \frac{\partial net_j}{\partial w_{ij}} = o_i$$

3) Now look at $\frac{\partial E}{\partial O_j}$.

① If node j is the output node, then $O_j = y$.

$$\frac{\partial E}{\partial O_j} = \frac{\partial E}{\partial y} = y - t$$

② For other nodes.

$$\delta_j = \frac{\partial E}{\partial O_j} = \sum_{\text{next layer}} \frac{\partial E}{\partial O_u} \cdot \frac{\partial O_u}{\partial O_j} = \sum_{\text{net layer}} \frac{\partial E}{\partial O_u} \cdot \frac{\partial O_u}{\partial \text{net}_u} \cdot \frac{\partial \text{net}_u}{\partial O_j}$$

$$= \sum_u \left[\frac{\partial E}{\partial O_u} \cdot O_u(1-O_u) \cdot O_j W_{ju} \right] \quad \text{Recursion!}$$

∴ Algorithm: Start with output node, calculate $\frac{\partial E}{\partial O_j}$ layer by layer BACKWARDS!

$$\frac{\partial E}{\partial W_{ij}} = \sum_u \frac{\partial E}{\partial O_u} O_u(1-O_u) W_{ju} O_j(1-O_j) \cdot O_j$$

Vanishing Gradient: δ_j vanishes after repeated multiplication.

As $0 < O_j < 1$. Especially when O_j is close to 0 or 1.

