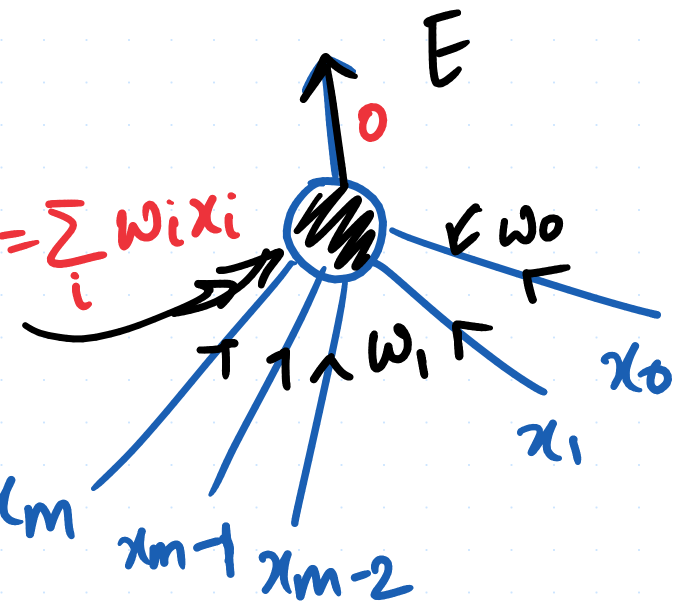
Lecture - 5

15/02/2026

Single sigmoid neuron & cross-entropy loss, derived for single data point, hence dropping the right suffix i.

$net = \sum_i w_i x_i$



$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial o} \cdot \frac{\partial o}{\partial net} \cdot \frac{\partial net}{\partial w_1}$$

$$E = -\underset{\text{target}}{t} \log \underset{\text{output}}{o} - \underbrace{(1-t)\log(1-o)}$$

target   output   Binary crossentropy loss function.

$$\frac{\partial E}{\partial o} = -\frac{t}{o} - \frac{(1-t)}{(1-o)}\cdot(-1)$$

$$= \frac{-t(1-o) + (1-t)o}{o(1-o)}$$

$$= \frac{-t + \cancel{to} + o - \cancel{to}}{o(1-o)} = -\frac{(t-o)}{o(1-o)}$$

$$\frac{\partial E}{\partial o} = -\frac{(t-o)}{o(1-o)}$$

$$\frac{\partial o}{\partial net} = o(1-o).$$

sigmoid's → derivation

output w.r.t. network.

$$\frac{\partial net}{\partial w_1} = \frac{\partial}{\partial w_1}(w_0 x_0 + w_1 x_1 + \dots + w_n x_n)$$

$$\boxed{\frac{\partial net}{\partial w_1} = x_1}.$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial o} \cdot \frac{\partial o}{\partial net} \cdot \frac{\partial net}{\partial w_1}$$

$$= -\frac{(t-o)}{\cancel{o(1-o)}} \times \cancel{o(1-o)} \times x_1$$

$$\boxed{\frac{\partial E}{\partial w_1} = -(t-o)\, x_1}$$

weight change.

$$\boxed{\Delta w_1 = -\eta \frac{\partial E}{\partial w_1} = \eta (t-o)\, x_1}$$

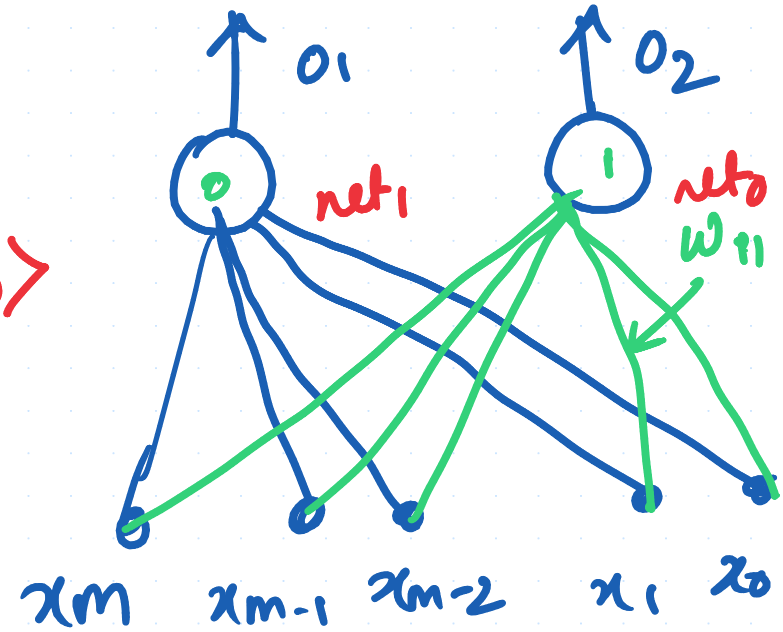
gradient of weight 1 w.r.t. the Error/ Loss term.

Sigmoid Case.

If there are multiple neurons in the o/p layer → Softmax as activation and Cross-entropy loss as the loss function.

$$O = \langle o_1, o_0 \rangle$$

$$NET = \langle net_1, net_0 \rangle$$



$$\frac{\partial w_{11}}{\partial w_{11}} = 1 \quad ✓$$

$$\frac{\partial w_{11}}{\partial w_{01}} = 0 \quad ✓$$

$$o_1 = \frac{e^{net_1}}{e^{net_1} + e^{net_0}} \qquad o_0 = \frac{e^{net_0}}{e^{net_1} + e^{net_0}}$$

$$\frac{\partial O}{\partial NET} = \begin{bmatrix} \frac{\partial O_0}{\partial net_0} & \frac{\partial O_1}{\partial net_0} \\ \frac{\partial O_0}{\partial net_1} & \frac{\partial O_1}{\partial net_1} \end{bmatrix}$$

Jacobian

$$= \begin{bmatrix} O_0(1-O_0) & -O_0 O_1 \\ -O_1 O_0 & O_1(1-O_1) \end{bmatrix}$$

refer to previous class notes.

Jacobian - first order partial derivative.

Hessian - 2nd order partial derivative

m-tensor → Higher order partial derivative

$$f(x,y) = x + y$$

$$J = \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \, \partial x} \\ \frac{\partial^2 f}{\partial x \, \partial y} & \frac{\partial^2 f}{\partial^2 y} \end{bmatrix}$$

$$f(x,y) = x + y$$

$$H = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$f(x,y,z) = x^2 + 2xy + z$$

$$J = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \dfrac{\partial f}{\partial z} \end{bmatrix}$$

$$= \begin{bmatrix} 2x + 2y \,, & 2x \,, & 1 \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x^2} \quad \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$$

Hessian

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x \cdot \partial x} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} \\[2ex] \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y \cdot \partial y} & \dfrac{\partial^2 f}{\partial y \partial z} \\[2ex] \dfrac{\partial^2 f}{\partial z \cdot \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z \partial z} \end{bmatrix}$$

$$H = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E = -t_1 \log o_1 - (1-t_1)\log(1-o_1)$$

$$\frac{\partial E}{\partial w_{11}} = -\frac{t_1}{o_1}\frac{\partial o_1}{\partial w_{11}} - \frac{t_0}{o_0}\frac{\partial o_0}{\partial w_{11}}$$

$$\frac{\partial o_1}{\partial w_{11}} = \frac{\partial o_1}{\partial net_1}\cdot\frac{\partial net_1}{\partial w_{11}} + \frac{\partial o_1}{\partial net_0}\cdot\frac{\partial net_0}{\partial w_{11}}$$

$$= o_1(1-o_1)x_1 + -o_0 o_0 \cdot 0$$

$$= o_1(1-o_1)x.$$

$$\frac{\partial o_0}{\partial w_{11}} = \frac{\partial o_0}{\partial net_1} - \frac{\partial net_1}{\partial w_{11}} + \frac{\partial o_0}{\partial net_0} \frac{\partial net_0}{\partial w_{11}}$$

$$= -o_0 o_1 x + 0.$$

$$\frac{\partial E}{\partial w_{11}} = -\frac{t_1}{o_1} \cdot \frac{\partial o_1}{\partial w_{11}} - \frac{t_0}{o_0} \frac{\partial o_0}{\partial w_{11}}$$

$$= -\frac{t_1}{\cancel{o_1}} \cdot \cancel{o_1}(1-o_1) \cdot x_1 - \frac{t_0}{\cancel{o_0}} (-\cancel{o_0} o_1) x_1$$

$$= -t_1(1-o_1)x_1 + t_0 o_1 x_1$$

$$= x_1(-t_1 + o_1) = -(t_1 - o_1)x_1$$

$$\boxed{\Delta w_{11} = -\eta \frac{\partial E}{\partial w_{11}} = \eta (t_1 - o_1) x_1}$$