Q1) $x \rightarrow \sigma(w^T x + b)$

let $z = \sigma(w^T x + b)$

To train Rosenblatt's perceptron, we min.

$$d = -y(\beta^T z + \beta_0)$$

$y \rightarrow$ true label

Now, $d = -y\left[\beta^T \sigma(w^T x + b) + \beta_0\right]$

For β, $\beta_{new} = \beta_{old} - \eta \frac{\partial d}{\partial \beta} = \beta_{old} - \eta\left\{-y\,\sigma(w^T x + b)\right\}$

$= [0.5]$

For β_0, $\beta_0^{new} \leftarrow \beta_0^{old} - \eta \{-y\}$ — $[0.25]$

For b, $b_{new} \leftarrow b_{old} - \eta \{-y \beta \sigma'(w^T x + b)\}$, where $\sigma'(\cdot) = \sigma(\cdot)(1 - \sigma(\cdot))$ $[.25]$

$w_{new} \leftarrow w_{old} - \eta \{-y \beta \sigma'(w^T x + b) x\}$ — $[.5]$

'η' is learning rate.

Q2) $E(\beta, \beta_0) = -y \log \hat{y} - (1-y) \log(1-\hat{y})$

$$\hat{y} = \sigma(\beta^T x x^T \beta + \beta^T x + \beta_0) = \sigma(\cdot)$$

$$\frac{\partial E}{\partial \beta} = \frac{-y}{\sigma(\cdot)} \sigma(\cdot)(1-\sigma(\cdot))[2xx^T\beta + x]$$

$$- \frac{(1-y)}{1-\sigma(\cdot)} \{-\sigma(\cdot)(1-\sigma(\cdot))\} [2xx^T\beta + x] \qquad - [0.5]$$

$$= -y[1-\sigma(\cdot)][2xx^T\beta + x] + (1-y)\sigma(\cdot)[2xx^T\beta + x]$$

$$\beta_{new} \leftarrow \beta_{old} - \eta \frac{\partial E}{\partial \beta} \quad \Bigg| \quad \beta_0^{new} \leftarrow \beta_0^{old} - \eta \frac{\partial E}{\partial \beta_0}$$

$$\frac{\partial E}{\partial \beta_0} = -y\left[1-\sigma(\cdot)\right] + (1-y)\,\sigma(\cdot)$$ [0.5]

Now, $\beta = [0, 0]^T$, $\beta_0 = 1$, the 1^{st} update for β will be

$$\beta \leftarrow -\eta \frac{\partial E}{\partial \beta} = -\eta\left[-y\left(1-\sigma(\cdot)\right)x + (1-y)\,\sigma(\cdot)\,x\right]$$ [.5]

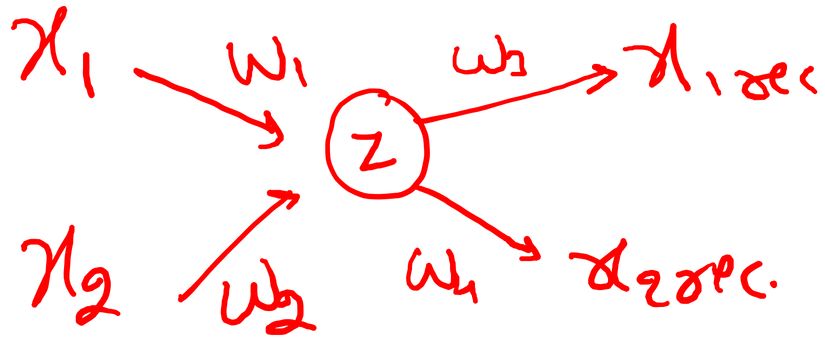
Putting $y=0$, $x = \begin{bmatrix}1\\2\end{bmatrix}$

$$= -\eta\left[-1\left[1-\sigma(1)\right]\begin{bmatrix}1\\2\end{bmatrix} + 0\right],$$

$$= -\eta\left[-1+\sigma(1)\right]\begin{bmatrix}1\\2\end{bmatrix}$$

$$\beta \leftarrow -\begin{bmatrix}1\\2\end{bmatrix}\left\{-\eta + \eta\cdot\frac{1}{1+e^{-1}}\right\}$$ [.5]

Q3)



$z = w_1 x_1 + w_2 x_2$

$x_{1rec} = w_3 z = w_3[w_1 x_1 + w_2 x_2]$ $\quad - [.5]$

$x_{2rec} = w_4 z = w_4[w_1 x_1 + w_2 x_2]$ $\quad - [.5]$

we want w_1 to be close to w_4, i.e. $(w_1 - w_4)^2$

Total cost function:
$$\min_{\substack{w_1, w_2 \\ w_3, w_4}} \; (x_{1rec} - x_1)^2 + (x_{2rec} - x_2)^2 + (w_1 - w_4)^2 + (w_2 - w_3)^2$$

$$\frac{\partial}{\partial w_1} = 2(x_{1rec} - x_1)\frac{\partial x_{1rec}}{\partial w_1} + 2(x_{2rec} - x_2)\frac{\partial x_{2rec}}{\partial w_1} + 2(w_1 - w_4)$$

$$= 2(x_{1rec} - x_1)\, w_3 x_1 + 2(x_{2rec} - x_2)\, w_4 x_1 + 2(w_1 - w_4)$$

$$\boxed{.5}$$

$w_1 \leftarrow w_1 - \eta \frac{\partial}{\partial w_1}$ update for w_1

Q4) O/P at 1st hidden node $= \sum_{i=1}^{3} U_i x_i = Z_1$ [.5]

O/P at 2nd hidden node : $\sum_{i=1}^{3} V_i x_i = Z_2$ [.5]

O/P at output node : $\sum_{i=1}^{2} Z_i W_i = \omega_1(U_1x_1+U_2x_2+U_3x_3) + \omega_2(V_1x_1+V_2x_2+V_3x_3)$

$\hat{y} = \sigma\left\{\sum_{i=1}^{2} Z_i W_i\right\}$ [.5]

Backprop from O/P layer $= \frac{\partial E}{\partial U_2} = \frac{\partial}{\partial U_2}\left\{e[-y\hat{y}]\right\}$

" " 1st hidden node $= \frac{\partial}{\partial U_2}\left\{e[-yZ_1]\right\}$

$$\frac{\partial E}{\partial u_2} = \frac{\partial}{\partial u_2} e^{-y\hat{y}} = -y e^{-y\hat{y}} \sigma'\left\{\sum_{i=1}^{2} w_i z_i\right\} w_1 x_2 \quad — ① \quad [.5]$$

$$\frac{\partial}{\partial u_2} e^{-y z_1} = -y e^{-y z_1} \quad w_1 x_2 \quad — ② \quad [.5]$$

$$[.5]$$

$$u_2^{new} \leftarrow u_2^{old} - \eta\left\{ ① + ② \right\}$$