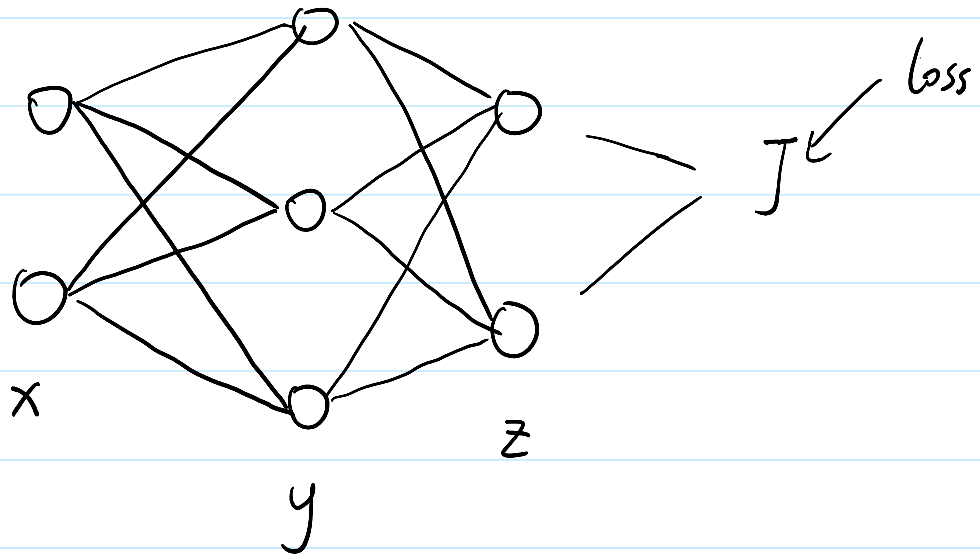
NN BP

Tuesday, October 15, 2019 9:59



$$\begin{cases} y_i = w_i^T x + b_i = \sum_j w_{ij} x_j + b_i \\ y_i = \mathrm{relu}(y_i) = \begin{cases} y_i, & y_i > 0 \\ 0, & \text{o.w.} \end{cases} \end{cases}$$

① FC layers : $Y = XW + B$

$X : N \times m$ $\qquad$ $x : 1 \times m$

$Y : N \times n$ $\qquad$ $y : 1 \times n$

$W : m \times n$

$B : n \times 1$ (broad cast to $N \times n$) $\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$

Gradiant calculation : (suppose $N=1$)

Known : $\frac{\partial J}{\partial Y} \in R^{N \times n}$

① $\frac{\partial J}{\quad} = \sum \frac{\partial J}{\quad} \cdot \frac{\partial y_j}{\quad} = \frac{\partial J}{\quad} \cdot 1 = \frac{\partial J}{\quad}$

① $\frac{\partial J}{\partial b_i} = \sum_j \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial b_i} = \frac{\partial J}{\partial y_i} \cdot 1 = \frac{\partial J}{\partial y_i}$

$$\Rightarrow \frac{\partial J}{\partial B} = \begin{bmatrix} \frac{\partial J}{\partial b_1} \\ \vdots \\ \frac{\partial J}{\partial b_n} \end{bmatrix} = \left(\frac{\partial J}{\partial y}\right)^T$$

Expand to batch_size = N:

$$\frac{\partial J}{\partial B} = \sum_{i=1}^{N} \frac{1}{N} \left(\frac{\partial J}{\partial Y(i)}\right)^T$$

② $\frac{\partial J}{\partial w_i} = \sum_j \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial w_i} = \frac{\partial J}{\partial y_j} \cdot x$

$$\frac{\partial J}{\partial w} = \left[\frac{\partial J}{\partial w_1}, \cdots, \frac{\partial J}{\partial w_n}\right]$$

$$= \left[\frac{\partial J}{\partial y_1} \cdot x, \cdots, \frac{\partial J}{\partial y_n} \cdot x\right]$$

$$= \left(\begin{bmatrix} \frac{\partial J}{\partial y_1} \\ \vdots \\ \frac{\partial J}{\partial y_n} \end{bmatrix}_{n\times 1} \cdot [x_1, \cdots x_m]_{1\times m}\right)^T$$

Expand to batch_size = N:

$$\frac{1}{N} \sum_N = \left(R^{N\times n\times 1} \times R^{N\times 1\times m}\right)^T$$

③ $$\frac{\partial J}{\partial x} = \sum_i \frac{\partial J}{\partial y_i} \cdot \frac{\partial y_i}{\partial x} = \sum_i \frac{\partial J}{\partial y_i} \cdot w_i$$

$$= \begin{bmatrix} w_1, & \cdots & w_n \end{bmatrix}_{m\times n} \times \begin{bmatrix} \frac{\partial J}{\partial y_1} \\ \vdots \\ \frac{\partial J}{\partial y_n} \end{bmatrix}_{n\times 1}$$

$$= W \times \frac{\partial J}{\partial y}$$

Expand to batch_size = N :

$$\frac{\partial J}{\partial X} = \underbrace{W}_{\text{broadcast to } N\times m\times n} \times \underbrace{\frac{\partial J}{\partial Y}}_{N\times n\times 1}$$