



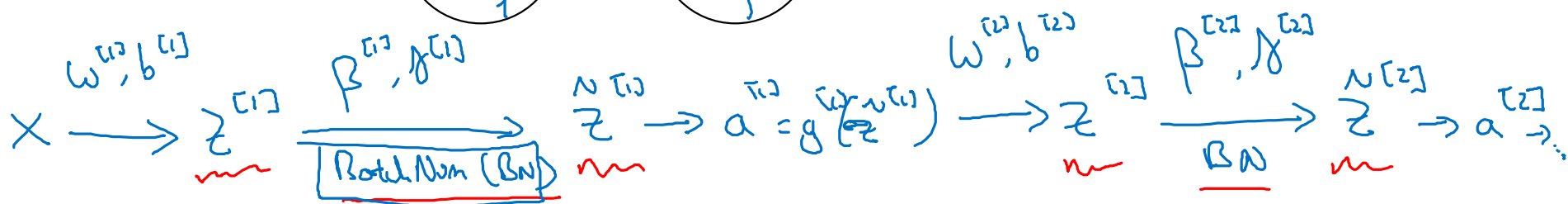
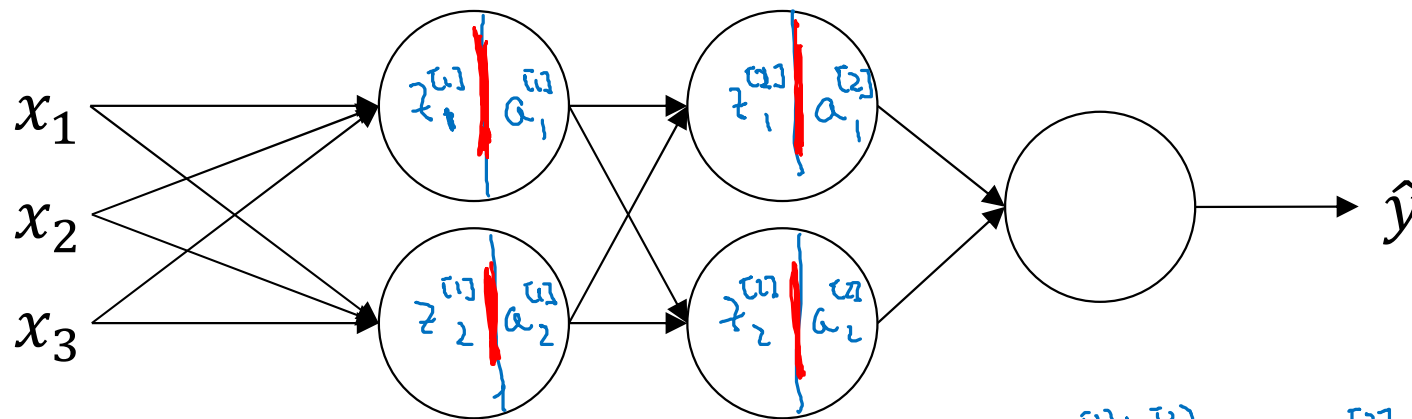
deeplearning.ai

# Batch Normalization

---

Fitting Batch Norm  
into a neural network

# Adding Batch Norm to a network



Parameters:  $\left\{ W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}, \dots, W^{(L)}, b^{(L)} \right\}$   
 $\rightarrow \underline{\beta}^{(1)}, \underline{\gamma}^{(1)}, \underline{\beta}^{(2)}, \underline{\gamma}^{(2)}, \dots, \underline{\beta}^{(L)}, \underline{\gamma}^{(L)}$   
 $\rightarrow \underline{\beta}$

$$d\beta^{(2)} \quad \beta = \beta - \alpha d\beta^{(2)}$$

tf.nn.batch-normalization ←

# Working with mini-batches

$$\tilde{X}^{\{1\}} \xrightarrow{W^{\{1\}}, b^{\{1\}}} z^{\{1\}} \xrightarrow[\text{BN}]{\beta^{\{1\}}, \gamma^{\{1\}}} \tilde{z}^{\{1\}} \rightarrow g^{\{1\}}(\tilde{z}^{\{1\}}) = a^{\{1\}} \xrightarrow{W^{\{2\}}, b^{\{2\}}} z^{\{2\}} \rightarrow \dots$$

$$\boxed{X^{\{2\}}} \rightarrow \underline{z^{\{2\}}} \xrightarrow[\text{BN}]{\beta^{\{2\}}, \gamma^{\{2\}}} \tilde{z}^{\{2\}} \rightarrow \dots$$

$$X^{\{2\}} \rightarrow \dots$$

Parameters:  $W^{\{2\}}, \cancel{b^{\{2\}}}, \beta^{\{2\}}, \gamma^{\{2\}}$

$\uparrow$   
 $(n^{\{2\}}, 1)$

$\uparrow$   
 $(n^{\{2\}}, 1)$

$\uparrow$   
 $(n^{\{2\}}, 1)$

$$\tilde{z}^{\{2\}}_{(n^{\{2\}}, 1)}$$

$$\rightarrow \underline{z^{\{2\}}} = W^{\{2\}} a^{\{1\}} + \cancel{b^{\{2\}}}$$

$\uparrow$

$$z^{\{2\}} = W^{\{2\}} a^{\{1\}}$$

$$z^{\{2\}}_{\text{norm}}$$

$$\rightarrow \tilde{z}^{\{2\}} = \gamma^{\{2\}} z^{\{2\}}_{\text{norm}} + \boxed{\beta^{\{2\}}}$$

Andrew Ng

# Implementing gradient descent

for  $t = 1 \dots \text{num Mini Batches}$   
Compute forward pass on  $X^{\{t\}}$ .

In each hidden layer, use BN to replace  $\underline{z}^{\{t\}}$  with  $\hat{\underline{z}}^{\{t\}}$ .

Use backprop to compute  $\underline{dw}^{\{t\}}$ ,  ~~$\underline{db}^{\{t\}}$~~ ,  $\underline{dp}^{\{t\}}$ ,  $\underline{df}^{\{t\}}$

Update params  $\left. \begin{array}{l} w^{\{t\}} := w^{\{t\}} - \alpha \underline{dw}^{\{t\}} \\ \beta^{\{t\}} := \beta^{\{t\}} - \alpha \underline{dp}^{\{t\}} \\ \gamma^{\{t\}} := \dots \end{array} \right\} \leftarrow$

Works w/ momentum, RMSprop, Adam.