

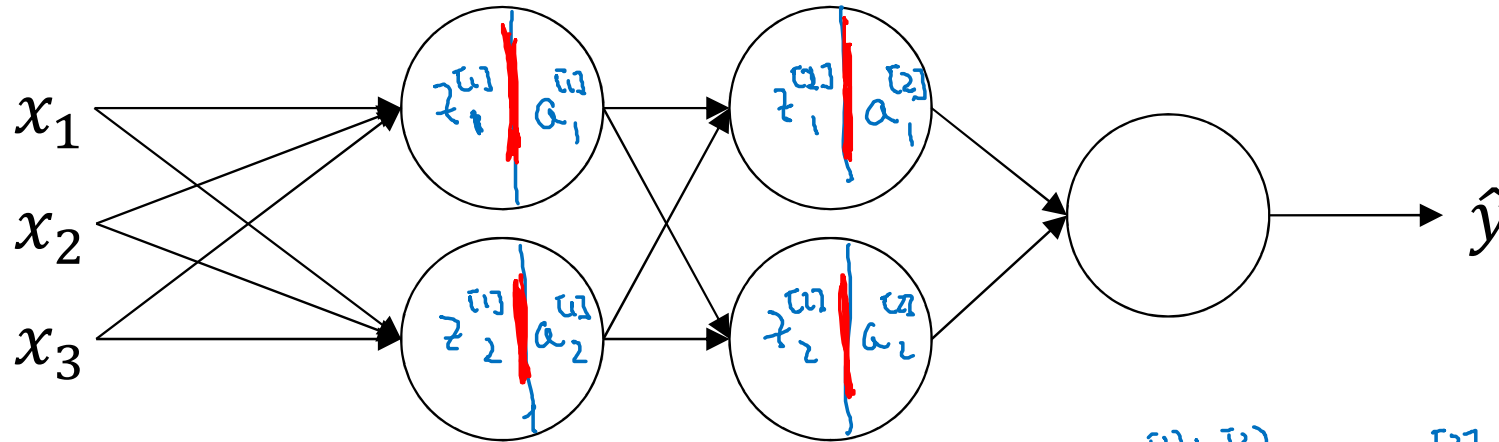


deeplearning.ai

Batch Normalization

Fitting Batch Norm
into a neural network

Adding Batch Norm to a network



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

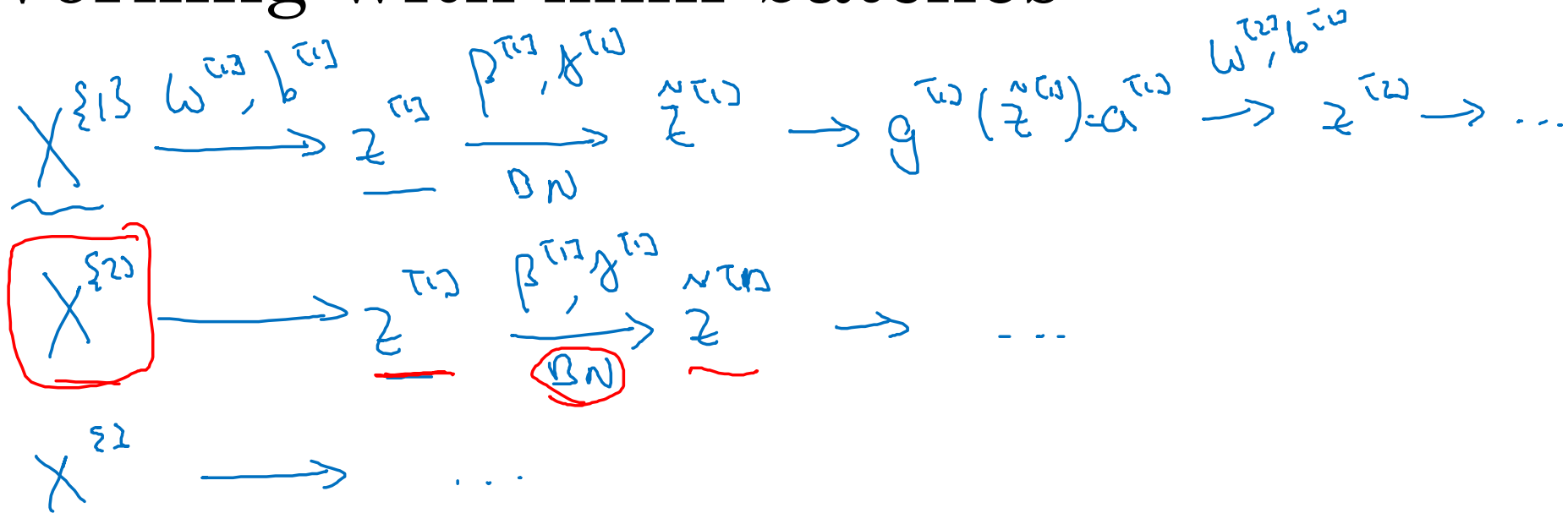
(Note: $\tilde{z}^{[1]} \rightarrow a^{[1]} = g(z^{[1]})$ is implied in the original image)

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

$$d\beta^{[l]} \quad \beta = \beta^{[0]} - \alpha d\beta^{[0]}$$

tf.nn.batch-normalization ←

Working with mini-batches



Parameters: $W^{\tau_1}, \cancel{b^{\tau_1}}, \beta^{\tau_1}, \gamma^{\tau_1}$.

$\begin{matrix} \uparrow & \uparrow & \uparrow \\ (n^{\tau_1}, 1) & (n^{\tau_1}, 1) & (n^{\tau_1}, 1) \end{matrix}$

$\begin{matrix} \uparrow \\ z^{\tau_1} \\ (n^{\tau_1}, 1) \end{matrix}$

$\rightarrow \underline{z^{\tau_1}} = W^{\tau_1} a^{\tau_1-1} + \cancel{b^{\tau_1}}$

\uparrow

$z^{\tau_1} = W^{\tau_1} a^{\tau_1-1}$

$z_{norm}^{\tau_1}$

$\rightarrow \tilde{z}^{\tau_1} = \gamma^{\tau_1} z_{norm}^{\tau_1} + \beta^{\tau_1}$

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 $\underline{\hat{z}}^{\{t\}}$.

Use backprop to compute $\underline{dw}^{\{t\}}$, ~~$\underline{db}^{\{t\}}$~~ , $\underline{d\beta}^{\{t\}}$, $\underline{d\gamma}^{\{t\}}$

Update params $\left. \begin{array}{l} w^{\{t\}} := w^{\{t-1\}} - \alpha \underline{dw}^{\{t\}} \\ \beta^{\{t\}} := \beta^{\{t-1\}} - \alpha \underline{d\beta}^{\{t\}} \\ \gamma^{\{t\}} := \dots \end{array} \right\} \leftarrow$

Works w/ momentum, RMSprop, Adam.