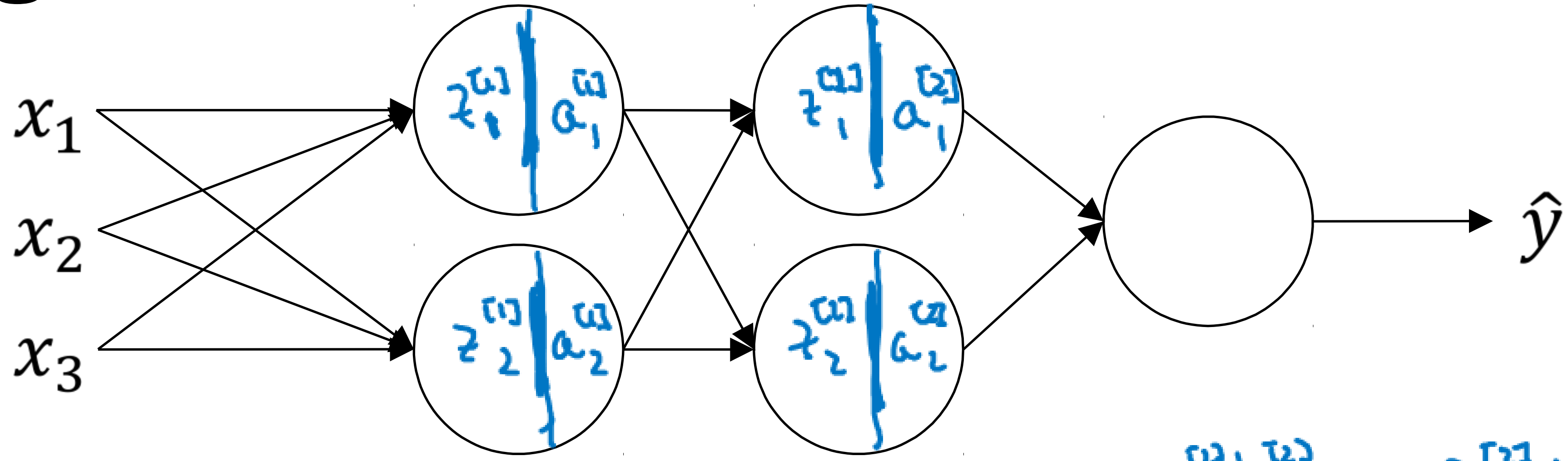
deeplearning.ai

Batch Normalization

Fitting Batch Norm into a neural network

Adding Batch Norm to a network


x_1
x_2
x_3
$z_1^{[1]}$ | $a_1^{[1]}$
$z_2^{[1]}$ | $a_2^{[1]}$
$z_1^{[2]}$ | $a_1^{[2]}$
$z_2^{[2]}$ | $a_2^{[2]}$
$\hat{y}$

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

Parametes: $\left.\begin{array}{l} W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, \ldots, W^{[L]}, b^{[L]}, \\ \rightarrow \beta^{[1]}, \gamma^{[1]}, \beta^{[2]}, \gamma^{[2]}, \ldots, \beta^{[L]}, \gamma^{[L]} \end{array}\right\}$

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

$\rightarrow \beta$

tf.nn.batch-normalization ←

Working with mini-batches

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

$$\boxed{X^{\{2\}}} \longrightarrow z^{[1]} \xrightarrow[\boxed{BN}]{\beta^{[1]},\, \gamma^{[1]}} \tilde{z}^{[1]} \rightarrow \dots$$

$$X^{\{3\}} \longrightarrow \dots$$

Parameters: $W^{[l]}, \cancel{b^{[l]}}, \beta^{[l]}, \gamma^{[l]}.$

$$\begin{array}{ccc} \cancel{b^{[l]}} & \beta^{[l]} & \gamma^{[l]} \\ | & | & | \\ (n^{[l]}, 1) & (n^{[l]}, 1) & (n^{[l]}, 1) \\ \uparrow & & \end{array}$$

$z^{[l]}$

$(n^{[l]}, 1)$

$$\rightarrow z^{[l]} = W^{[l]} a^{[l-1]} + \boxed{\cancel{b^{[l]}}}$$

$$\uparrow$$

$$z^{[l]} = W^{[l]} a^{[l-1]}$$

$$z^{[l]}_{norm}$$

$$\rightarrow \tilde{z}^{[l]} = \gamma^{[l]} z^{[l]}_{norm} + \boxed{\beta^{[l]}} \leftarrow$$

Implementing gradient descent

```
for $t = 1 \ldots \text{numMiniBatches}$
    Compute forward prop on $X^{\{t\}}$.
        In each hidden layer, use BN to replace $\underline{z^{[l]}}$ with $\underline{\tilde{z}^{[l]}}$.
    Use backprop to compute $\underline{dW^{[l]}}$, $\cancel{db^{[l]}}$, $\underline{d\beta^{[l]}}$, $\underline{d\gamma^{[l]}}$
    Update params $\left.\begin{aligned} W^{[l]} &:= W^{[l]} - \alpha\, dW^{[l]} \\ \beta^{[l]} &:= \beta^{[l]} - \alpha\, d\beta^{[l]} \\ \gamma^{[l]} &:= \ldots \end{aligned}\right\} \leftarrow$
```

Works w/ momentum, RMSprop, Adam.