Batch Normalization

Fitting Batch Norm into a neural network

Adding Batch Norm to a network

$$X \xrightarrow{W^{[1]}, b^{[1]}} Z^{[1]} \xrightarrow[\boxed{\text{Batch Norm (BN)}}]{\beta^{[1]}, \gamma^{[1]}} \tilde{Z}^{[1]} \rightarrow a^{[1]} = g^{[1]}(\tilde{Z}^{[1]}) \xrightarrow{W^{[2]}, b^{[2]}} Z^{[2]} \xrightarrow[BN]{\beta^{[2]}, \gamma^{[2]}} \tilde{Z}^{[2]} \rightarrow a^{[2]} \rightarrow \cdots$$
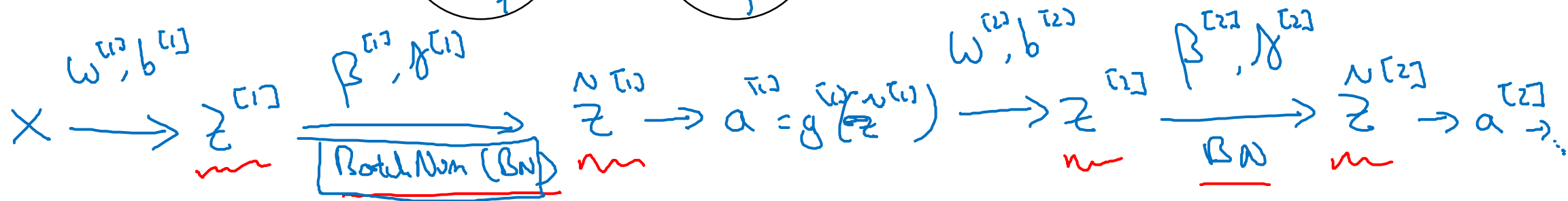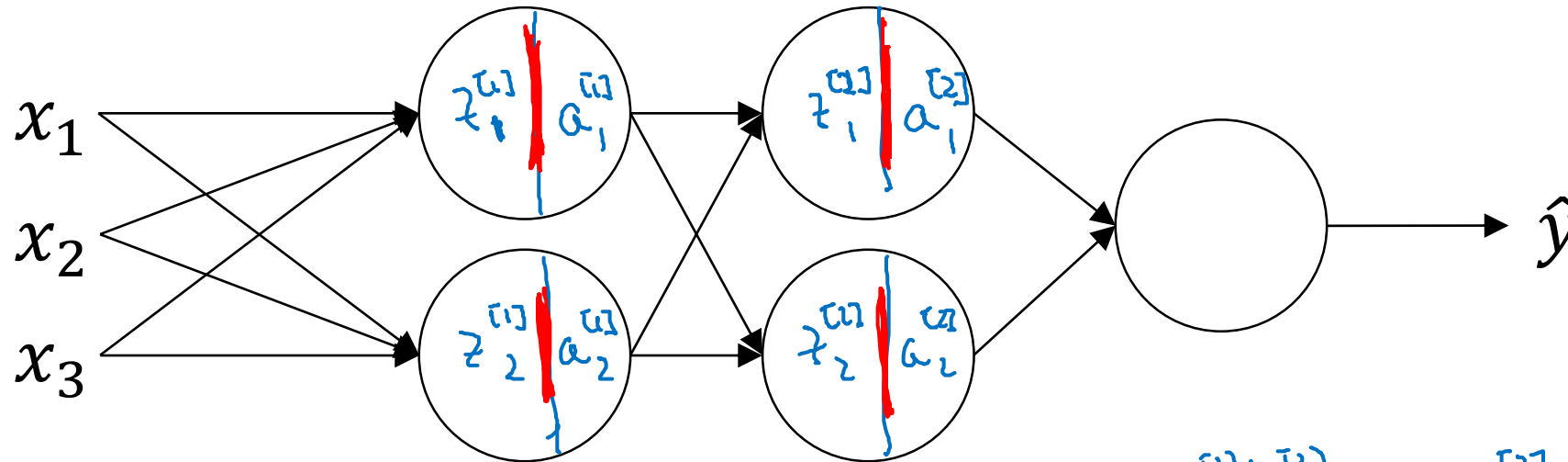
$$\text{Parameters: } \left. \begin{array}{l} W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, \ldots, W^{[L]}, b^{[L]}, \\ \to \beta^{[1]}, \gamma^{[1]}, \beta^{[2]}, \gamma^{[2]}, \ldots, \beta^{[L]}, \gamma^{[L]} \end{array} \right\}$$

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

tf.nn.batch-normalization

$$\to \beta$$

Andrew Ng

Working with mini-batches

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

$$\boxed{X^{\{2\}}} \to Z^{[1]} \xrightarrow[\boxed{BN}]{\beta^{[1]}, \gamma^{[1]}} \tilde{Z}^{[1]} \to \dots$$

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

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

$$\qquad\qquad (n^{[\ell]}, 1) \quad (n^{[\ell]}, 1) \quad (n^{[\ell]}, 1)$$

$$Z^{[\ell]}$$
$$(n^{[\ell]}, 1)$$

$$\to Z^{[\ell]} = W^{[\ell]} a^{[\ell-1]} + \boxed{\cancel{b^{[\ell]}}}$$

$$Z^{[\ell]} = W^{[\ell]} a^{[\ell-1]}$$

$$Z^{[\ell]}_{norm}$$

$$\to \tilde{Z}^{[\ell]} = \gamma^{[\ell]} Z^{[\ell]}_{norm} + \boxed{\beta^{[\ell]}}$$

Andrew Ng

Implementing gradient descent

for $t = 1 \dots$ num MiniBatches

   Compute forward prop on $X^{\{t\}}$.

     In each hidden layer, use BN to replace $Z^{[l]}$ with $\tilde{Z}^{[l]}$.

   Use backprop to compute $dW^{[l]}, \cancel{db^{[l]}}, d\beta^{[l]}, d\gamma^{[l]}$

   Update params
$$\left. \begin{array}{l} W^{[l]} := W^{[l]} - \alpha \, dW^{[l]} \\ \beta^{[l]} := \beta^{[l]} - \alpha \, d\beta^{[l]} \\ \gamma^{[l]} := \dots \end{array} \right\} \leftarrow$$

Works w/ momentum, RMSprop, Adam.

Andrew Ng