Batch Normalization
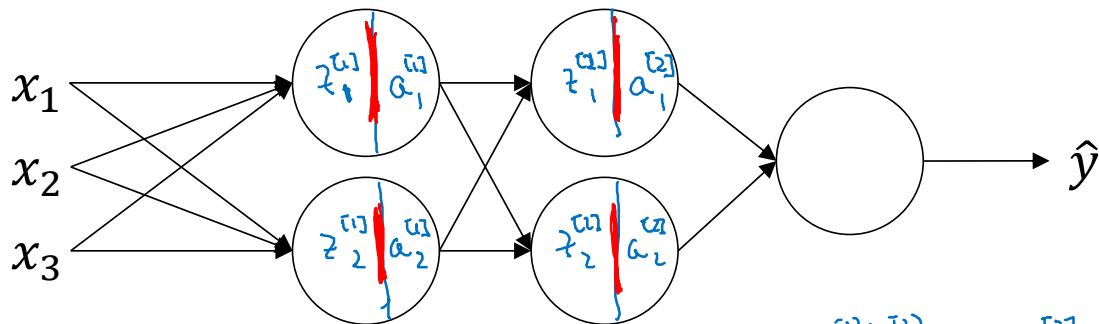
deeplearning.ai

Fitting Batch Norm into a neural network

Adding Batch Norm to a network

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

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

$\rightarrow \beta$

$d\beta^{[l]}$

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

tf.nn.batch-normalization $\leftarrow$

Working with mini-batches

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

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

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

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

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

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

no point

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

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

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

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

Andrew Ng

Implementing gradient descent

for t = 1 .... num MiniBatches

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

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

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

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

can Works w/ momentum, RMSprop, Adam.