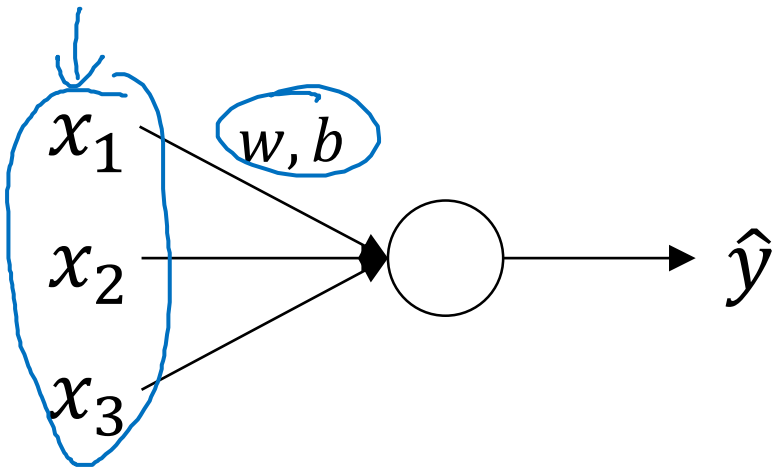
deeplearning.ai

Batch Normalization

Normalizing activations in a network

Normalizing inputs to speed up learning



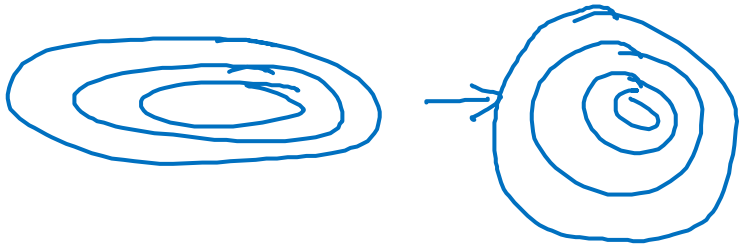
$$\mu = \frac{1}{m} \sum_i x^{(i)}$$

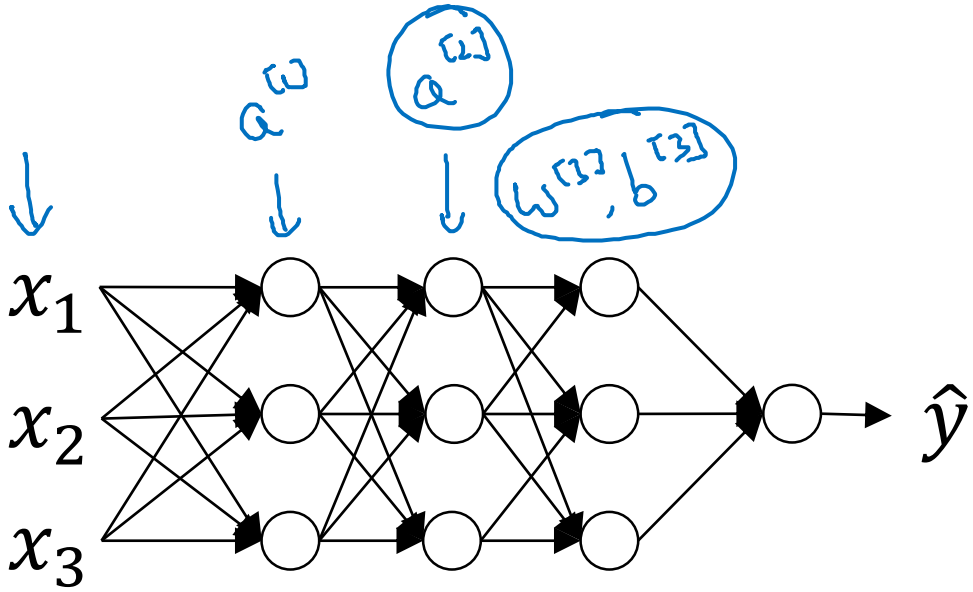
$$X = X - \mu$$

$$\sigma^2 = \frac{1}{m} \sum_i x^{(i)2}$$

$$X = X / \sigma^2$$

← element-wise





Can we normalize $a^{[2]}$ so as to train $w^{[3]}, b^{[3]}$ faster

Normalize $z^{[2]}$

Implementing Batch Norm

Given some intermediate values in NN $z^{(1)}, \ldots, z^{(m)}$ — $z^{[l](i)}$

$\mu = \frac{1}{m} \sum_i z^{(i)}$

$\sigma^2 = \frac{1}{m} \sum_i (z_i - \mu)^2$

$z^{(i)}_{norm} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$ ←

$\tilde{z}^{(i)} = \gamma z^{(i)}_{norm} + \beta$ ← learnable parameters of model.

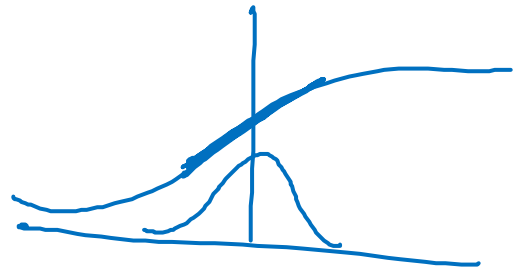
If $\gamma = \sqrt{\sigma^2 + \varepsilon}$ ←

$\beta = \mu$ ←

then $\tilde{z}^{(i)} = z^{(i)}$

X ←

$z^{(i)}$ ←



Use $\tilde{z}^{[l](i)}$ instd of $z^{[l](i)}$.