

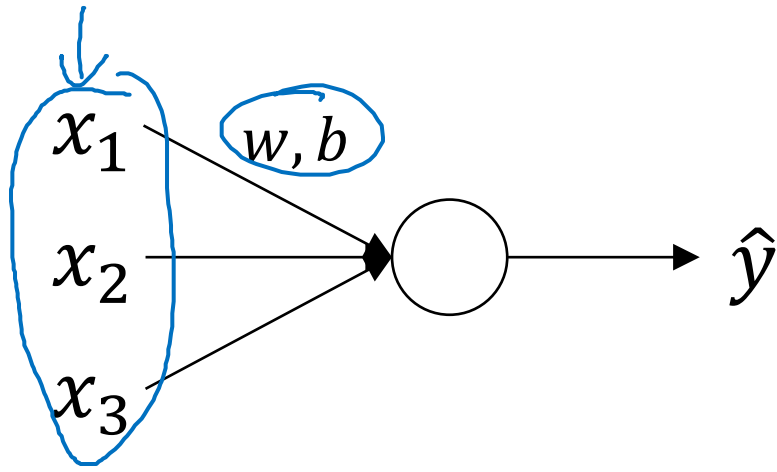


deeplearning.ai

Batch Normalization

Normalizing activations
in a network

Normalizing inputs to speed up learning

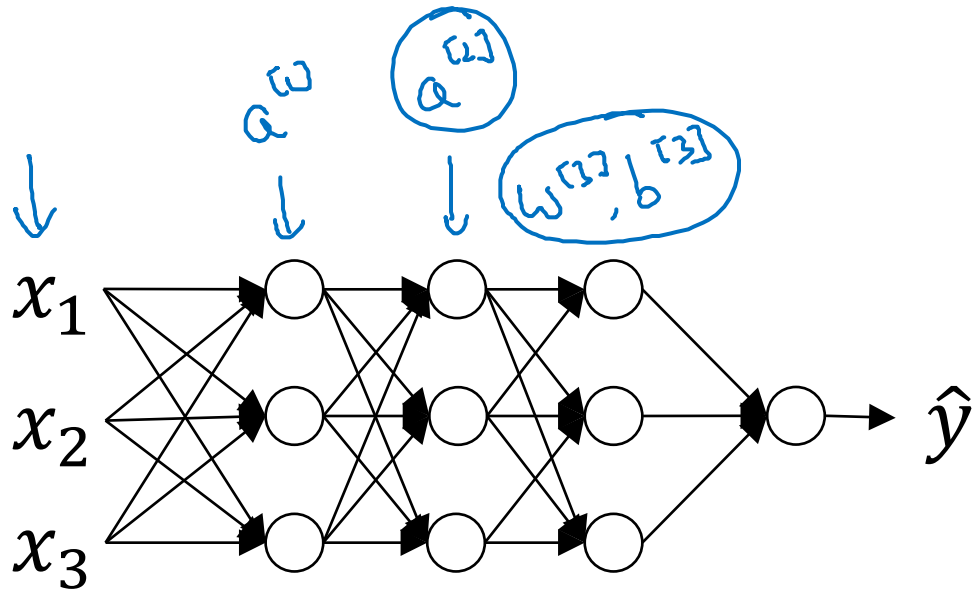
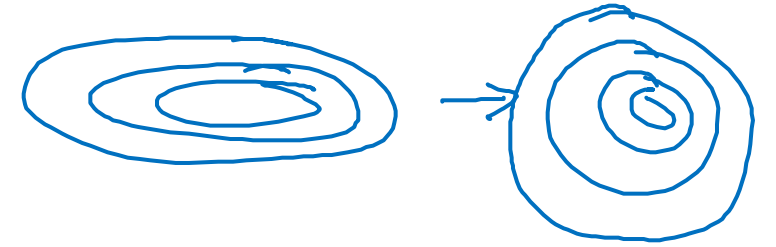


$$\mu = \frac{1}{n} \sum_i x^{(i)}$$

$$X = X - \mu$$

$$\sigma^2 = \frac{1}{n} \sum_i x^{(i)2} \quad \leftarrow \text{element-wise}$$

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



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

Normalize $\frac{z^{[2]}}{\uparrow}$

Implementing Batch Norm

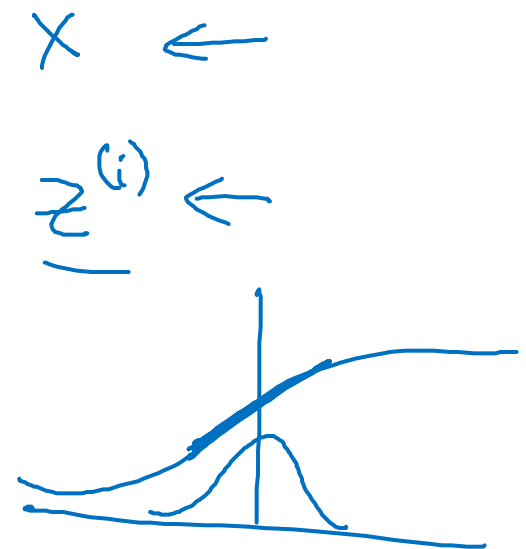
Given some intermediate values in NN

$$\begin{matrix} \downarrow & \downarrow \\ z^{(1)} & \dots, z^{(m)} \\ \underbrace{\hspace{10em}} & \\ z^{[l]}(i) \end{matrix}$$

$$\left[\begin{aligned} \mu &= \frac{1}{m} \sum_i z^{(i)} \\ \sigma^2 &= \frac{1}{m} \sum_i (z_i - \mu)^2 \\ z_{\text{norm}}^{(i)} &= \frac{z^{(i)} - \boxed{\mu}}{\boxed{\sqrt{\sigma^2 + \epsilon}}} \\ \hat{z}^{(i)} &= \underbrace{\gamma}_{\downarrow} z_{\text{norm}}^{(i)} + \underbrace{\beta}_{\downarrow} \end{aligned} \right.$$

$$\left[\begin{aligned} \text{If } \gamma &= \boxed{\sqrt{\sigma^2 + \epsilon}} \leftarrow \\ \beta &= \boxed{\mu} \leftarrow \\ \text{then } \hat{z}^{(i)} &= z^{(i)} \end{aligned} \right.$$

learnable parameters of model.



Use $\hat{z}^{[l]}(i)$ instead of $z^{[l]}(i)$.