



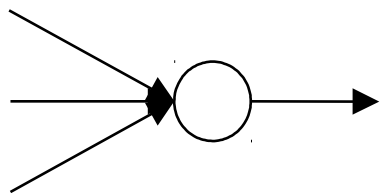
**deeplearning.ai**

# Batch Normalization

---

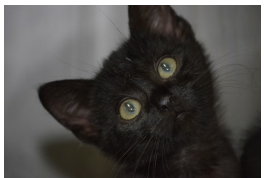
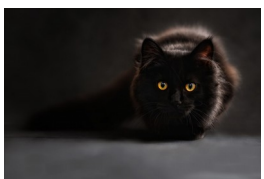
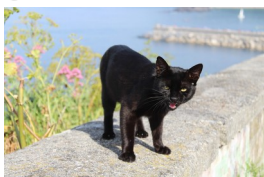
Why does  
Batch Norm  
work?

# Learning on shifting input distribution



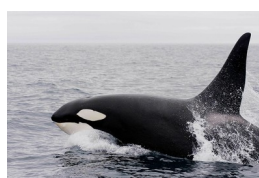
Cat

$y = 1$



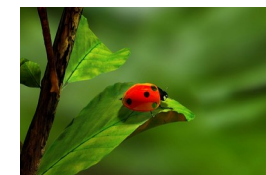
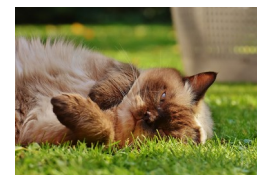
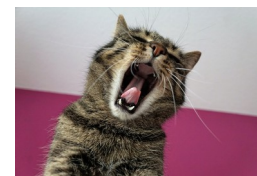
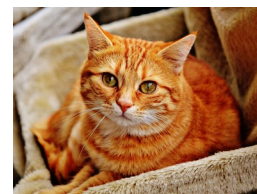
Non-Cat

$y = 0$



$y = 1$

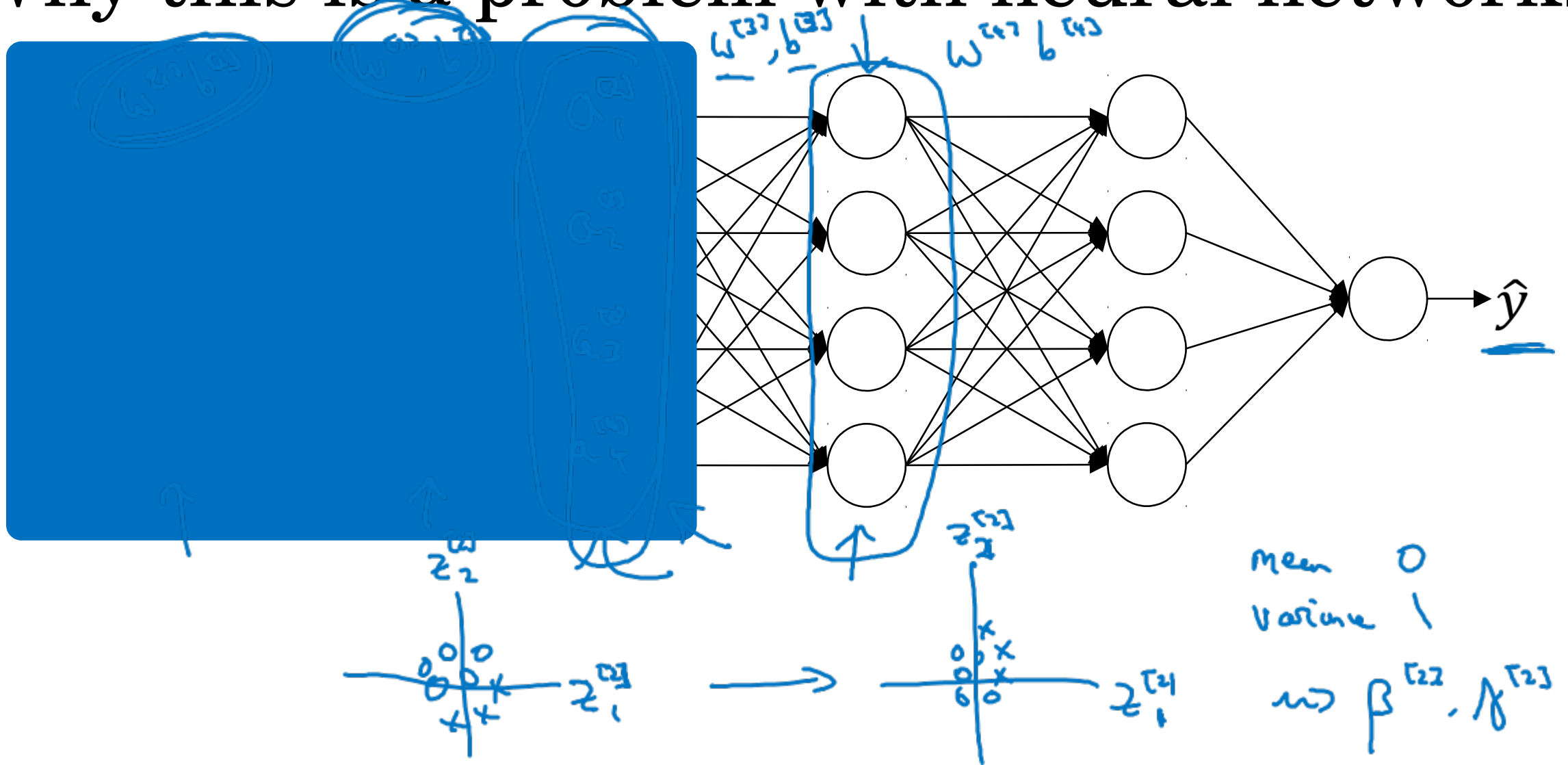
$y = 0$



"Covariate shift"

$\underline{x} \rightarrow y$

# Why this is a problem with neural networks?



# Batch Norm as regularization

- Each mini-batch is scaled by the mean/variance computed on just that mini-batch.  $\mu, \sigma^2$
- This adds some noise to the values within that mini-batch. So similar to dropout, it adds some noise to each layer's activations.  $\mu, \sigma^2$
- This has a slight regularization effect.

mini-batch : 64  $\longrightarrow$  512