



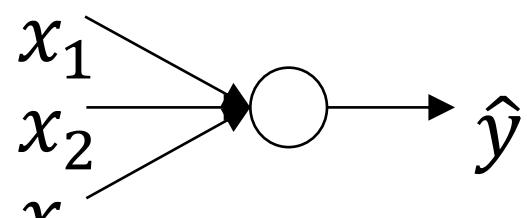
deeplearning.ai

# Batch Normalization

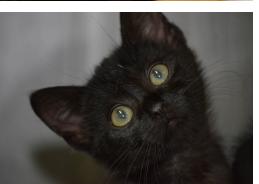
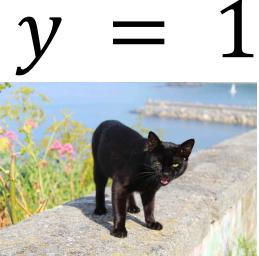
---

Why does  
Batch Norm work?

# Learning on shifting input distribution



Cat

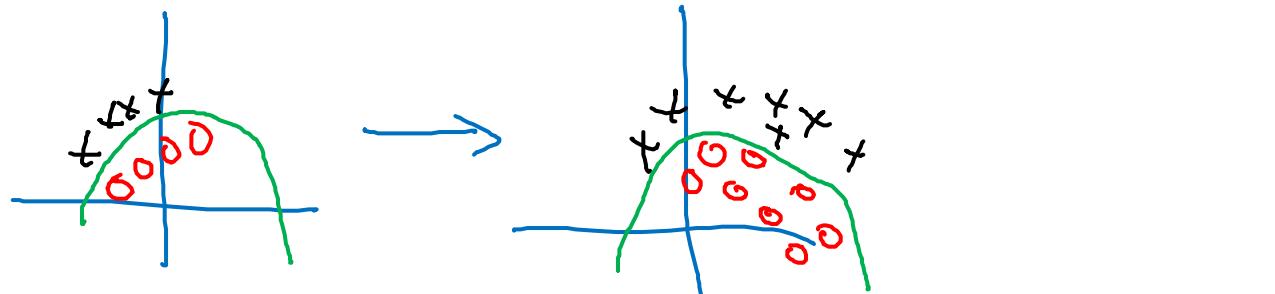


Non-Cat

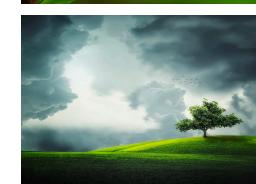


$y = 1$

$y = 0$



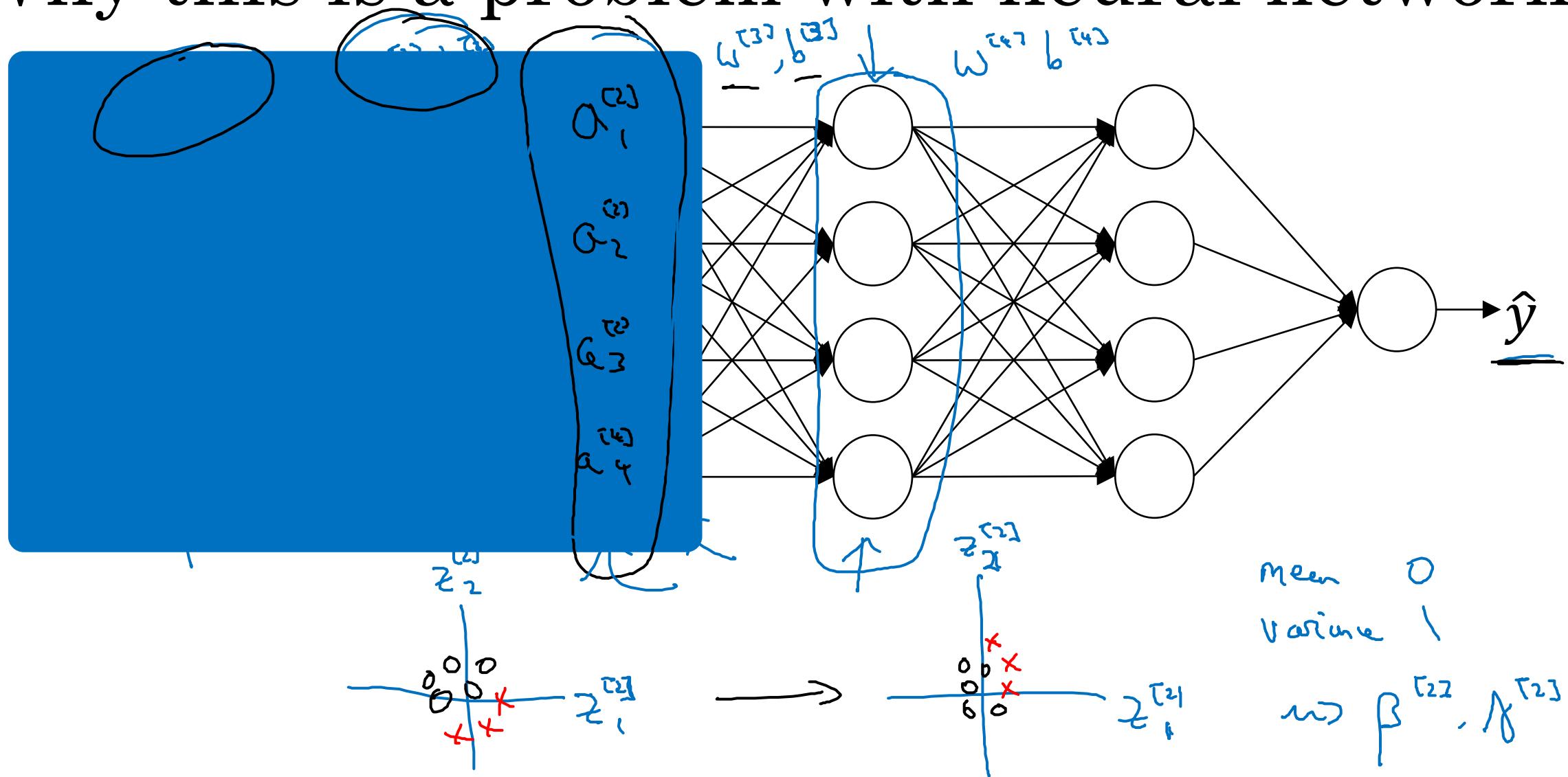
$y = 1$        $y = 0$



"Covariate shift"

$X \rightarrow Y$

# Why this is a problem with neural networks?



# Batch Norm as regularization

X

- Each mini-batch is scaled by the mean/variance computed on just that mini-batch.  
 $\xrightarrow{\hat{z}^{[l]}}$   $\mu, \sigma^2$   $\{z^{[l]}\}$   
 $\hat{z}^{[l]}$   $\underline{64}, \underline{128}$
- This adds some noise to the values  $z^{[l]}$  within that minibatch. So similar to dropout, it adds some noise to each hidden layer's activations.  
 $\mu, \sigma^2$
- This has a slight regularization effect.

mini-batch : 64  $\longrightarrow$  512