

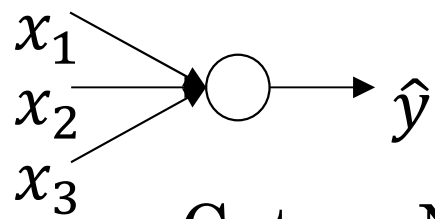


deeplearning.ai

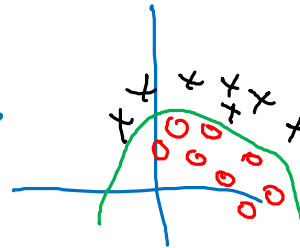
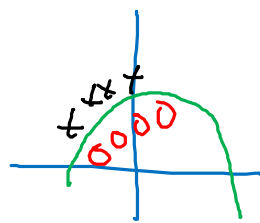
Batch Normalization

Why does
Batch Norm work?

Learning on shifting input distribution



Cat $y = 1$ ✓ Non-Cat $y = 0$



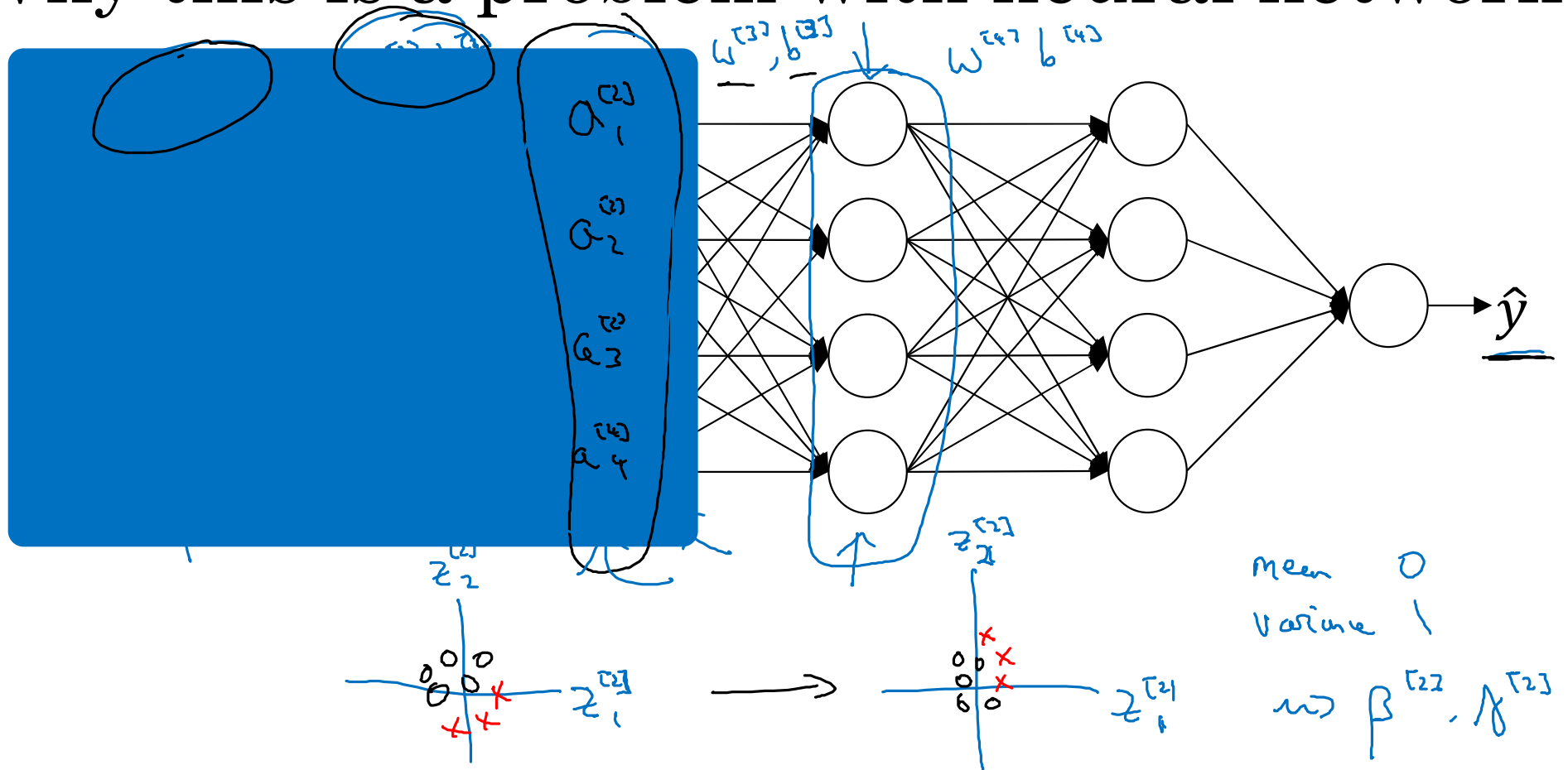
$y = 1$ ✓ $y = 0$



"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.
- 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.
- This has a slight regularization effect.

X

$X^{\{t\}}$
 $z^{[l]}$

$\tilde{z}^{[l]}$

64, 128

μ, σ^2

mini-batch : 64 \longrightarrow 512