deeplearning.ai

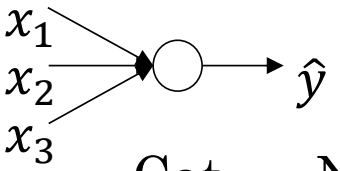
Batch Normalization

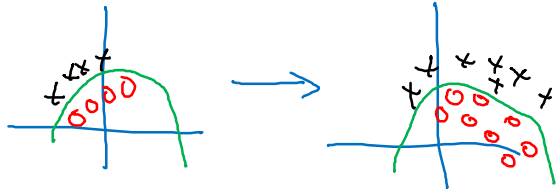
Why does Batch Norm work?

on each mini-batch

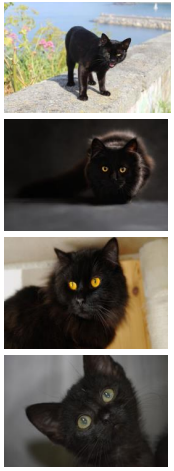
first reason: BN makes weights, later or deeper than your network. (speed up)

Learning on shifting input distribution





Cat $y = 1$

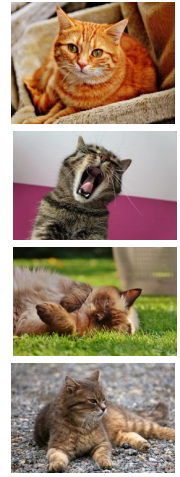


Non-Cat $y = 0$

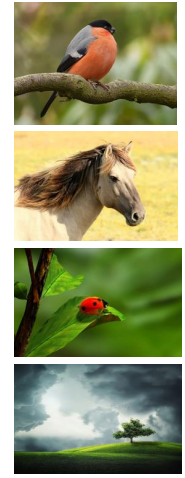
协变量移位
"Covariate shift"

$X \rightarrow y$

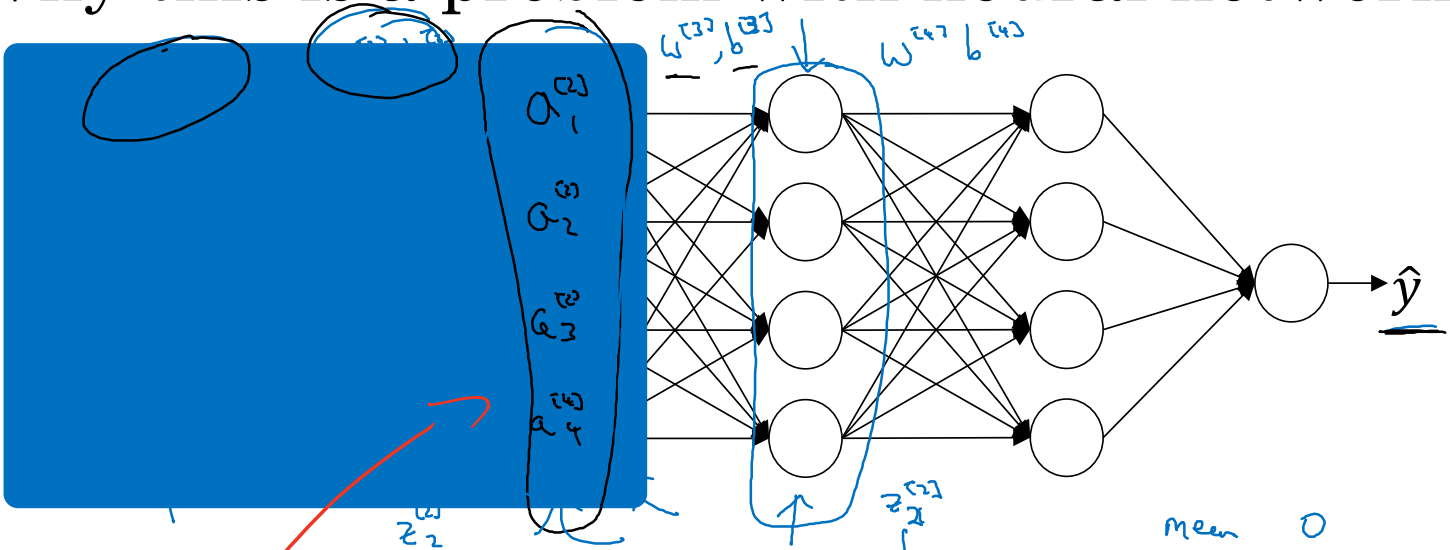
$y = 1$



$y = 0$



Why this is a problem with neural networks?



from the perspective of the third hidden layer, these hidden unit values are changing all the time → covariate shift.

forces the later layers to adapt to as early as layer changes is reduced

mean 0
variance 1

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.