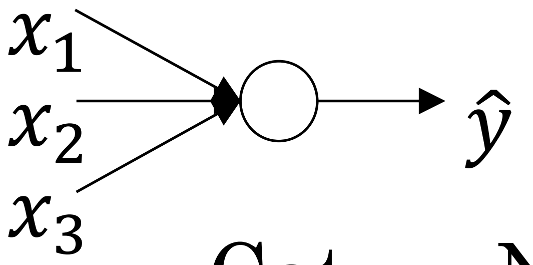
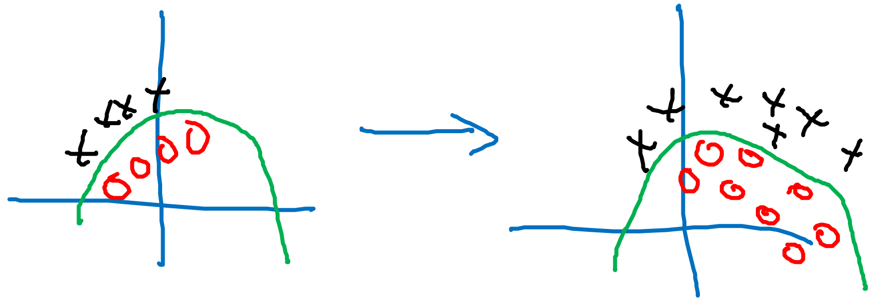
deeplearning.ai

Batch Normalization

Why does Batch Norm work?

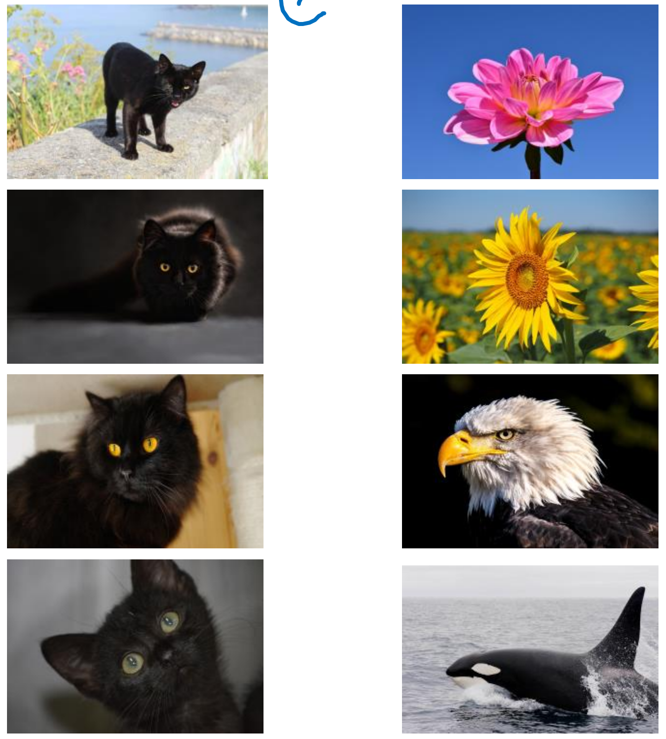
Learning on shifting input distribution





A second reason why batch norm works, is it makes weights, later or deeper than your network.

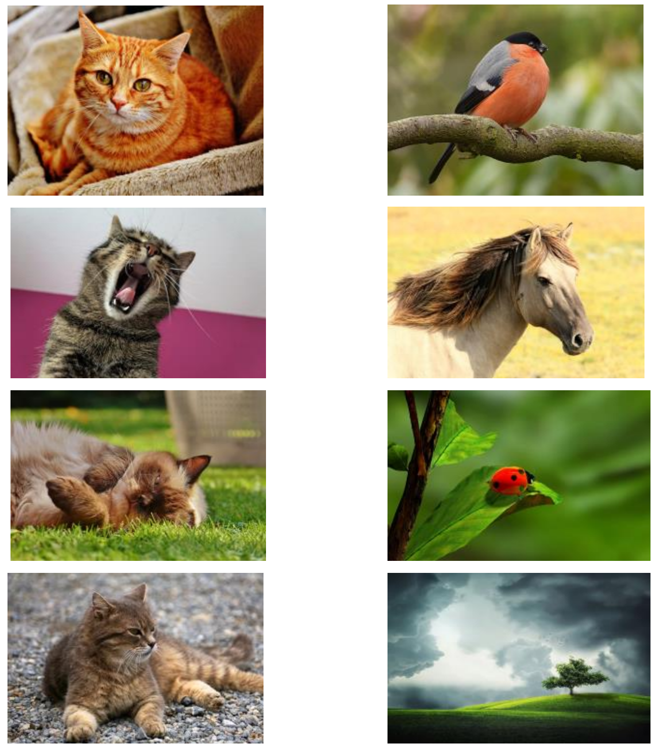
Cat — $y = 1$ ✓ Non-Cat — $y = 0$



"Covariate shift"

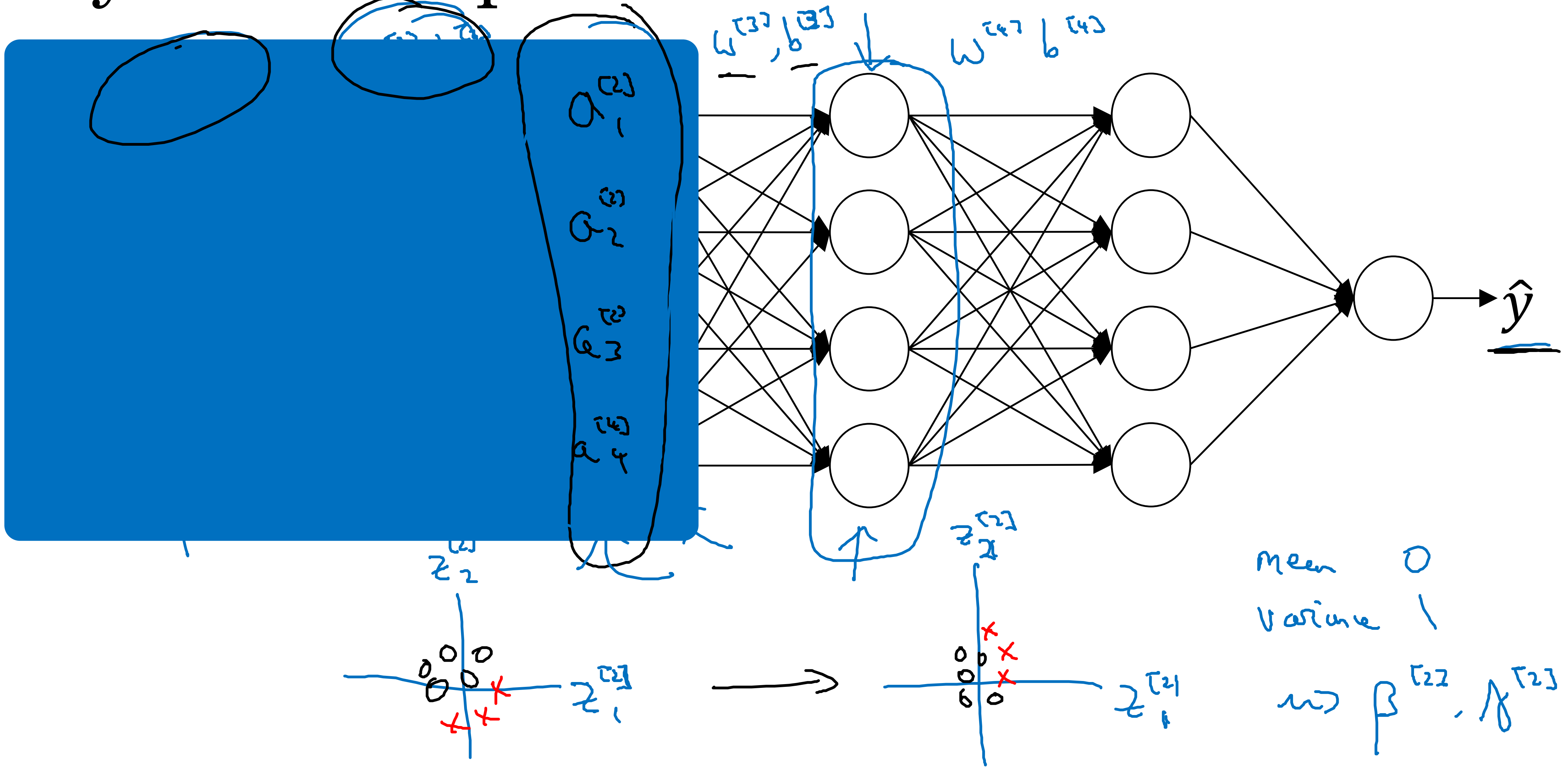
$\underset{\sim}{X} \longrightarrow y$

$y = 1$ ✓ $y = 0$



Andrew Ng

Why this is a problem with neural networks?



So from the perspective of the third hidden layer, these hidden unit values are changing all the time, and so it's suffering from the problem of covariate shift.
So the Batch Norm does, is it reduces the amount that the distribution of these hidden unit values shifts around.

But the take away is that batch norm means that, especially from the perspective of one of the later layers of the neural network, the earlier layers don't get to shift around as much, because they're constrained to have the same mean and variance.
And so this makes the jobs of learning on the later layers easier.

Andrew Ng

Batch Norm as regularization

X

$X^{\{t\}}$

- Each mini-batch is scaled by the mean/variance computed on just that mini-batch.

$\tilde{z}^{[l]}$ $64, 128$ $z^{[l]}$

- 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.

μ, σ^2 are noisy

- This has a slight regularization effect.

mini-batch : $64 \longrightarrow 512$ one strange property of dropout, which is that by using a bigger mini-batch size, you reduce the regularization effect.

Because by adding noise to the hidden units, it's forcing the downstream hidden units not to rely on too much on any one hidden unit.
And so similar to dropout, it adds noise to the hidden layers and therefore has a very slight regularization effect.
Because the noise added is quite small, this is not a huge regularization effect and you might choose to use batch norm together with dropout if you want the more powerful regularization effect of dropout.

Andrew Ng