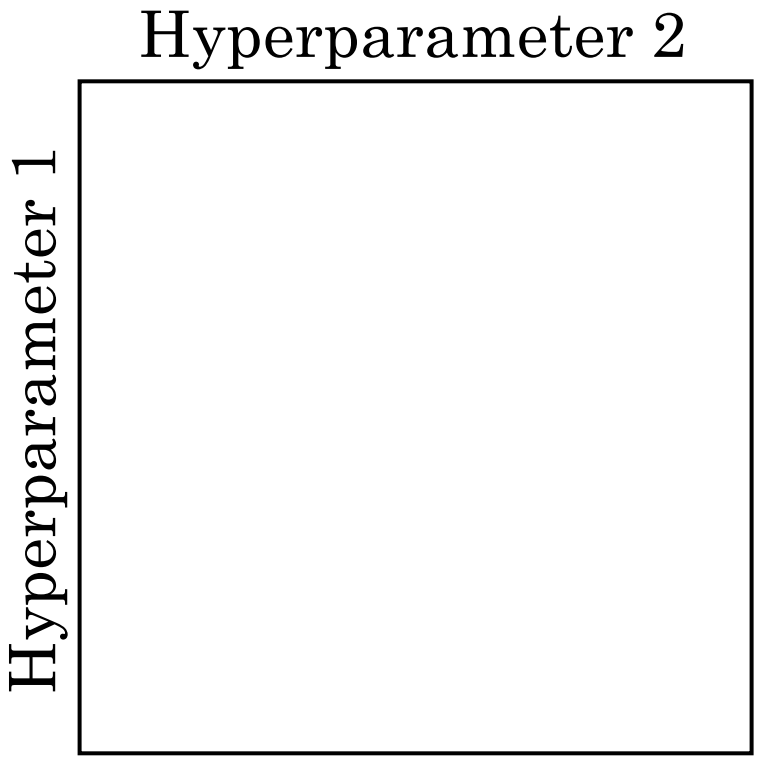
deeplearning.ai

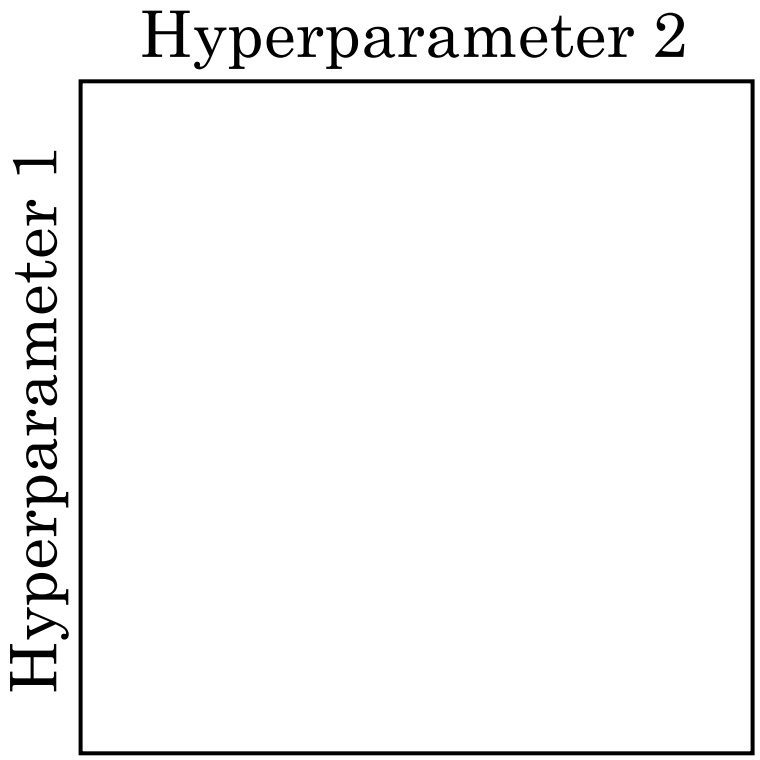
Hyperparameter tuning

Tuning process

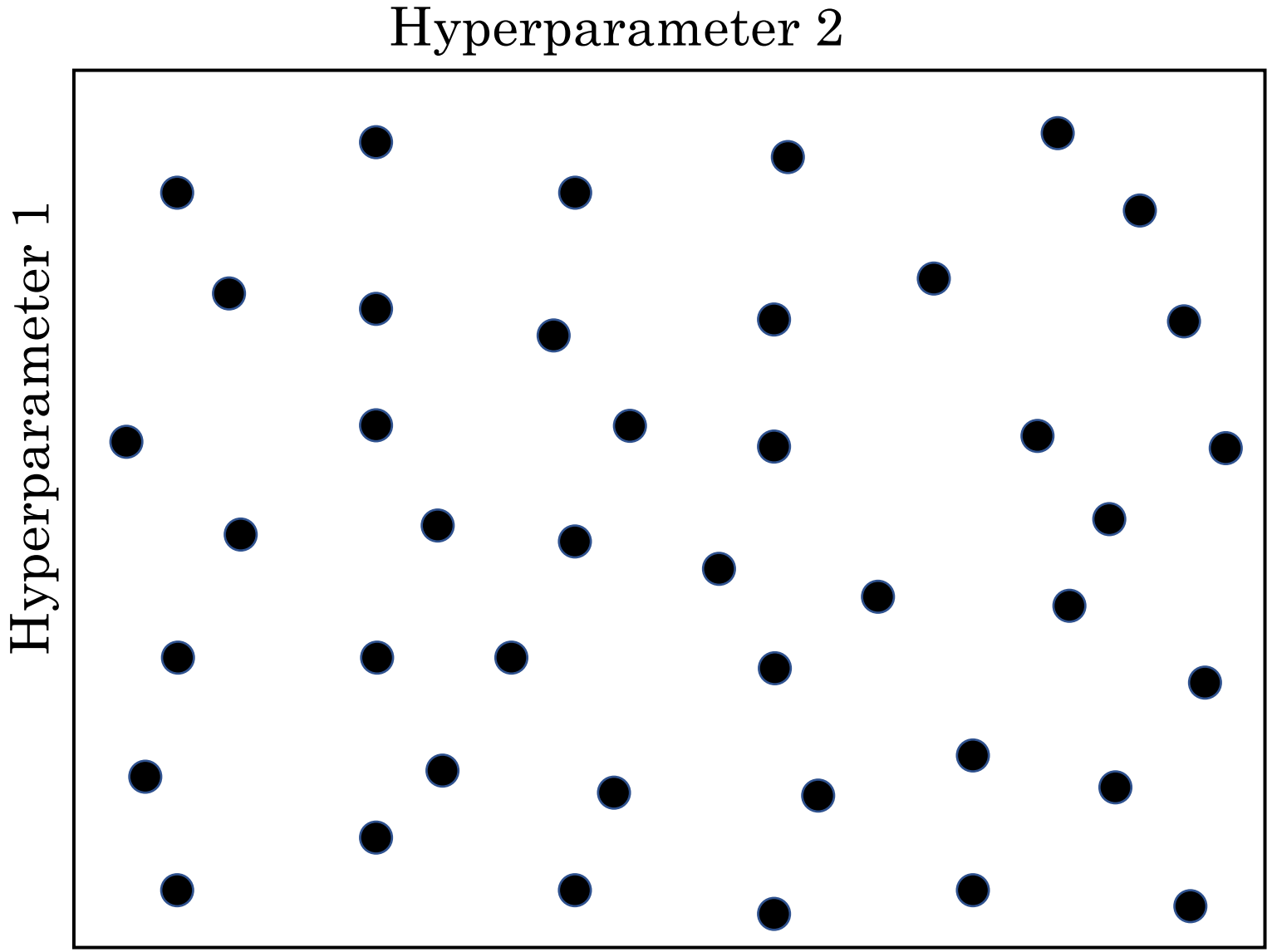
Hyperparameters

Try random values: Don’t use a grid





Coarse to fine



Andrew Ng

deeplearning.ai

Hyperparameter tuning

Using an appropriate scale to pick hyperparameters

Picking hyperparameters at random

Appropriate scale for hyperparameters

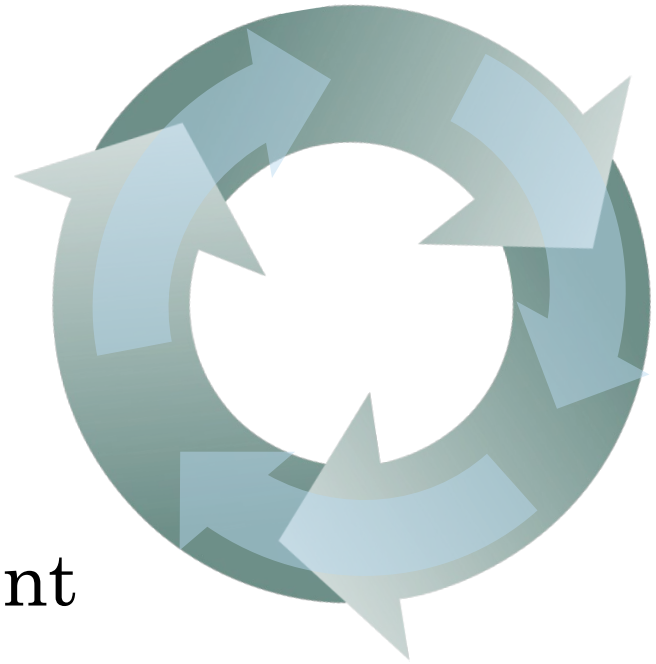
Hyperparameters for exponentially weighted averages

deeplearning.ai

Hyperparameters tuning in practice: Pandas vs. Caviar

Re-test hyperparameters occasionally

Idea



Experiment

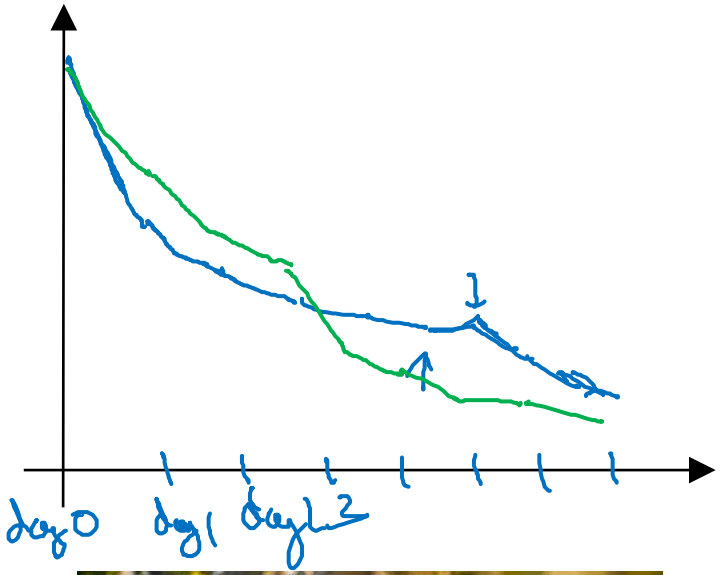
Code

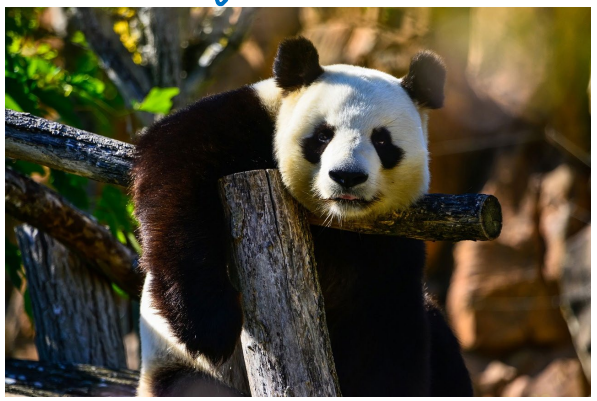
- NLP, Vision, Speech, Ads, logistics, ….

- Intuitions do get stale. Re-evaluate occasionally.

Andrew Ng

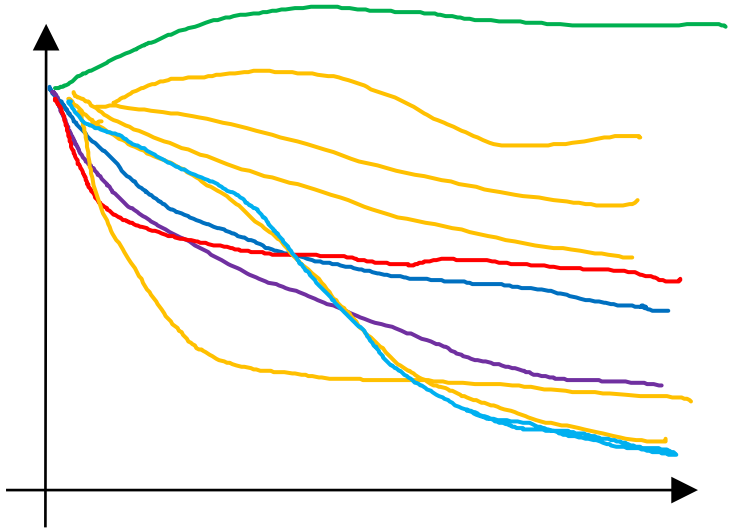
Babysitting one model





Panda

Training many models in parallel



Caviar

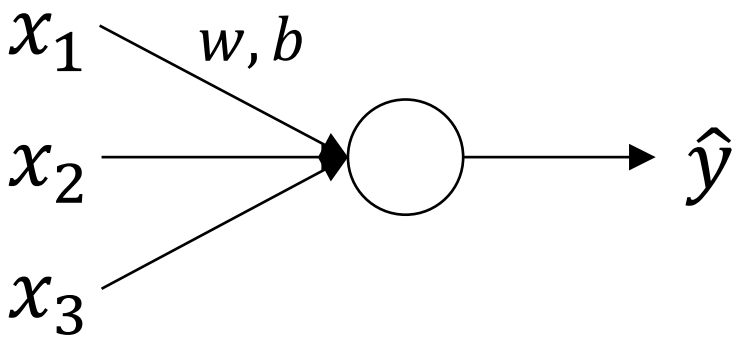
Andrew Ng

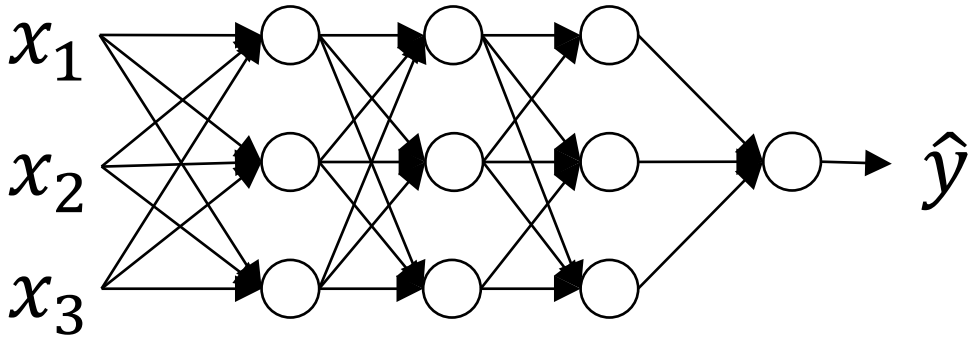
deeplearning.ai

Batch Normalization

Normalizing activations in a network

Normalizing inputs to speed up learning





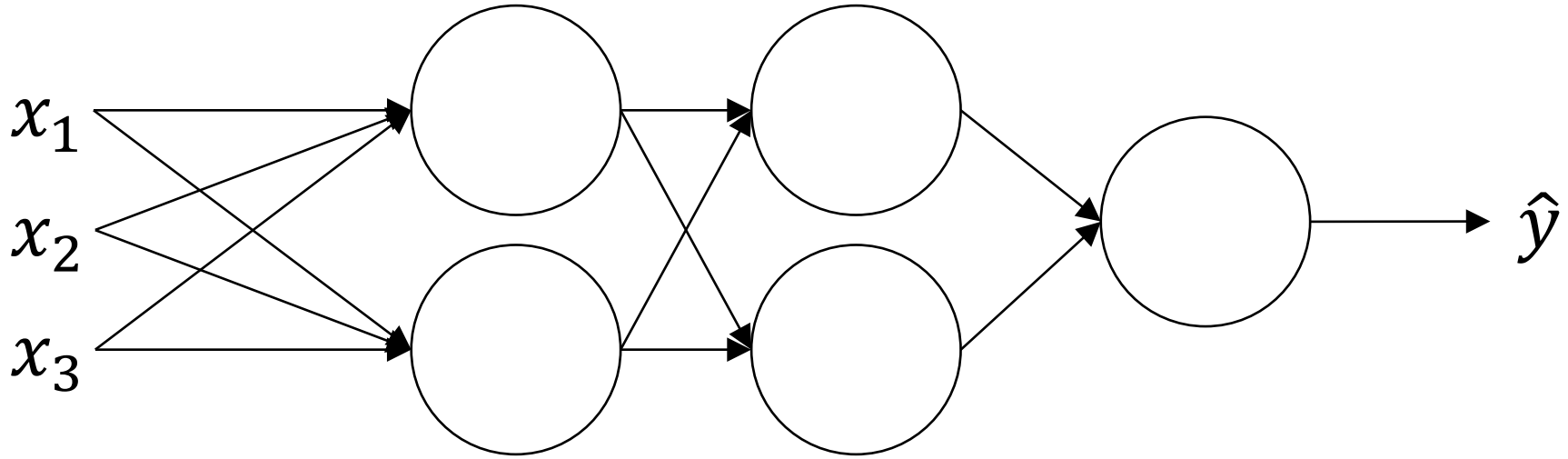
Implementing Batch Norm

deeplearning.ai

Batch Normalization

Fitting Batch Norm into a neural network

Adding Batch Norm to a network



Working with mini-batches

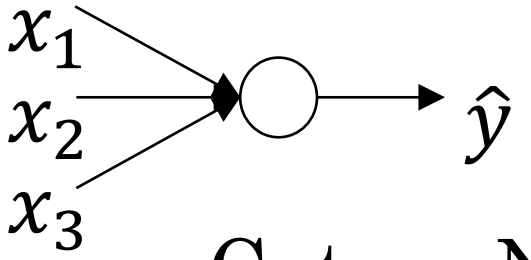
Implementing gradient descent

deeplearning.ai

Batch Normalization

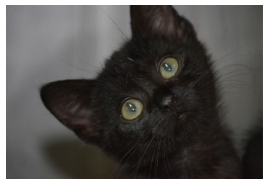
Why does Batch Norm work?

Learning on shifting input distribution



Cat
$y = 1$

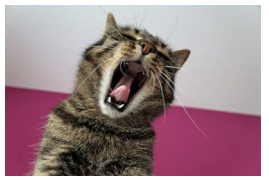


Non-Cat
$y = 0$

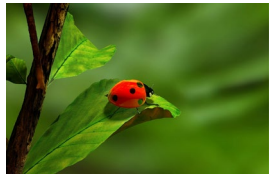



$y = 1$

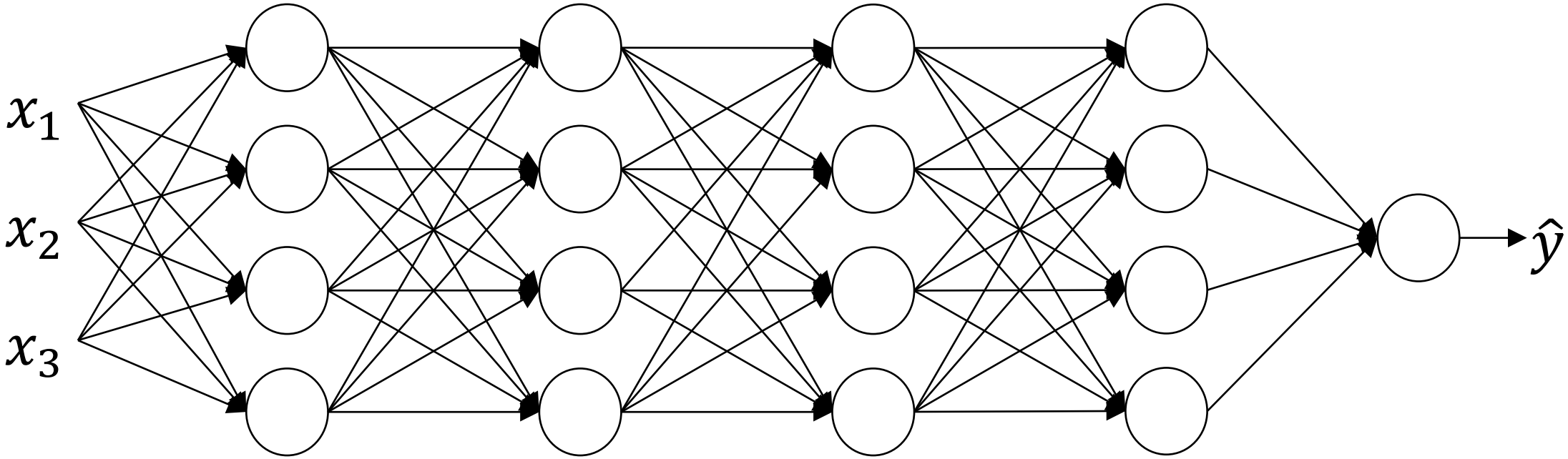
$y = 0$






Andrew Ng

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.

deeplearning.ai

Batch Normalization

Batch Norm at test time

Batch Norm at test time

$$\mu = \frac{1}{m}\sum_i z^{(i)}$$

$$\sigma^2 = \frac{1}{m}\sum_i (z^{(i)} - \mu)^2$$

$$z^{(i)}_{\text{norm}} = \frac{z^{(i)} - \mu}{\sqrt{\sigma^2 + \varepsilon}}$$

$$\tilde{z}^{(i)} = \gamma z^{(i)}_{\text{norm}} + \beta$$

Andrew Ng

deeplearning.ai

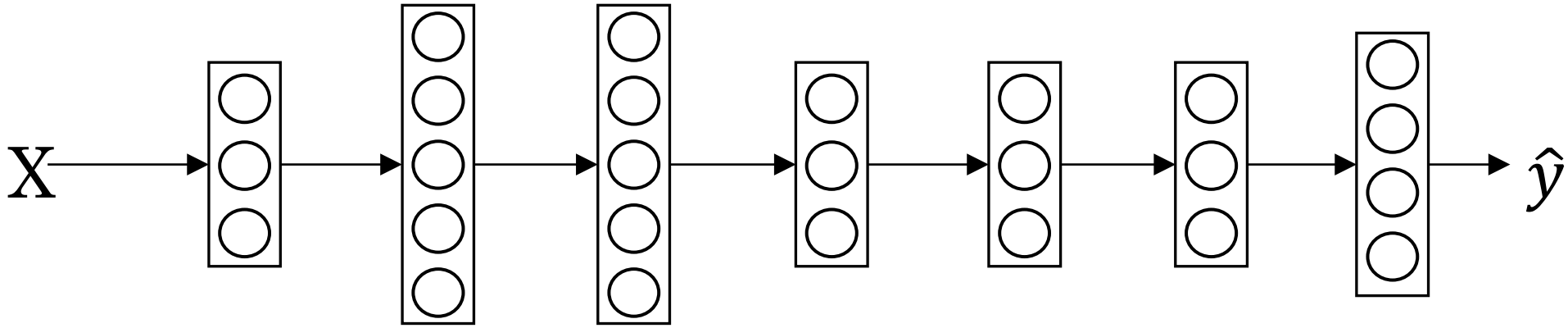
Multi-class classification

Trying a softmax classifier

Understanding softmax

Loss function

Summary of softmax classifier



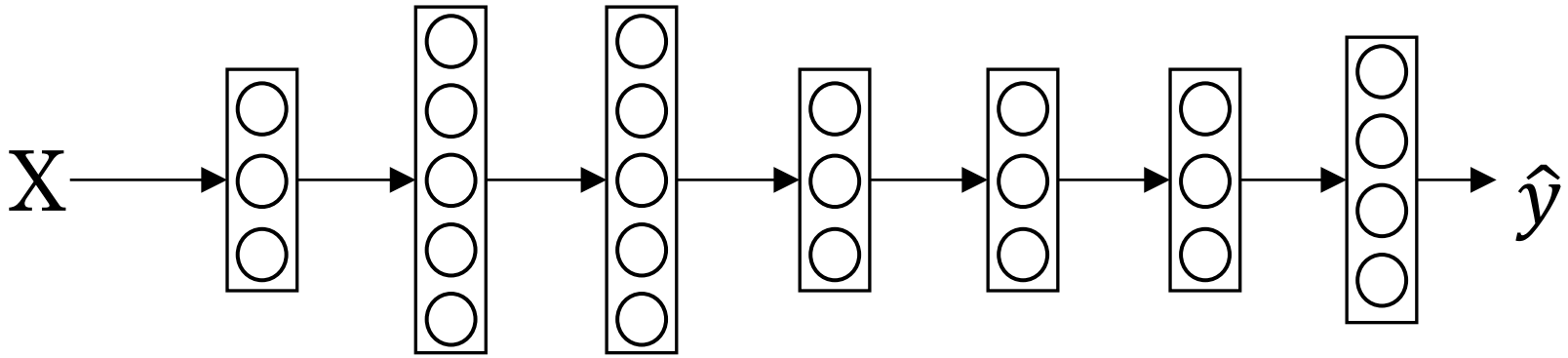
deeplearning.ai

Multi-class classification

Softmax regression

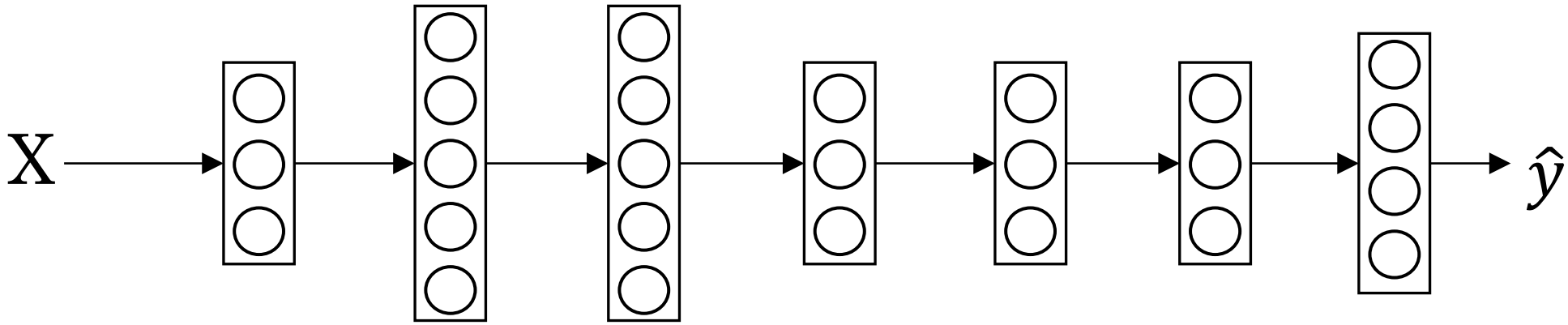
Recognizing cats, dogs, and baby chicks

3 1 2 0 3 2 0 1



Andrew Ng

Softmax layer



Softmax examples