deeplearning.ai

Setting up your ML application

Train/dev/test sets

Applied ML is a highly iterative process

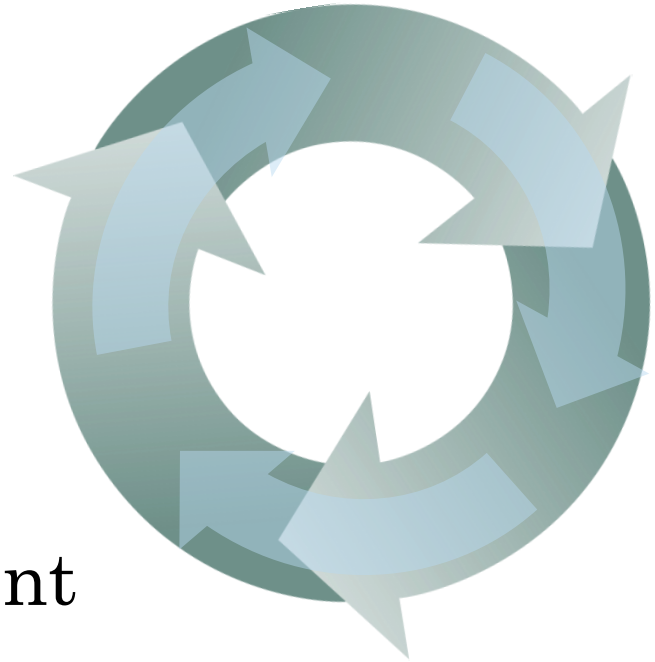
layers

hidden units

learning rates

activation functions

…



Train/dev/test sets

Mismatched train/test distribution

Training set:
Cat pictures from webpages

Dev/test sets:
Cat pictures from users using your app

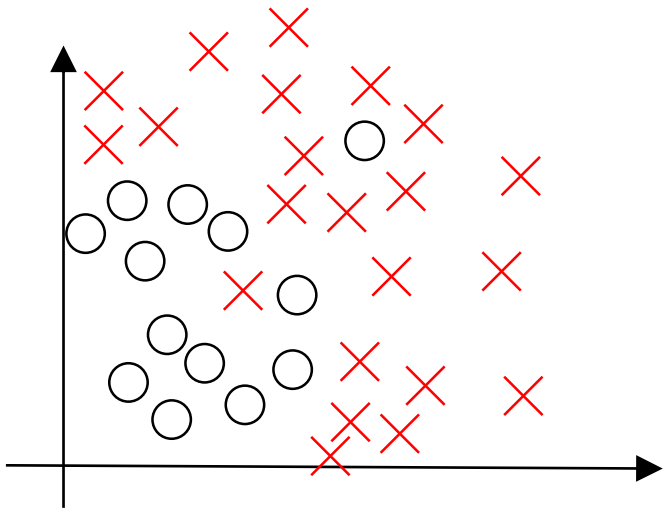
Not having a test set might be okay. (Only dev set.)

deeplearning.ai

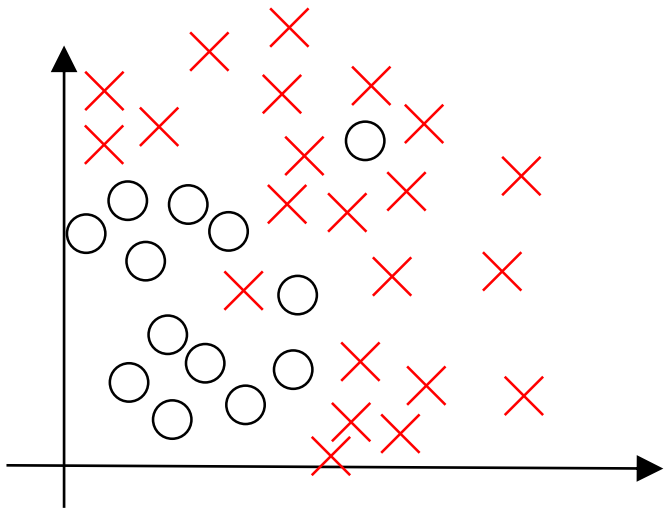
Setting up your ML application

Bias/Variance

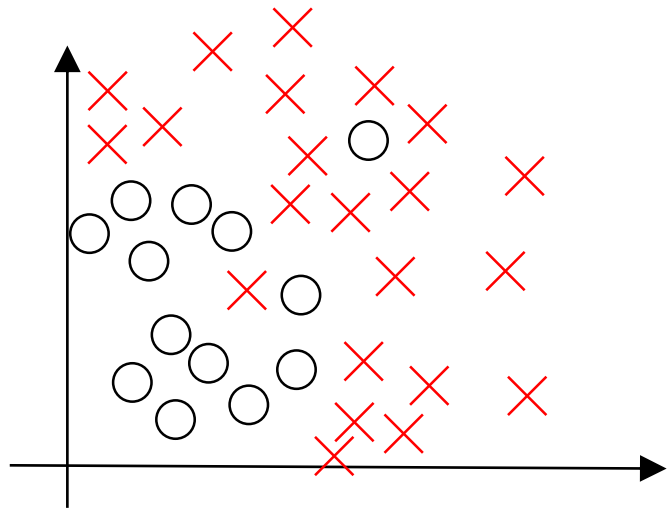
Bias and Variance



high bias



“just right”



high variance

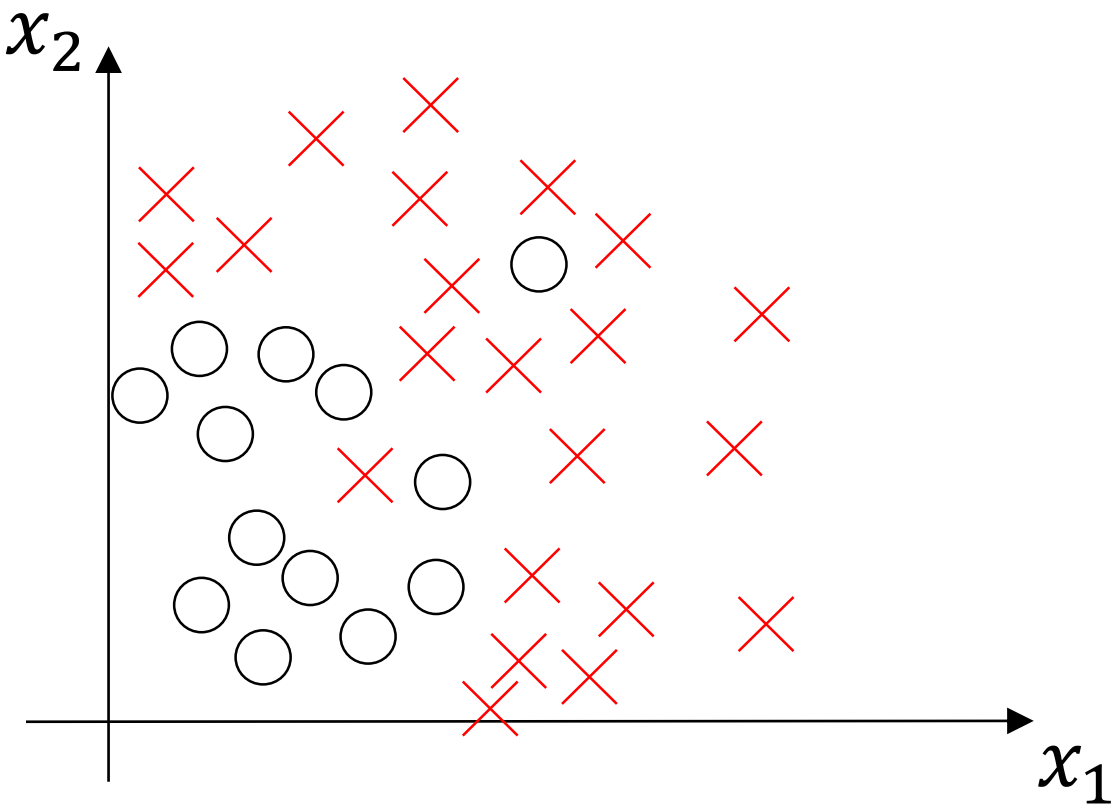
Bias and Variance

Cat classification

Train set error:

Dev set error:

High bias and high variance



deeplearning.ai

Setting up your ML application

Basic “recipe” for machine learning

Basic “recipe” for machine learning

Basic recipe for machine learning

deeplearning.ai

Regularizing your neural network

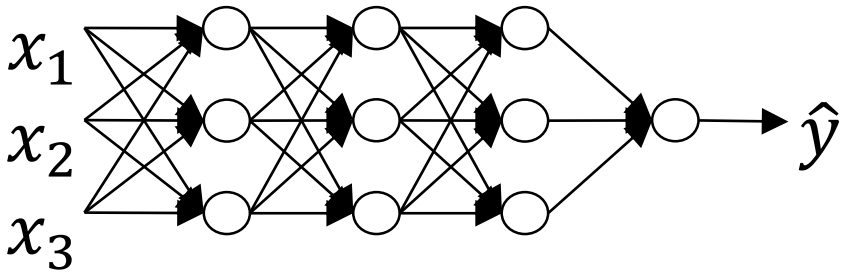
Regularization

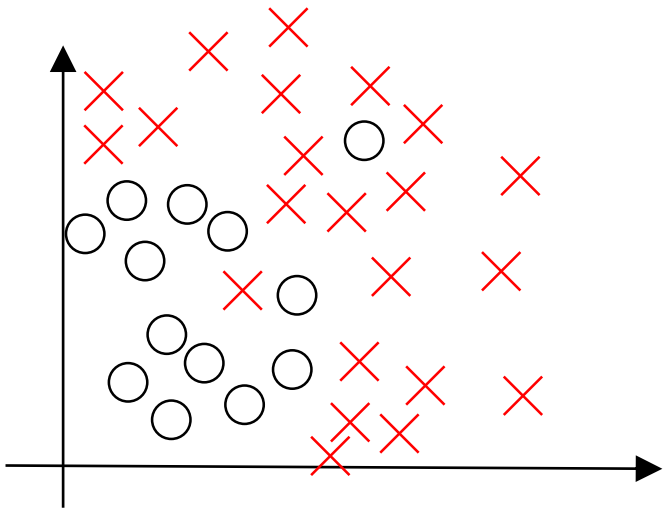
Logistic regression

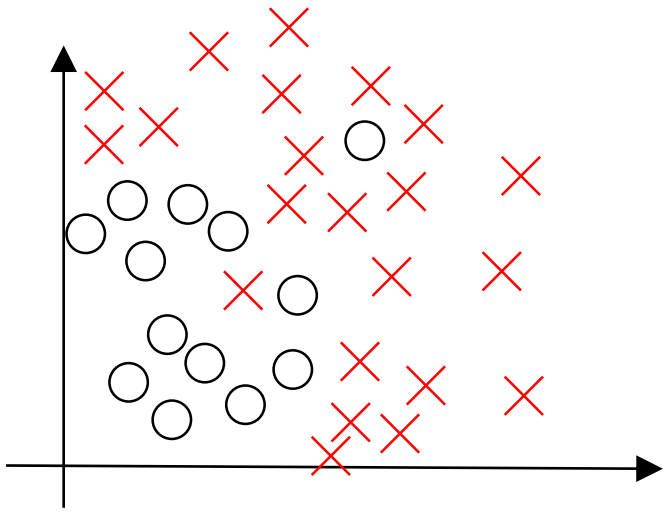
$$\min_{w,b} J(w,b)$$

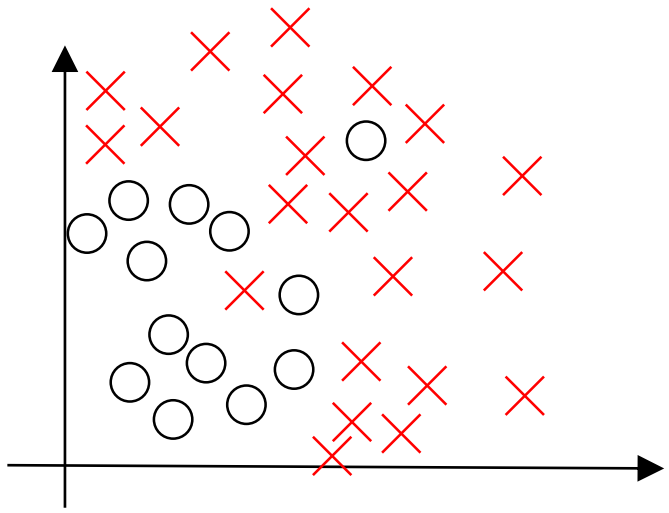
Neural network

How does regularization prevent overfitting?









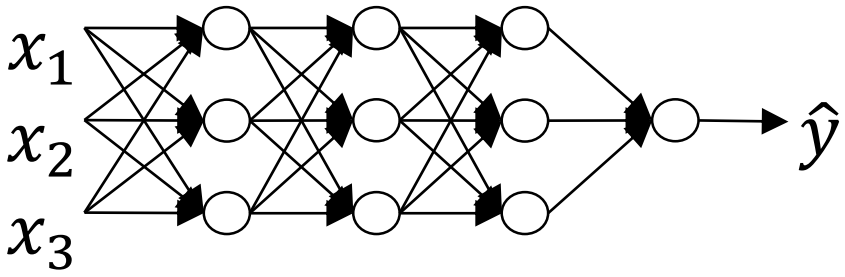
How does regularization prevent overfitting?

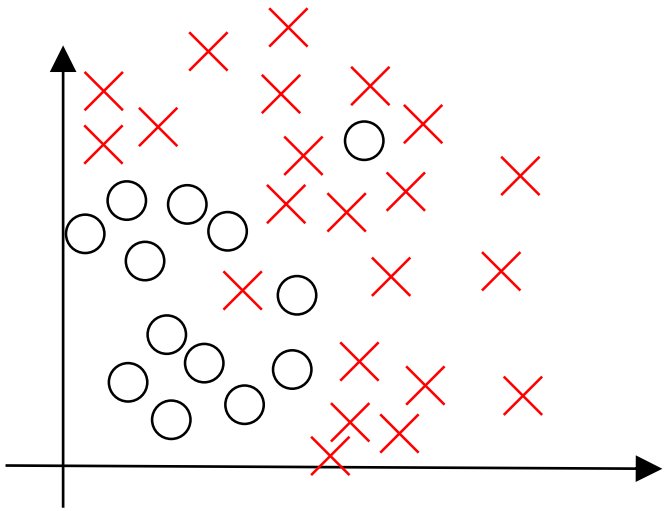
deeplearning.ai

Regularizing your neural network

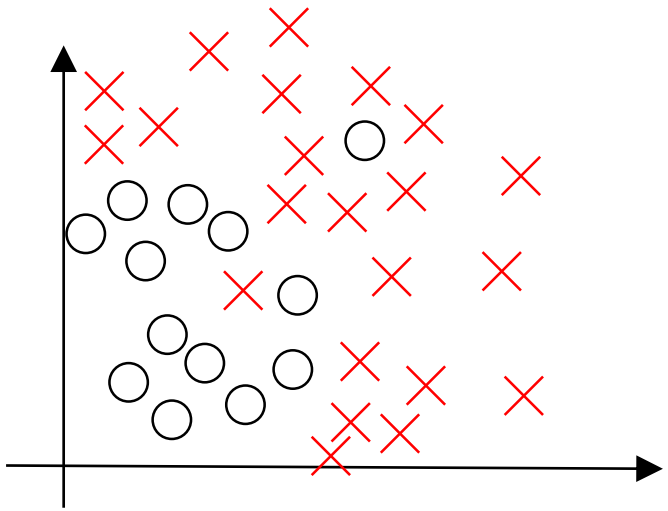
Why regularization reduces overfitting

How does regularization prevent overfitting?

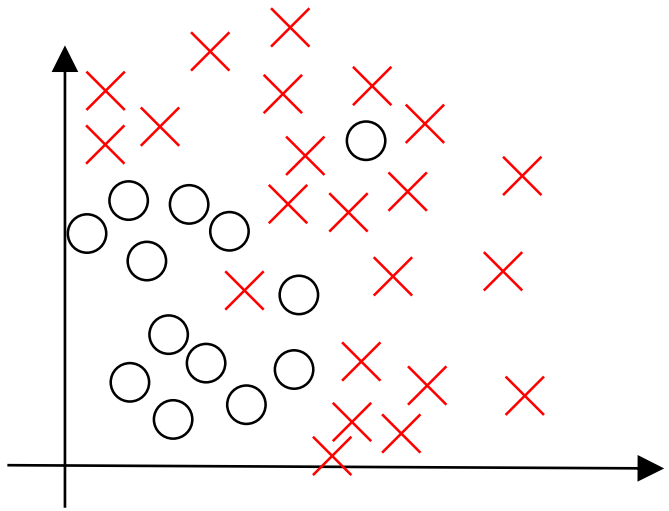




high bias



“just right”



high variance

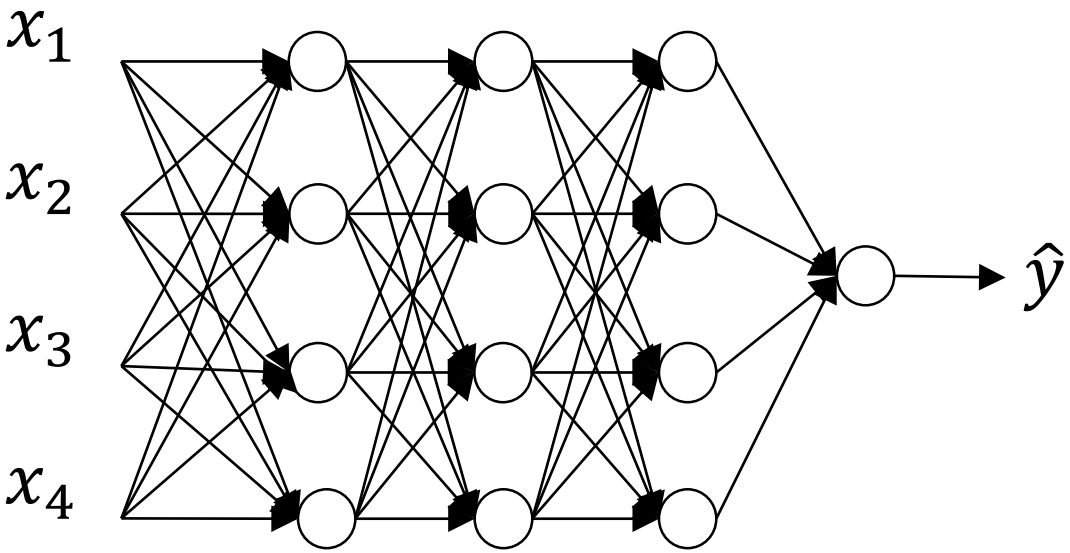
How does regularization prevent overfitting?

deeplearning.ai

Regularizing your neural network

Dropout regularization

Dropout regularization



Implementing dropout (“Inverted dropout”)

Making predictions at test time

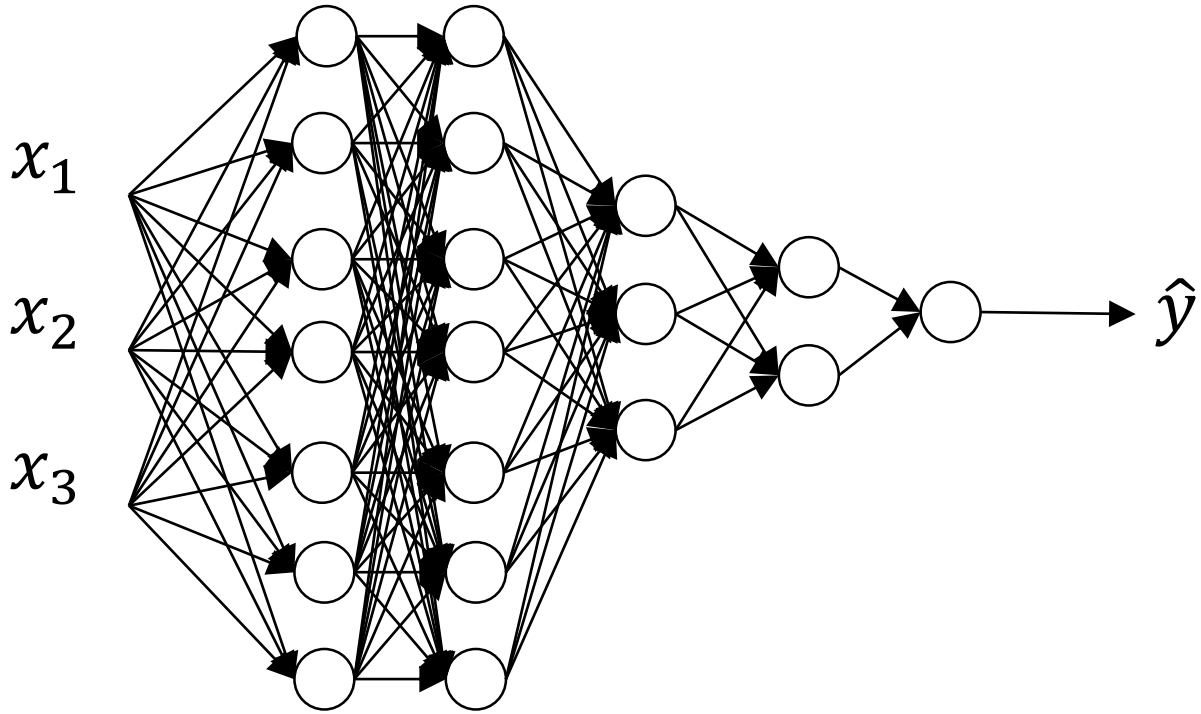
deeplearning.ai

Regularizing your neural network

Understanding dropout

Why does drop-out work?

Intuition: Can't rely on any one feature, so have to spread out weights.



deeplearning.ai

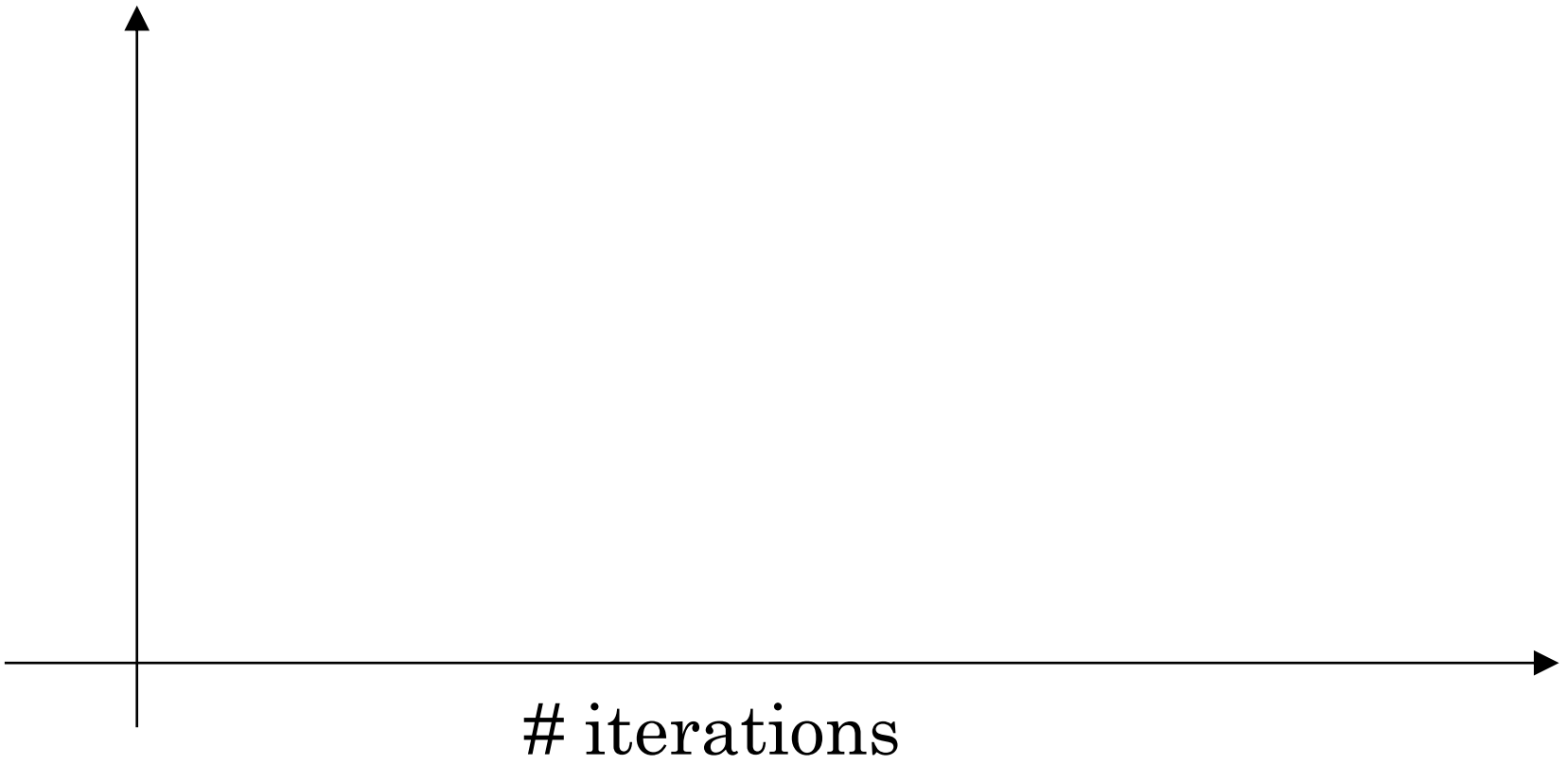
Regularizing your neural network

Other regularization methods

Data augmentation

4

Early stopping

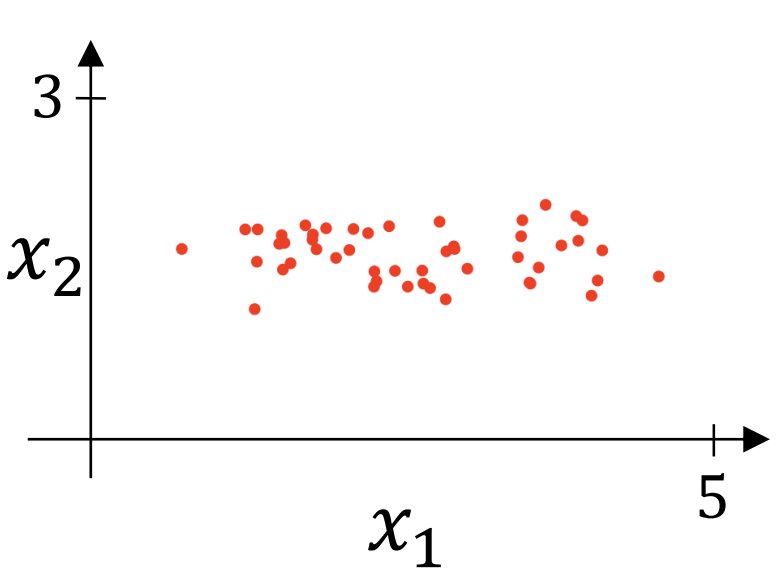


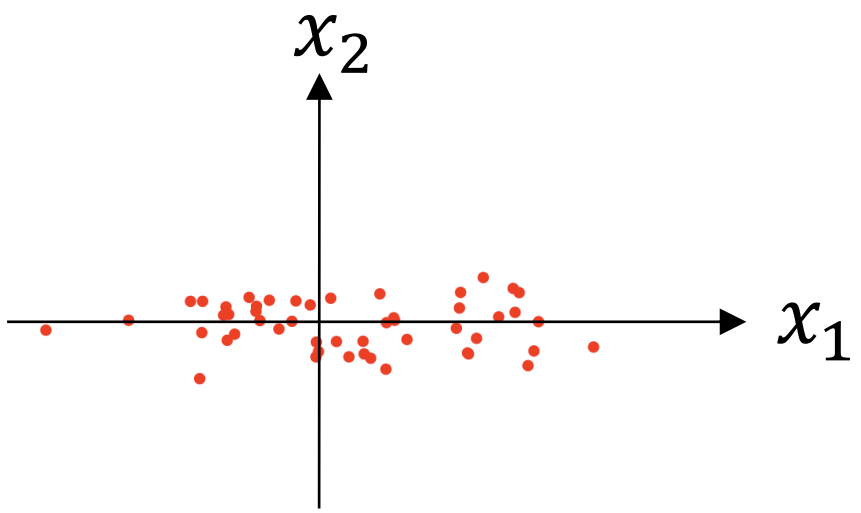
deeplearning.ai

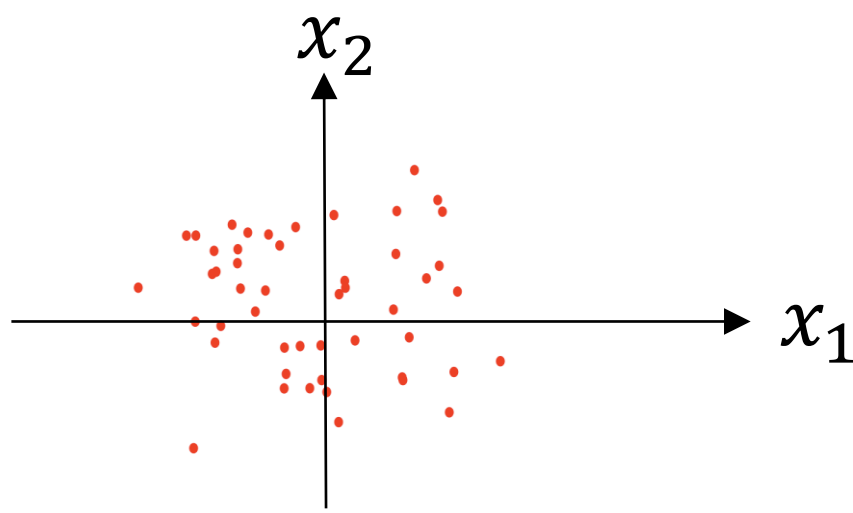
Setting up your optimization problem

Normalizing inputs

Normalizing training sets



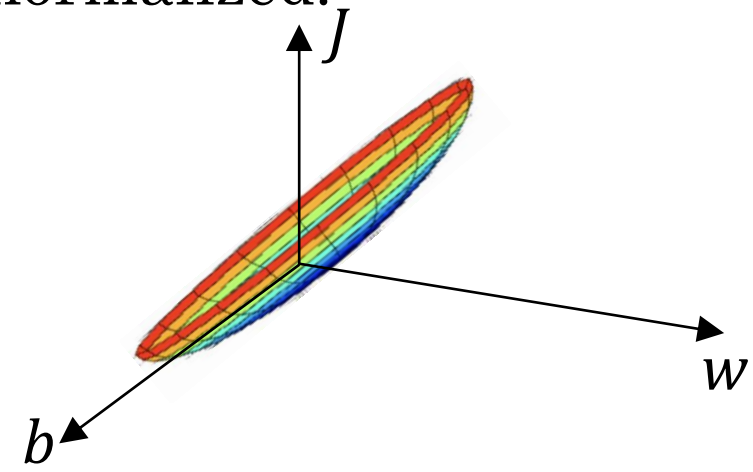


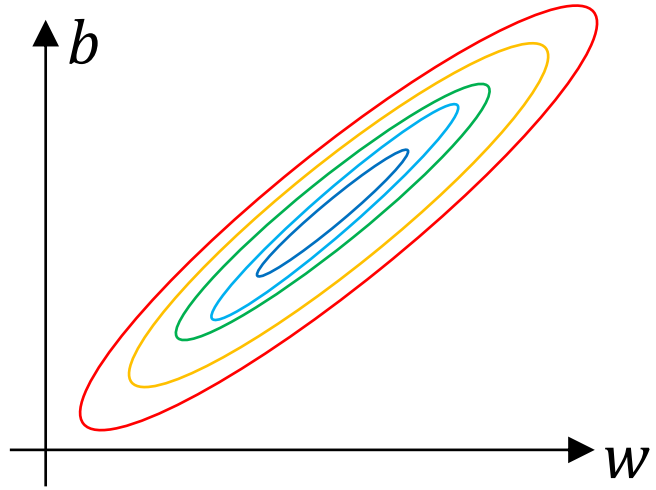


Why normalize inputs?

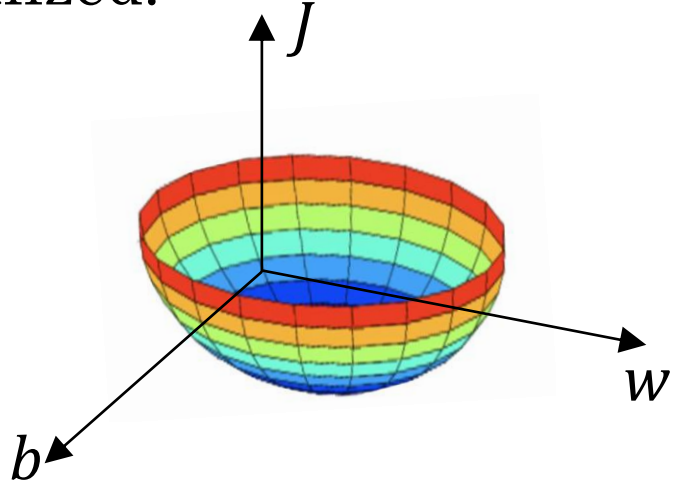
$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m} \mathcal{L}\left(\hat{y}^{(i)}, y^{(i)}\right)$$

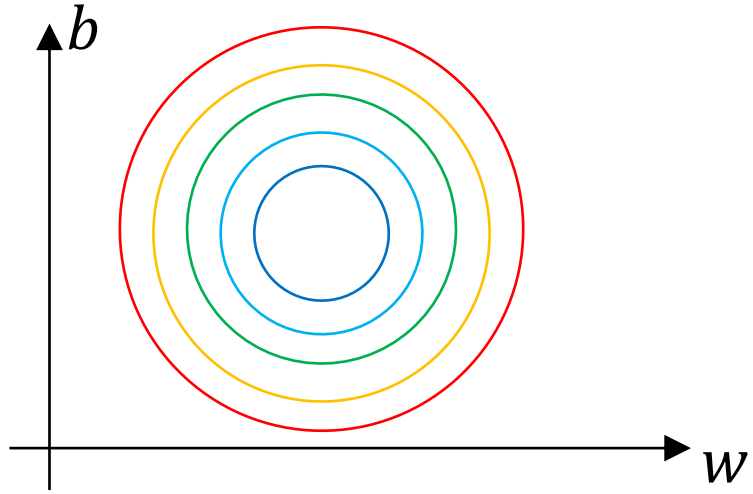
Unnormalized:





Normalized:





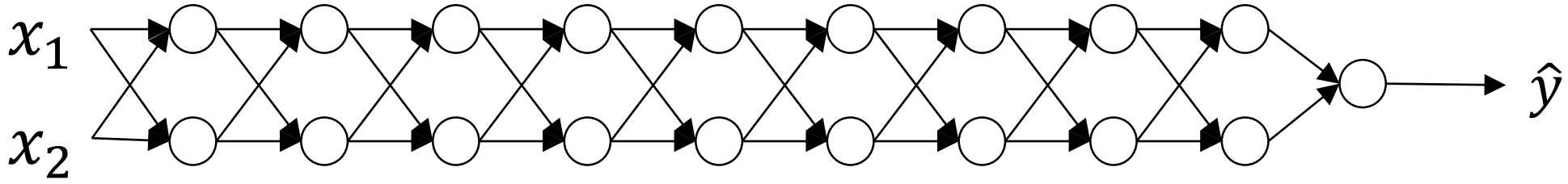
Andrew Ng

deeplearning.ai

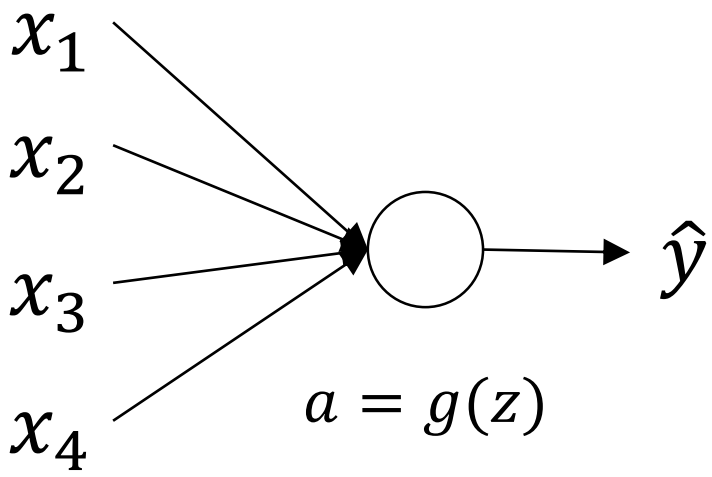
Setting up your optimization problem

Vanishing/exploding gradients

Vanishing/exploding gradients



Single neuron example

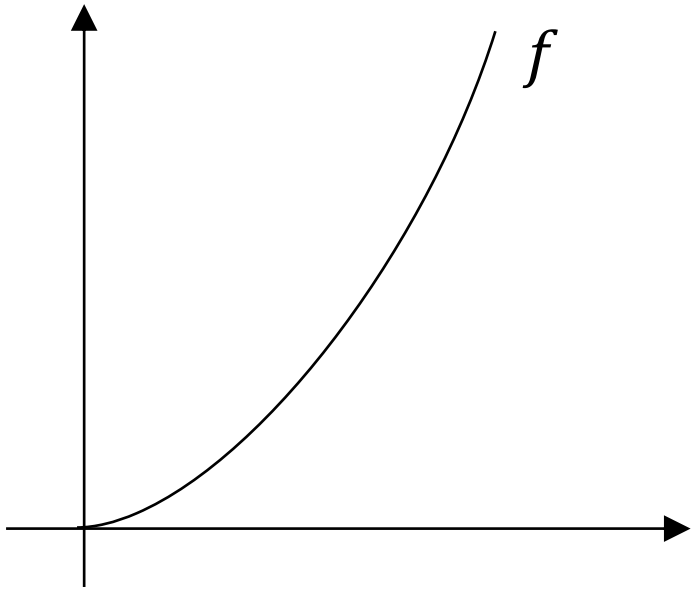


deeplearning.ai

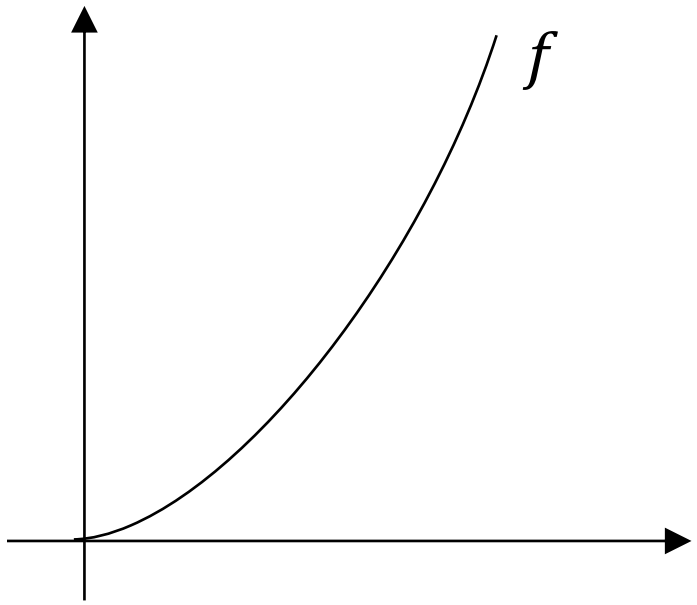
Setting up your optimization problem

Numerical approximation of gradients

Checking your derivative computation



Checking your derivative computation



deeplearning.ai

Setting up your optimization problem

Gradient Checking

Gradient check for a neural network

Take $W^{[1]}, b^{[1]}, \ldots, W^{[L]}, b^{[L]}$ and reshape into a big vector θ.

Take $dW^{[1]}, db^{[1]}, \ldots, dW^{[L]}, db^{[L]}$ and reshape into a big vector $d\theta$.

Gradient checking (Grad check)

deeplearning.ai

Setting up your optimization problem

Gradient Checking implementation notes

Gradient checking implementation notes

- Don’t use in training – only to debug

- If algorithm fails grad check, look at components to try to identify bug.

- Remember regularization.

- Doesn’t work with dropout.

- Run at random initialization; perhaps again after some training.