Intro to Neural Networks

Data Science Immersive

Henry Cham

FLATIRON SCHOOL

Quote:

AI is the new electricity. Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don’t think AI will transform in the next several years.

-Andrew Ng

Question:

What are some current use cases for neural networks?

Current use cases:

Supervised Learning

Input(x)	Output (y)	Application
Home features	Price	Real Estate
Ad, user info	Click on ad? (0/1)	Online Advertising
Image	Object (1,…,1000)	Photo tagging
Audio	Text transcript	Speech recognition
English	Chinese	Machine translation
Image, Radar info	Position of other cars	Autonomous driving

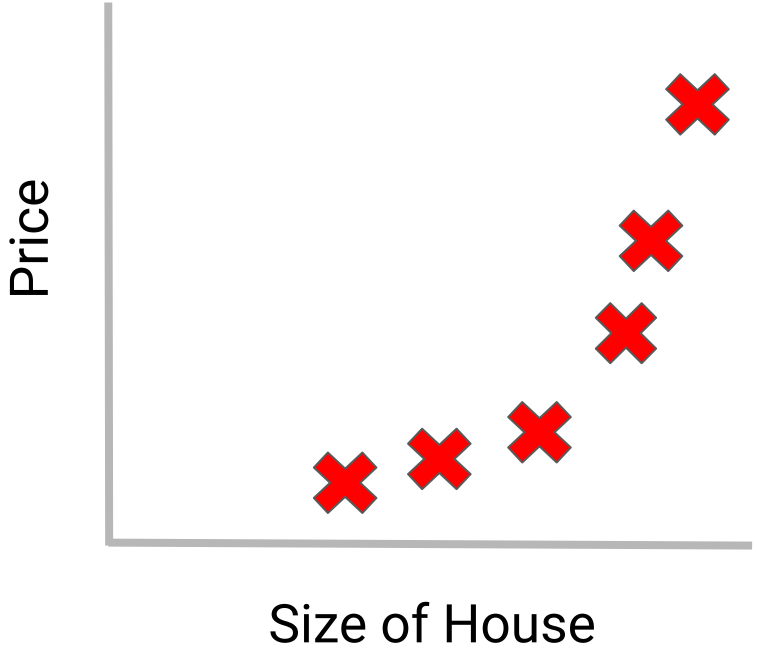
Andrew Ng, Neural Networks and Deep Learning, Coursera

Aims:

- Examine the intuition of why neural networks can solve complex problems
- Explain the mathematics operations underlying how artificial neurons predict and learn

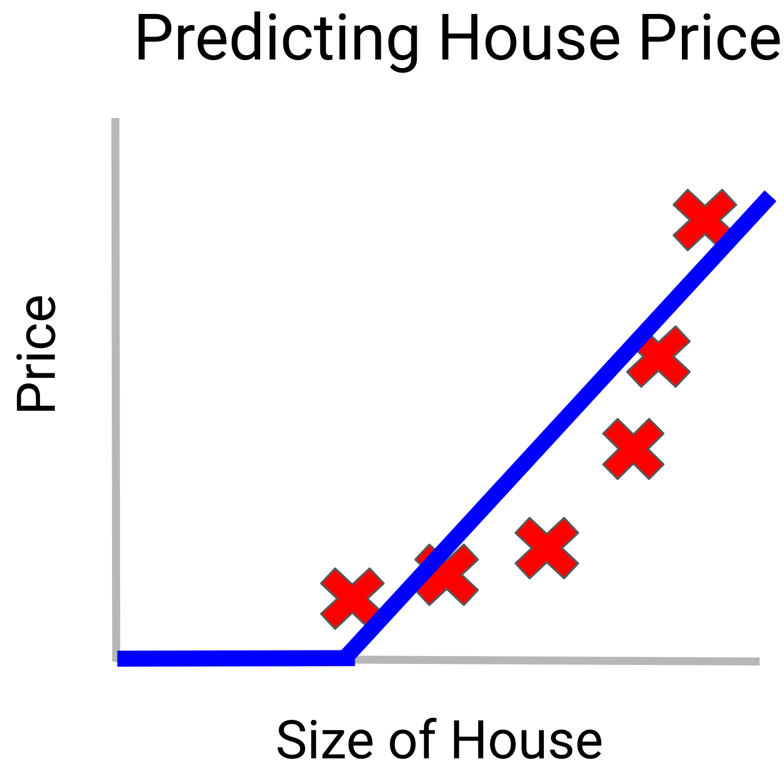
Why do neural networks work?

Predicting House Price



Andrew Ng, Neural Networks and Deep Learning, Coursera

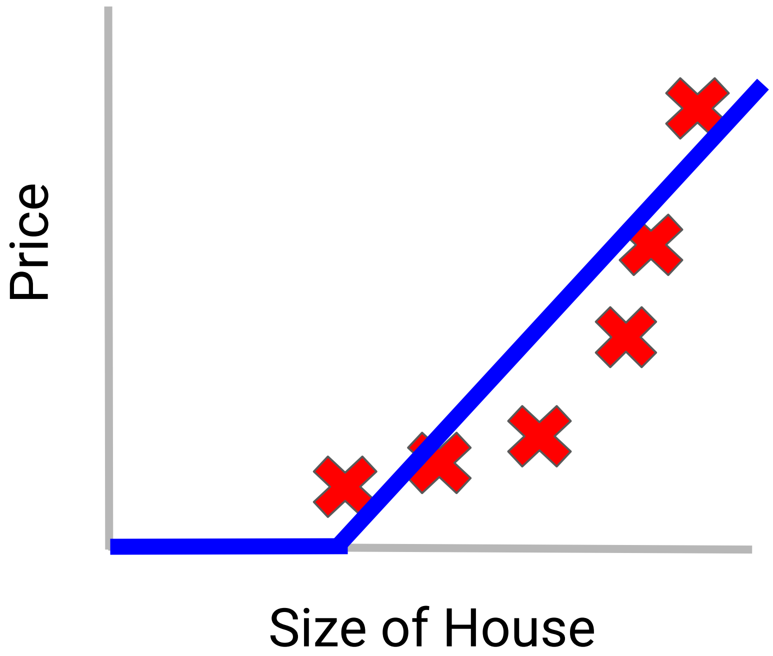
Why do neural networks work?



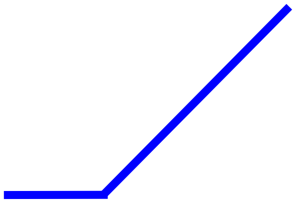
Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

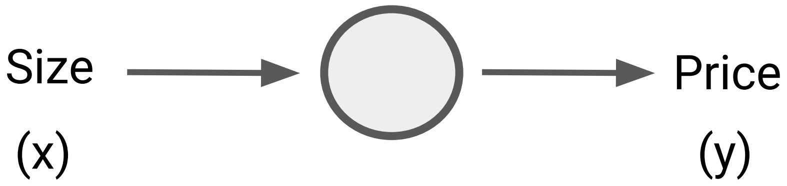
Predicting House Price



Rectified Linear (Relu)



Neuron



Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

Predicting House Price

size

# of bedrooms

zip code

wealth

(x)

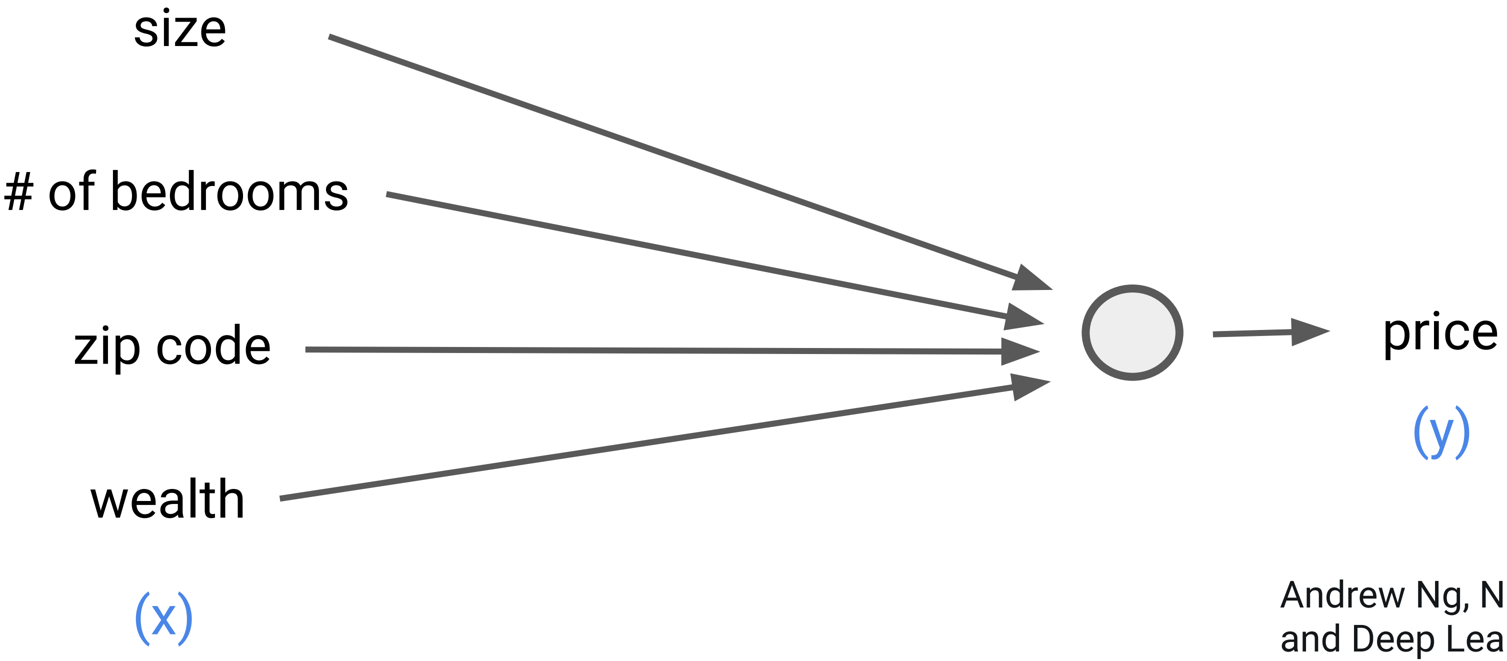
price

(y)

Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

Predicting House Price

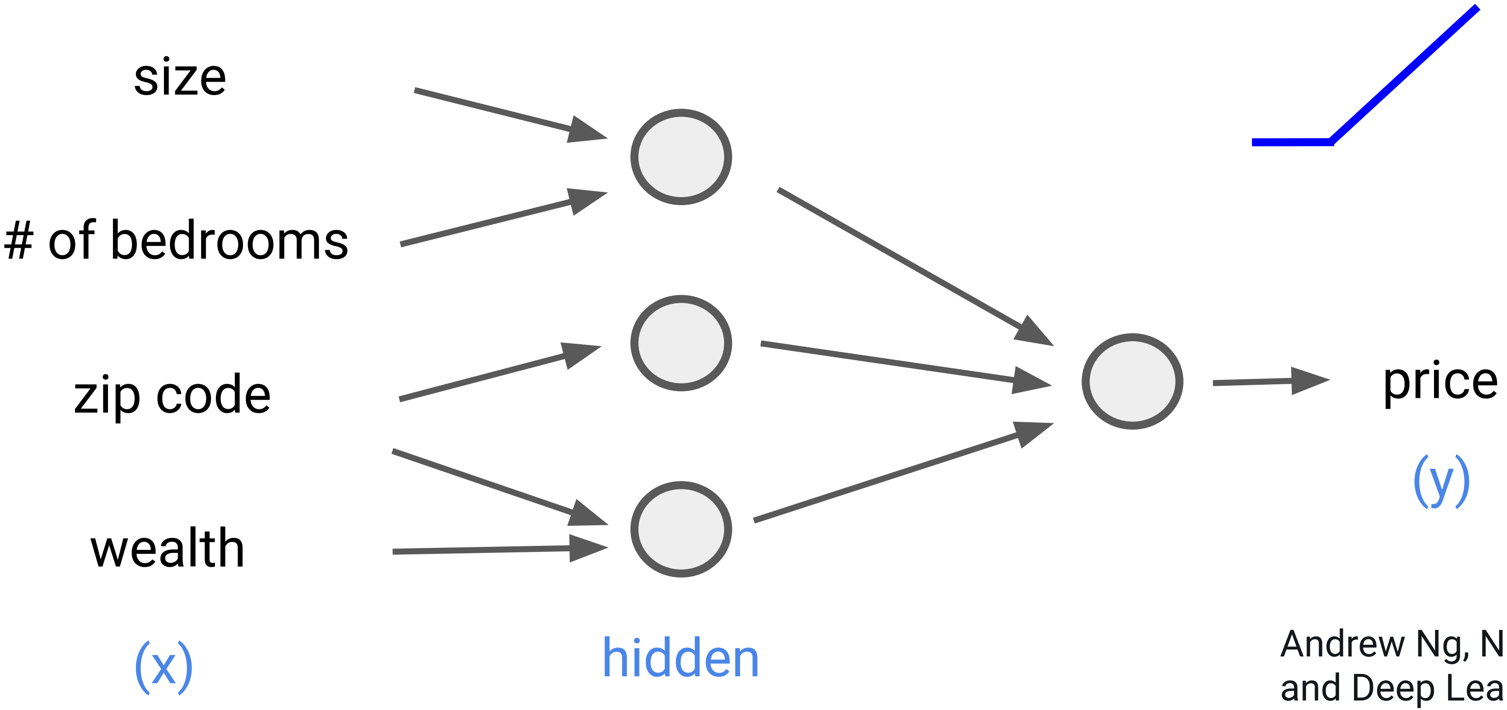


Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

Predicting House Price

Rectified Linear (Relu)

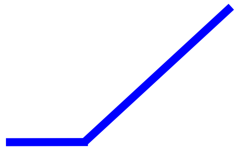


Andrew Ng, Neural Networks and Deep Learning, Coursera

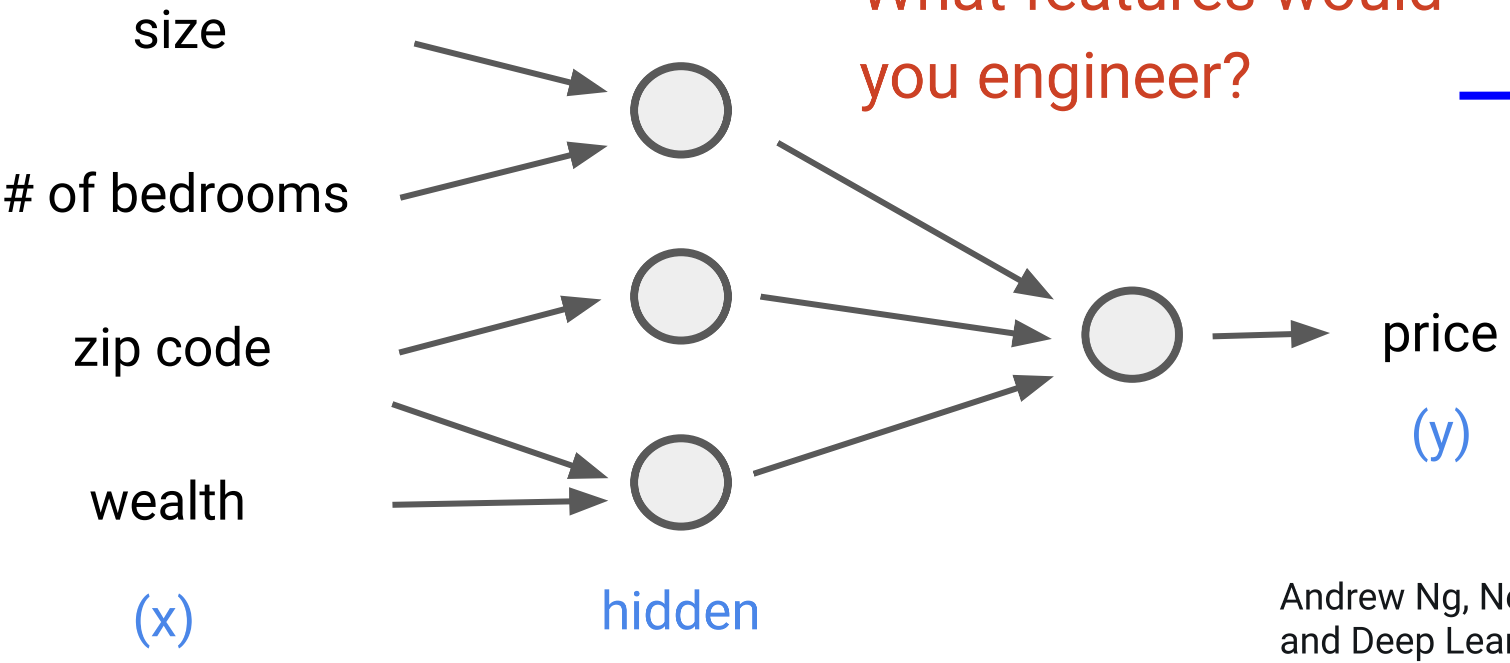
Why do neural networks work?

Predicting House Price

Rectified Linear (Relu)

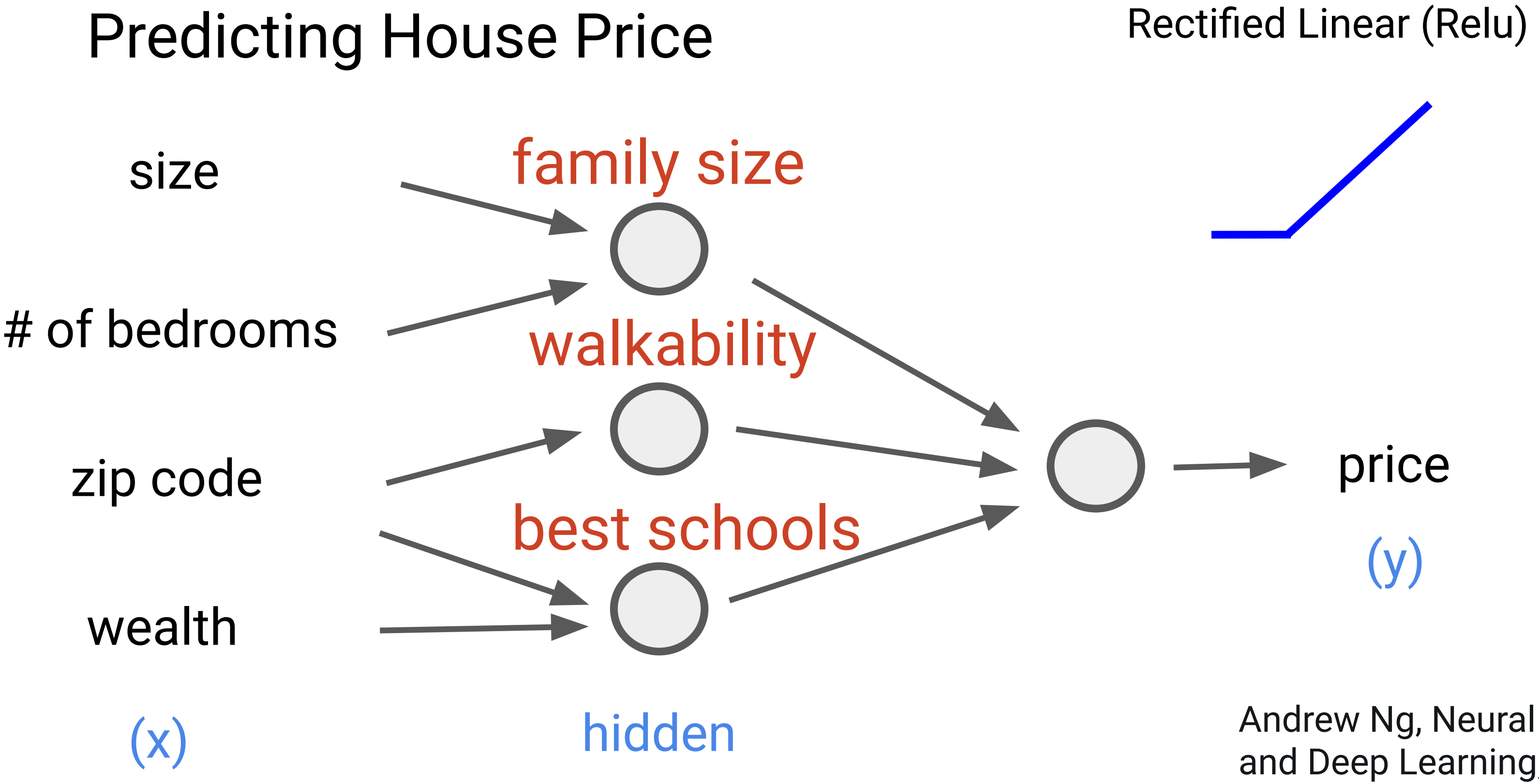


What features would you engineer?



Andrew Ng, Neural Networks and Deep Learning, Coursera

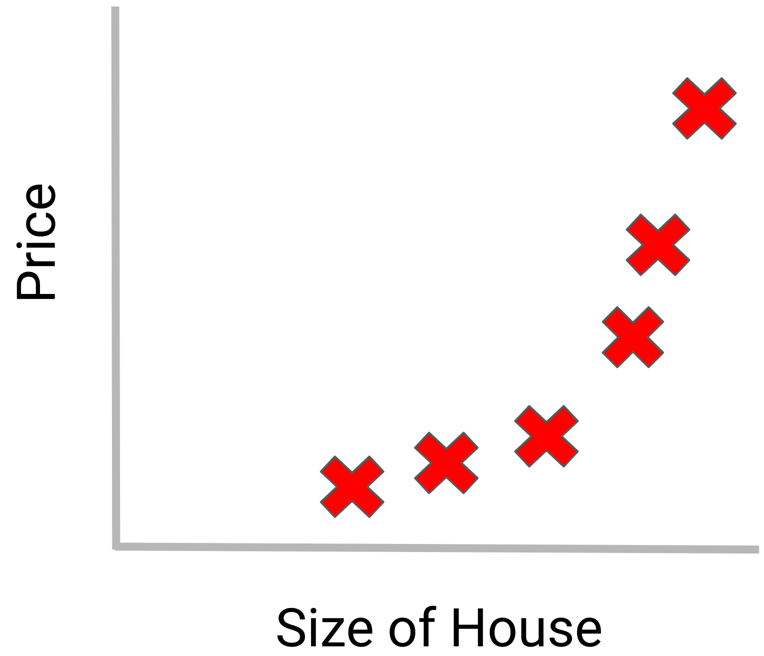
Why do neural networks work?



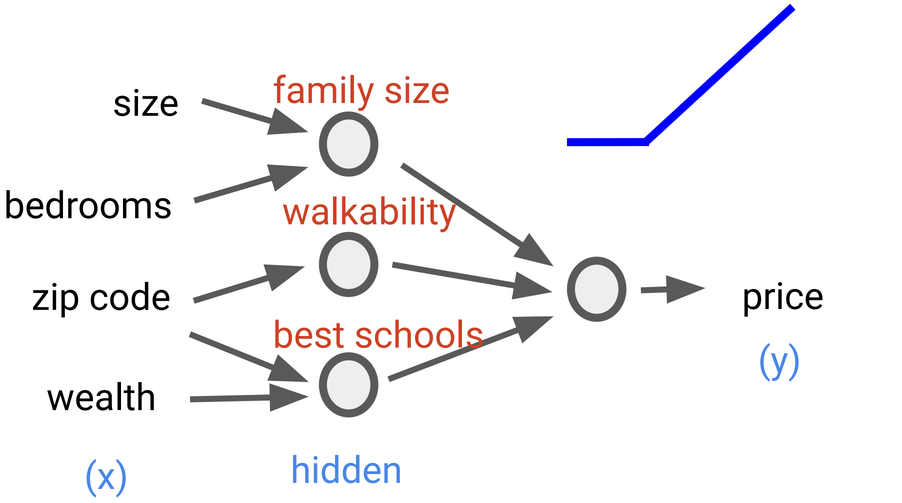
Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

Predicting House Price



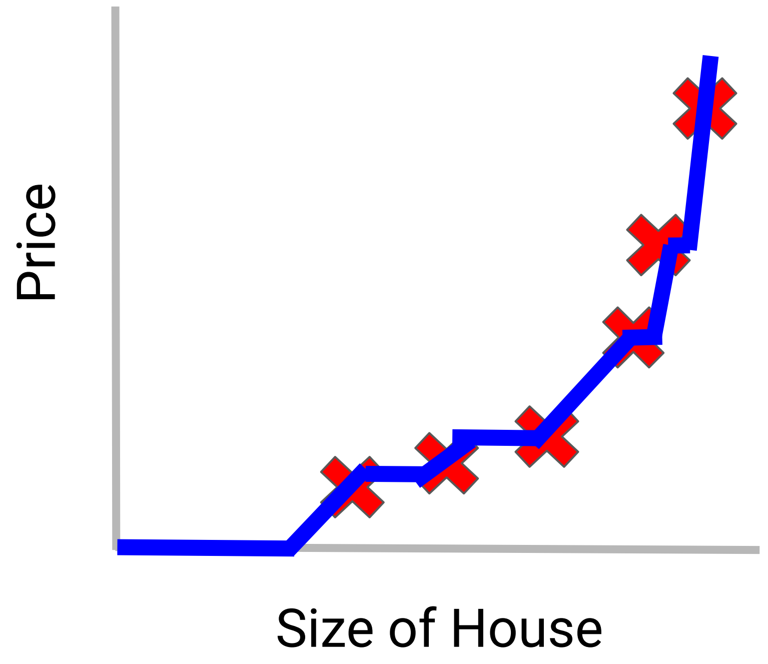
Rectified Linear (Relu)



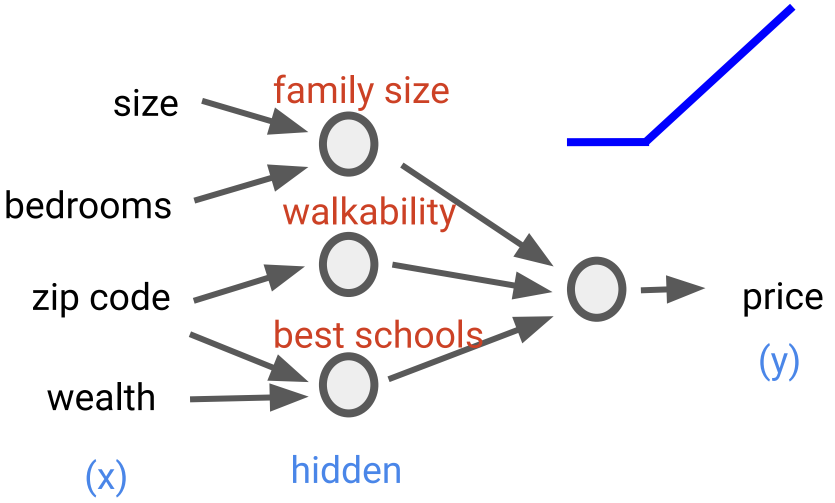
Andrew Ng, Neural Networks and Deep Learning, Coursera

Why do neural networks work?

Predicting House Price



Rectified Linear (Relu)

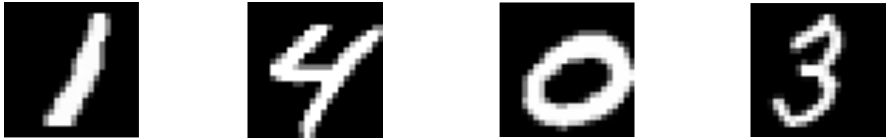


Andrew Ng, Neural Networks and Deep Learning, Coursera

What do hidden layers capture?

MNIST

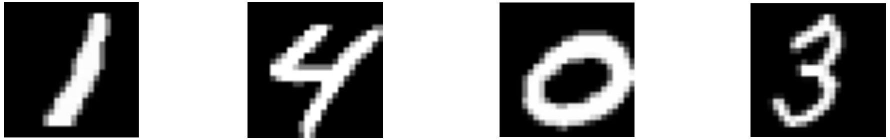
Input



What do hidden layers capture?

MNIST

Input



What do the hidden layers of a network train on MNIST data capture?

What do hidden layers capture?

MNIST

Input

Hidden Neuron

A

What do hidden layers capture?

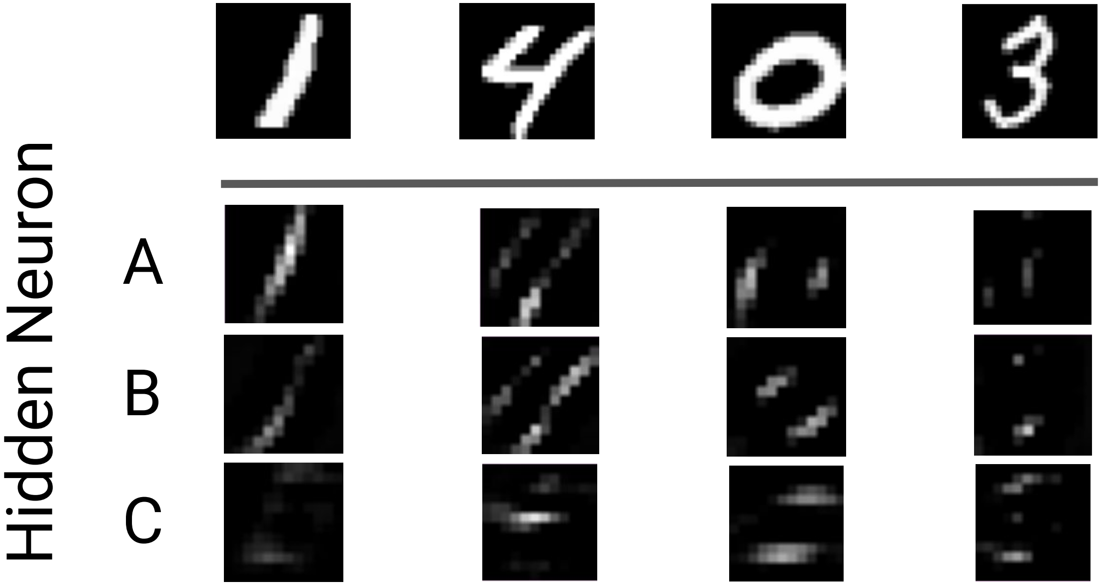
MNIST

Input

What do hidden layers capture?

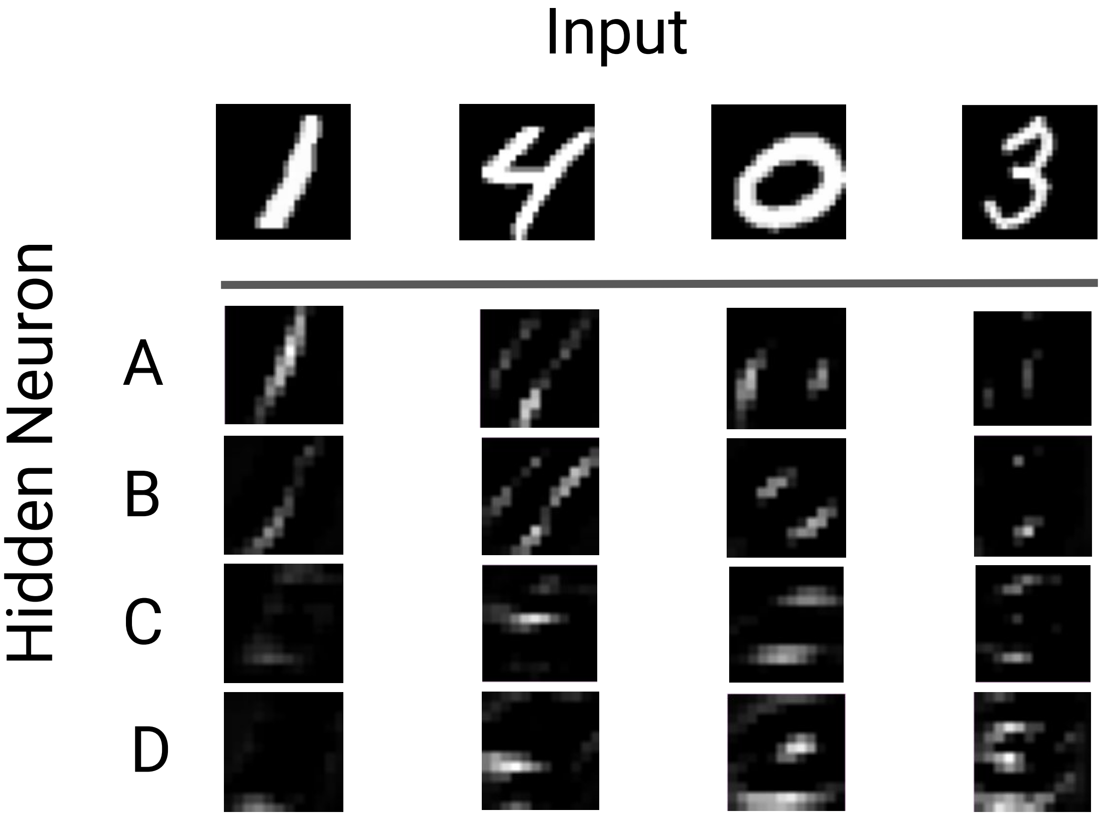
MNIST

Input

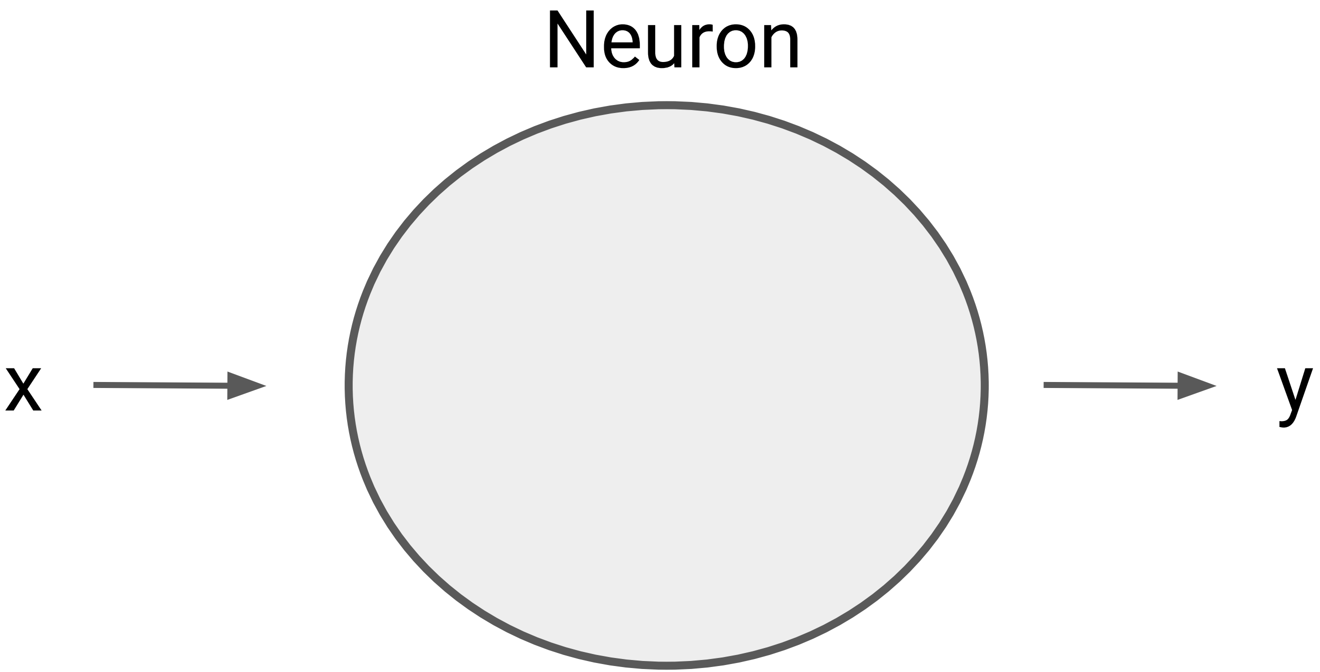


What do hidden layers capture?

MNIST

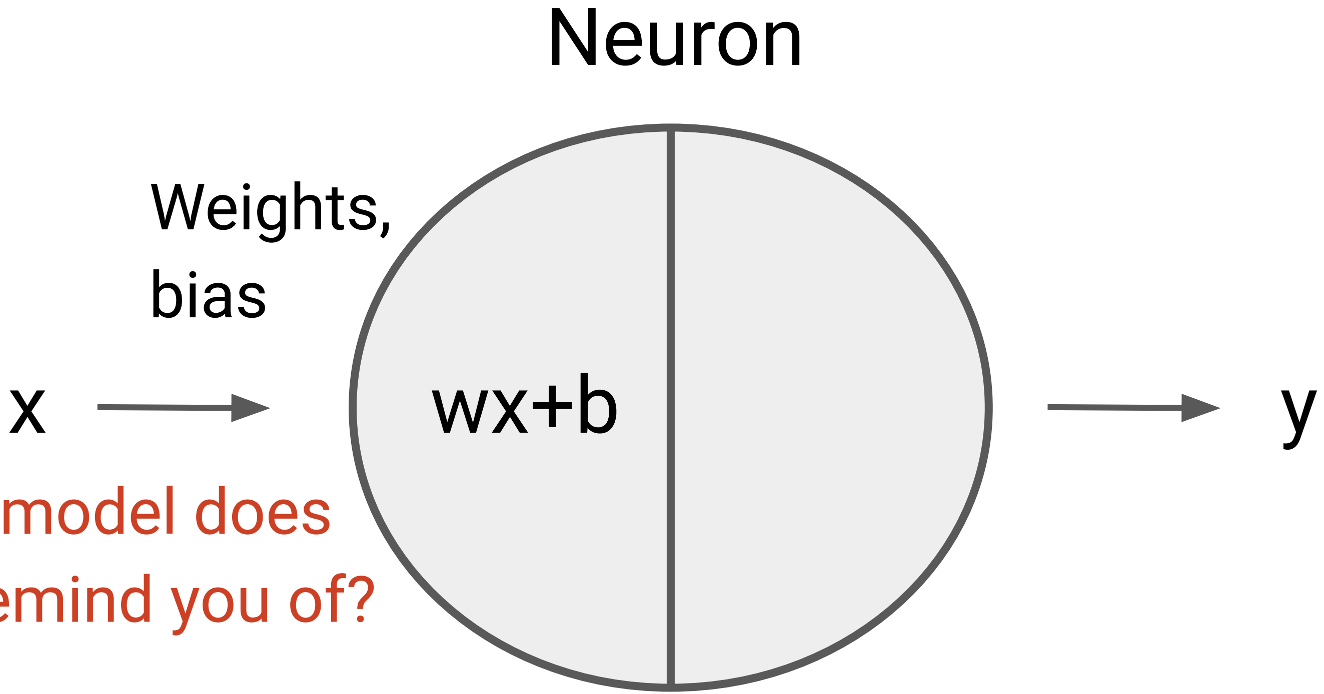


Simple Neural Network



Andrew Ng, Neural Networks and Deep Learning, Coursera

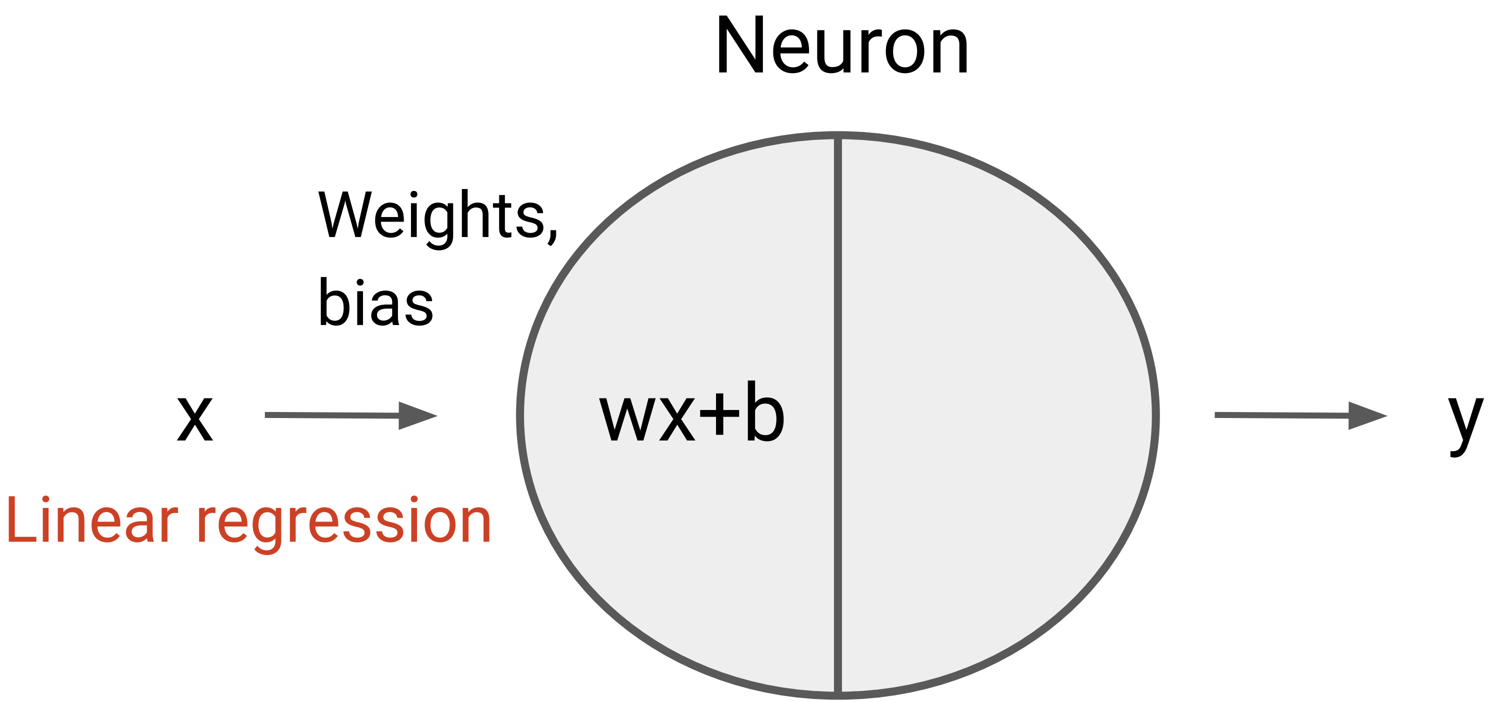
Simple Neural Network



What model does this remind you of?

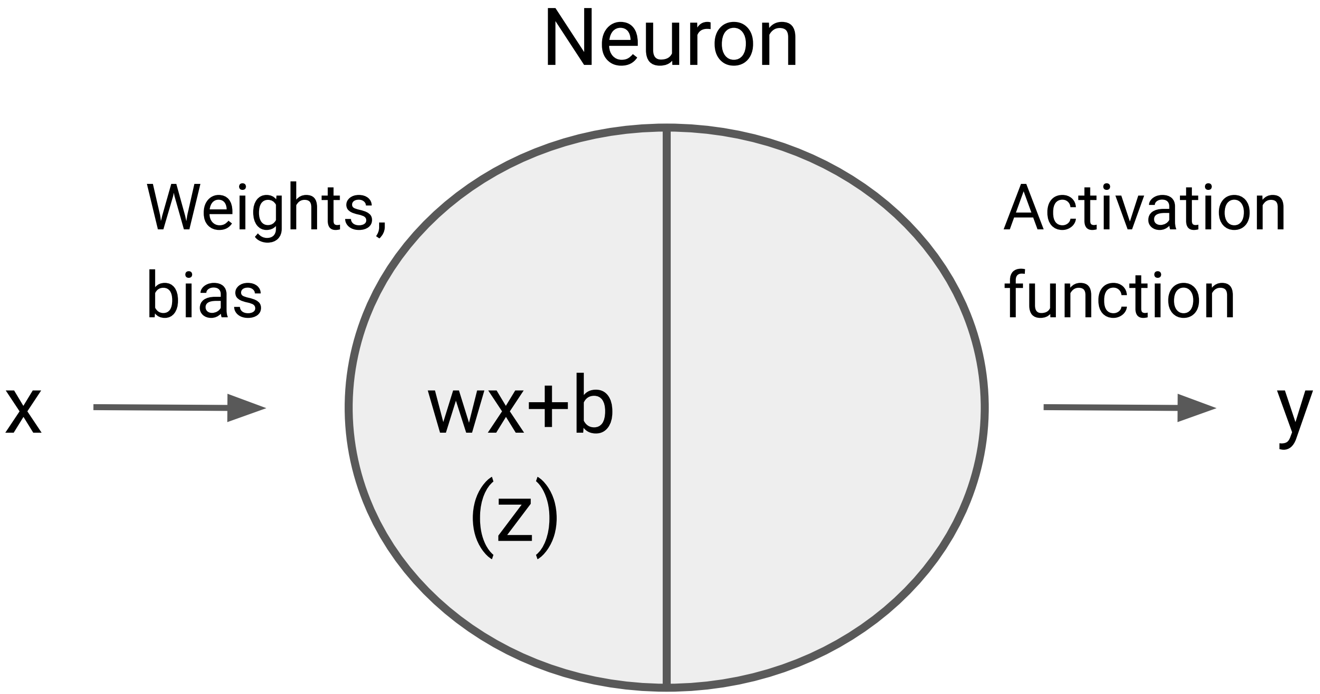
Andrew Ng, Neural Networks and Deep Learning, Coursera

Simple Neural Network



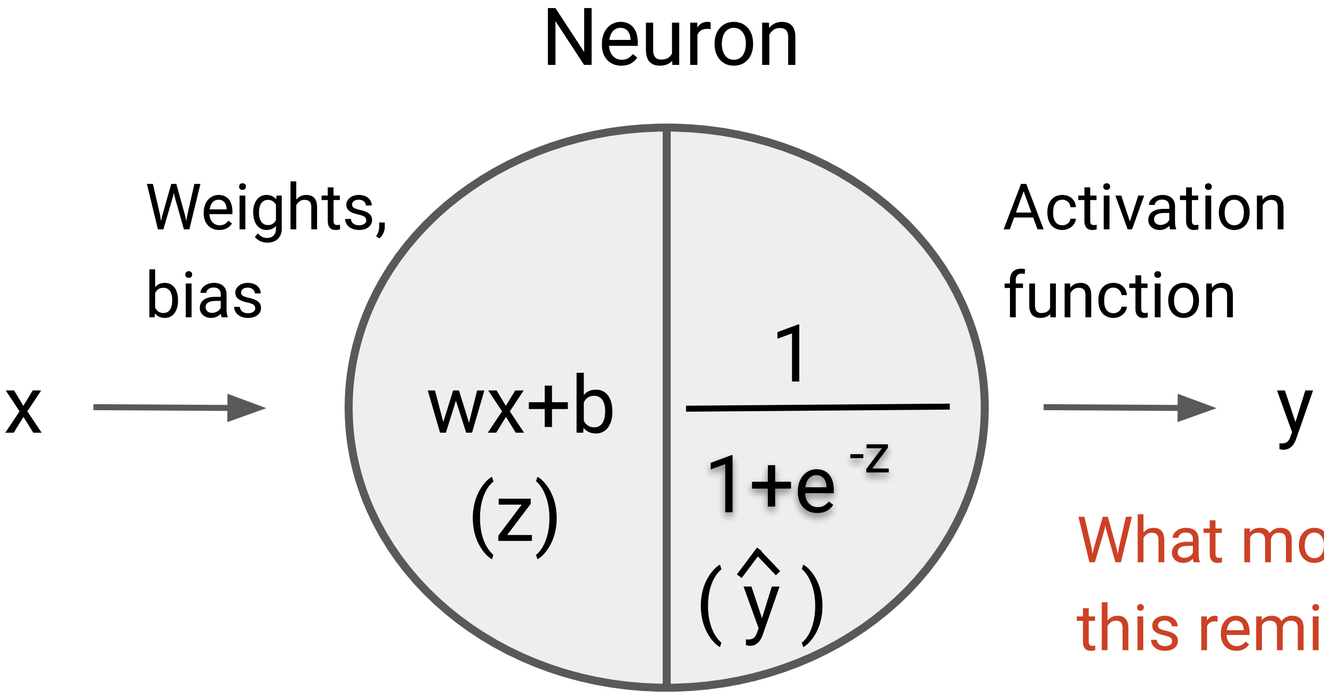
Andrew Ng, Neural Networks and Deep Learning, Coursera

Simple Neural Network



Andrew Ng, Neural Networks and Deep Learning, Coursera

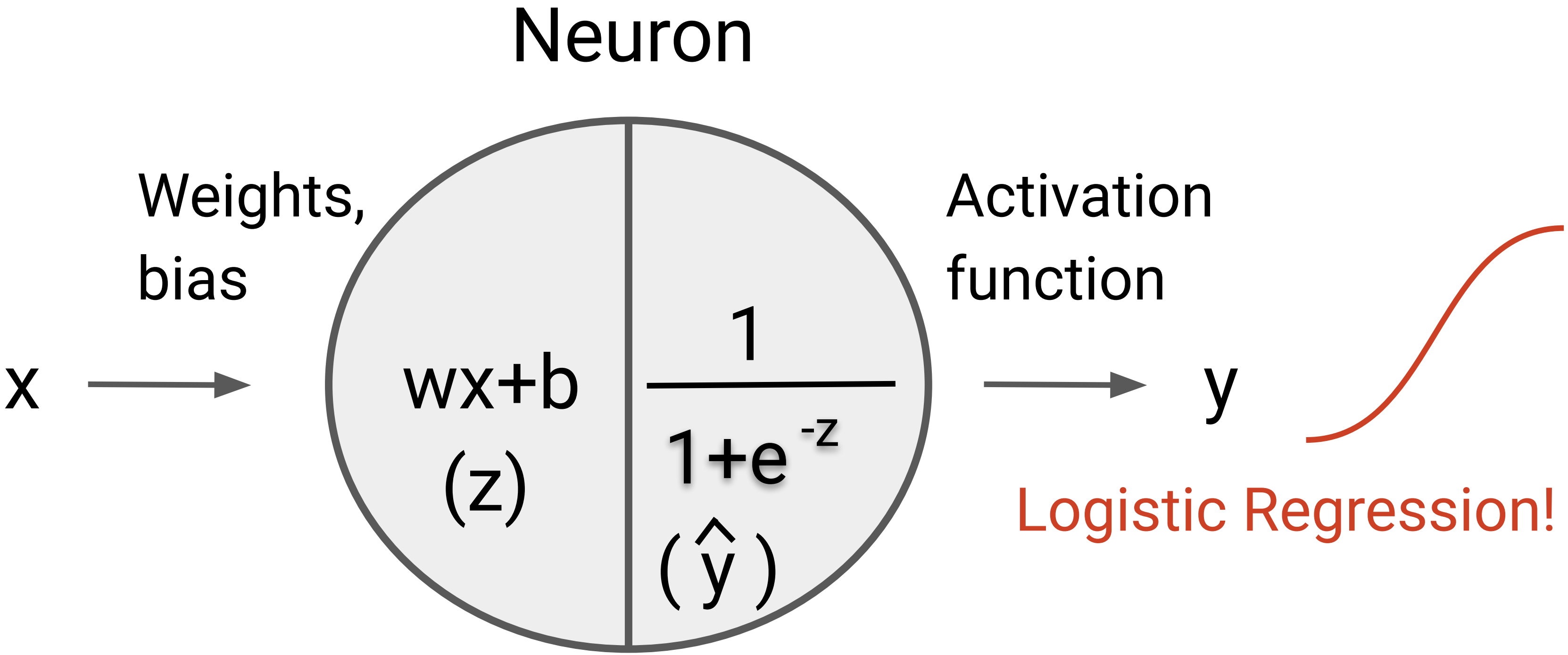
Simple Neural Network



What model does this remind you of?

Andrew Ng, Neural Networks and Deep Learning, Coursera

Simple Neural Network

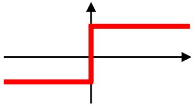
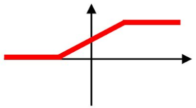
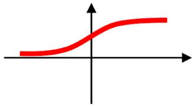
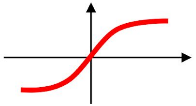
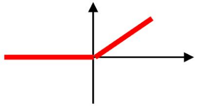


Andrew Ng, Neural Networks
and Deep Learning, Coursera

Neural Network Activation Functions

What are some good activation functions for neural networks?

Neural Network Activation Functions

Activation function	Equation	Example	1D Graph
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = max(0, z)$	Multi-layer Neural Networks	

Copyright © Sebastian Raschka 2016
(http://sebastianraschka.com)

Neural Network Forward Pass

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine
x	$z=wx+b$	$\hat{y}=\dfrac{1}{1+e^{-z}}$	y

Neural Network Forward Pass

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine
x	$z=wx+b$	$\hat{y}=\dfrac{1}{1+e^{-z}}$	y
$x=.12$	$w=15$, $b=-.4$ random initial values	Find $\hat{y}$	

Neural Network Forward Pass

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y
$x=.12$	$w=15, b=-.4$	$\hat{y}=.80$	
	random initial values		

Neural Network Forward Pass

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y
$x=.12$	$w=15, b=-.4$ random initial values	$\hat{y}=.80$	If $y = 1$, what was the error in the predict?

Logistic error function

Predicted ($\hat{y}$)	Real (y)	error	Squared error
.01	1	.99	.980
.001	1	.999	.998

Logistic error function

Predicted ($\hat{y}$)	Real (y)	error	Squared error	Negative Log error
.01	1	.99	.980	4.6
.001	1	.99	.998	6.9

$$J = \begin{cases} -\ln(\hat{y}) & \text{if } y = 1 \\ -\ln(1 - \hat{y}) & \text{if } y = 0 \end{cases}$$

J is the logistic regression loss (error)

Neural Network Loss Minimization

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15, b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

Neural Network Loss Minimization

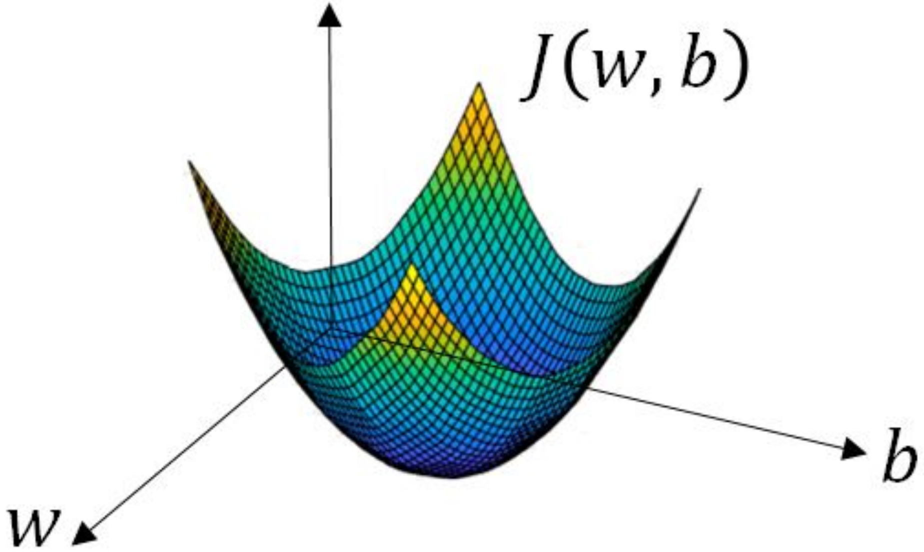
Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\dfrac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15$, $b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

How should we change *w* and *b* to minimize the error?

Neural Network Gradient Descent

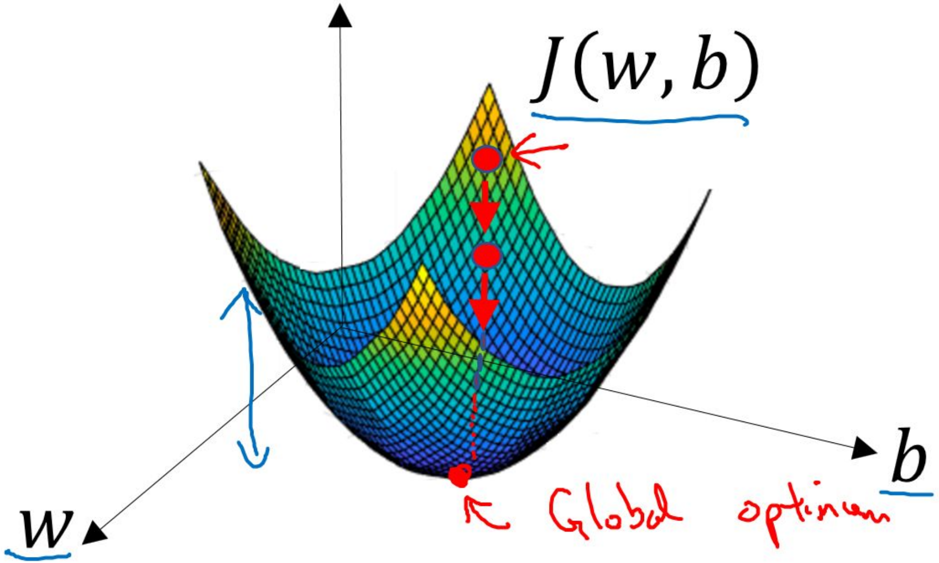
We can find the *w* and *b* that minimizes the error using gradient descent.



Andrew Ng, Neural Networks and Deep Learning, Coursera

Neural Network Gradient Descent

We can find the *w* and *b* that minimizes the error using gradient descent.



Andrew Ng, Neural Networks and Deep Learning, Coursera

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z = wx + b$	$\hat{y} = \frac{1}{1+e^{-z}}$	y	J
$x = .12$	$w = 15$, $b = -.4$	$\hat{y} = .80$	$y = 1$	$J = 1.6$

How do we calculate how changing *w* and *b* affects the error?

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15$, $b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15$, $b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}$$

$$\frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15$, $b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\hat{y}(1-\hat{y}) \qquad \frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15, b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{dw}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}\frac{dz}{dw} \qquad \frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\hat{y}(1-\hat{y}) \qquad \frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15$, $b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{dw}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}x \qquad \frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\hat{y}(1-\hat{y}) \qquad \frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15, b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{db}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}\frac{dz}{db} \qquad \frac{dJ}{dw}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}x$$

$$\frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\hat{y}(1-\hat{y}) \qquad \frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15, b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$$\frac{dJ}{db}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}1 \qquad \frac{dJ}{dw}=\frac{dJ}{d\hat{y}}\frac{d\hat{y}}{dz}x \qquad \frac{dJ}{dz}=\frac{dJ}{d\hat{y}}\hat{y}(1-\hat{y}) \qquad \frac{dJ}{d\hat{y}}=-\frac{1}{\hat{y}}$$

Neural Network Gradient Descent

Simple logistic network to predict the quality of wine

alcohol	weight bias	activation	Good wine	Loss
x	$z=wx+b$	$\hat{y}=\frac{1}{1+e^{-z}}$	y	J
$x=.12$	$w=15,\ b=-.4$	$\hat{y}=.80$	$y=1$	$J=1.6$

$\frac{dJ}{db}=-.2$ $\frac{dJ}{dw}=-.024$ $\frac{dJ}{dz}=-.2$ $\frac{dJ}{d\hat{y}}=-1.25$

Neural Network Gradient Descent

old weight bias

$w=15, b=-.4$

gradients

$$\frac{dJ}{db}=-.2 \qquad \frac{dJ}{dw}=-.024$$

update rule

$$w := w - \alpha \frac{dJ}{dw} \qquad b := b - \alpha \frac{dJ}{db}$$

α is the learning rate

Neural Network Gradient Descent

old weight bias

$w=15,\ b=-.4$

gradients

$$\frac{dJ}{db} = -.2 \qquad \frac{dJ}{dw} = -.024$$

update rule

$$w := w - \alpha \frac{dJ}{dw} \qquad b := b - \alpha \frac{dJ}{db}$$

α is the learning rate

If $\alpha = .1$, what is the new *w* and *b*?

Neural Network Gradient Descent

old weight bias

$w=15, b=-.4$

gradients

$\frac{dJ}{db}=-.2 \quad \frac{dJ}{dw}=-.024$

update rule

$w:= w - \alpha \frac{dJ}{dw} \qquad b:= b - \alpha \frac{dJ}{db}$

α is the learning rate

new weight bias

$w=14.9976, b=-.38$

Summary

- Neural networks create advanced features to make predictions
- A forward pass of a neural network require weight, bias, and an activation function
- Neural networks minimizes the error in the prediction by using gradient descent to change the weights and biases