Introduction

To Deep Learning

Materials at https://github.com/dansbecker/ODSC-Intro-to-DL-Workshop

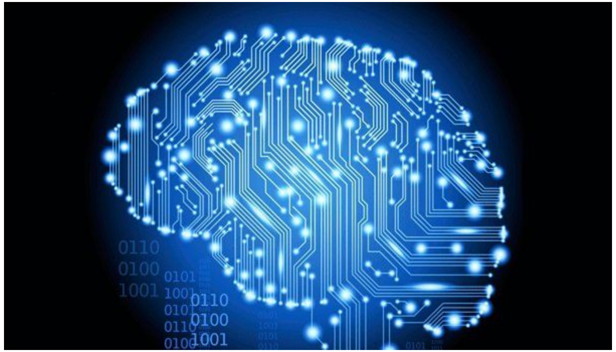
Quick Survey

Topics

- Key Concepts
- The Deep Learning Landscape
- Your First Model in Keras
- Computer Vision and Convolutional Networks
- Transfer Learning

What About The Sexy Stuff?

AlphaGo



Key Concepts

The Problem With Linear (GLM) Models

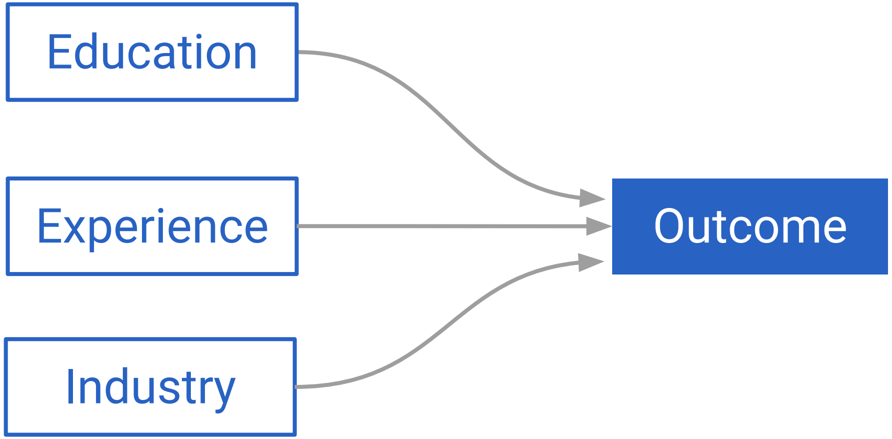
What's it buying us?

Weights

```
= ___ + ___ * years of education
      + ___ * years of experience
      + ___ * geographic indicators
      + ___ * industry indicators
      + …
```

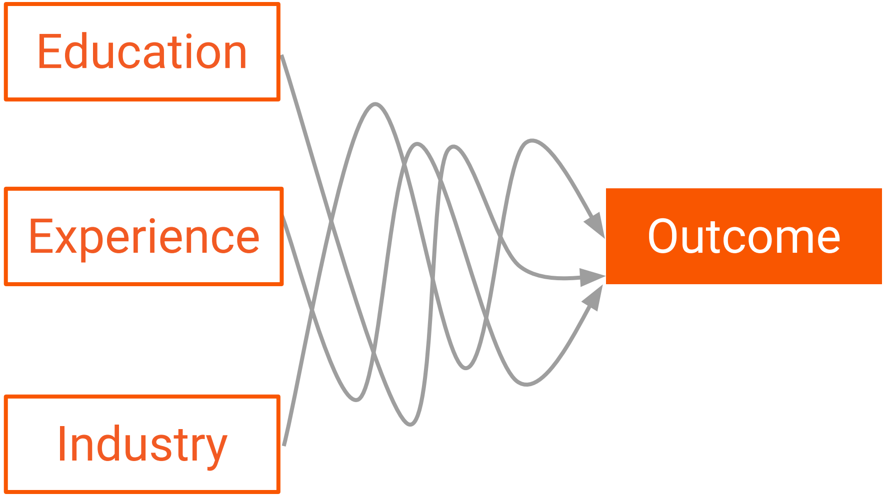
Why We Need Machine Learning

What the Linear Model Captures

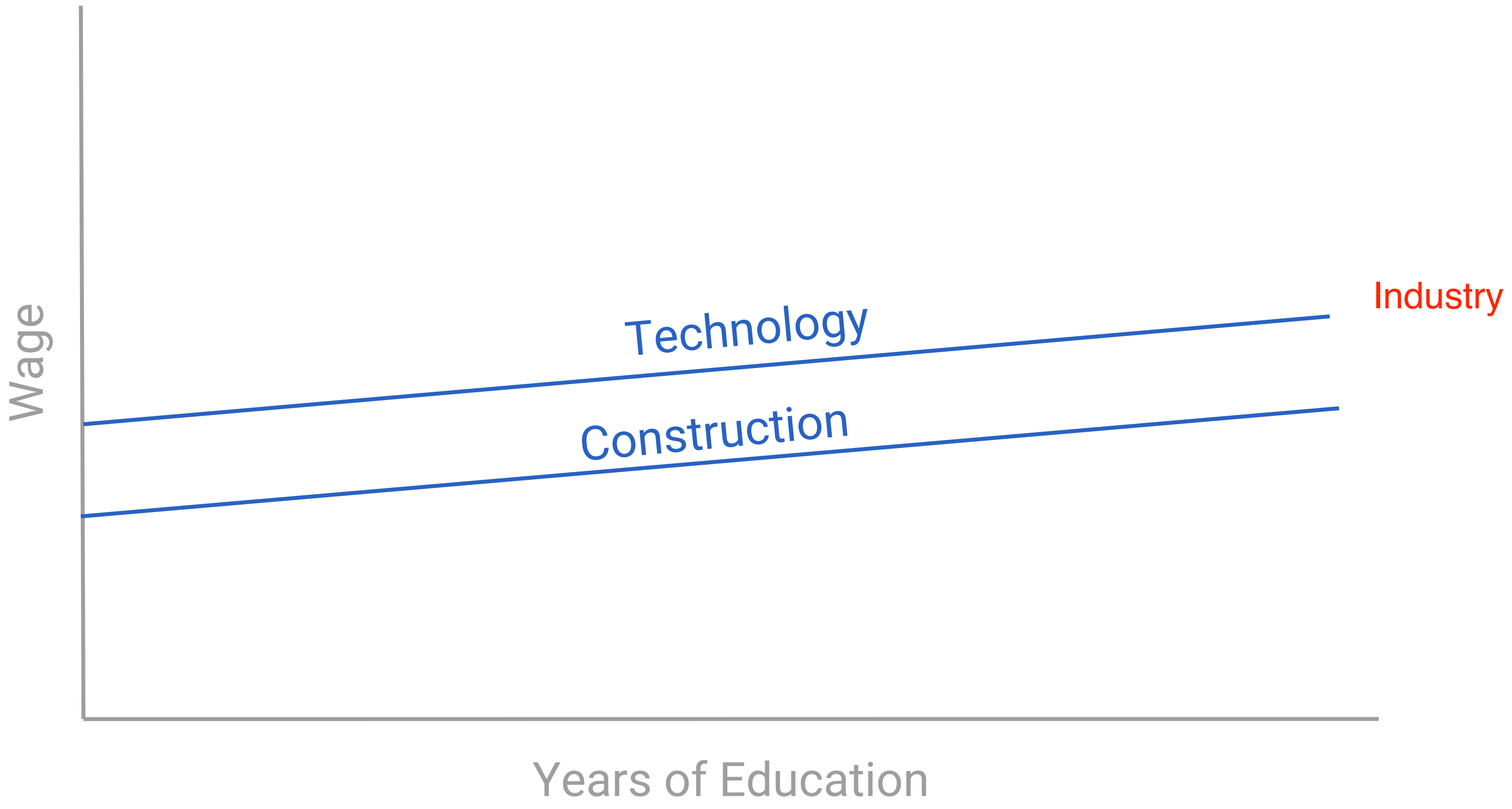


Reality

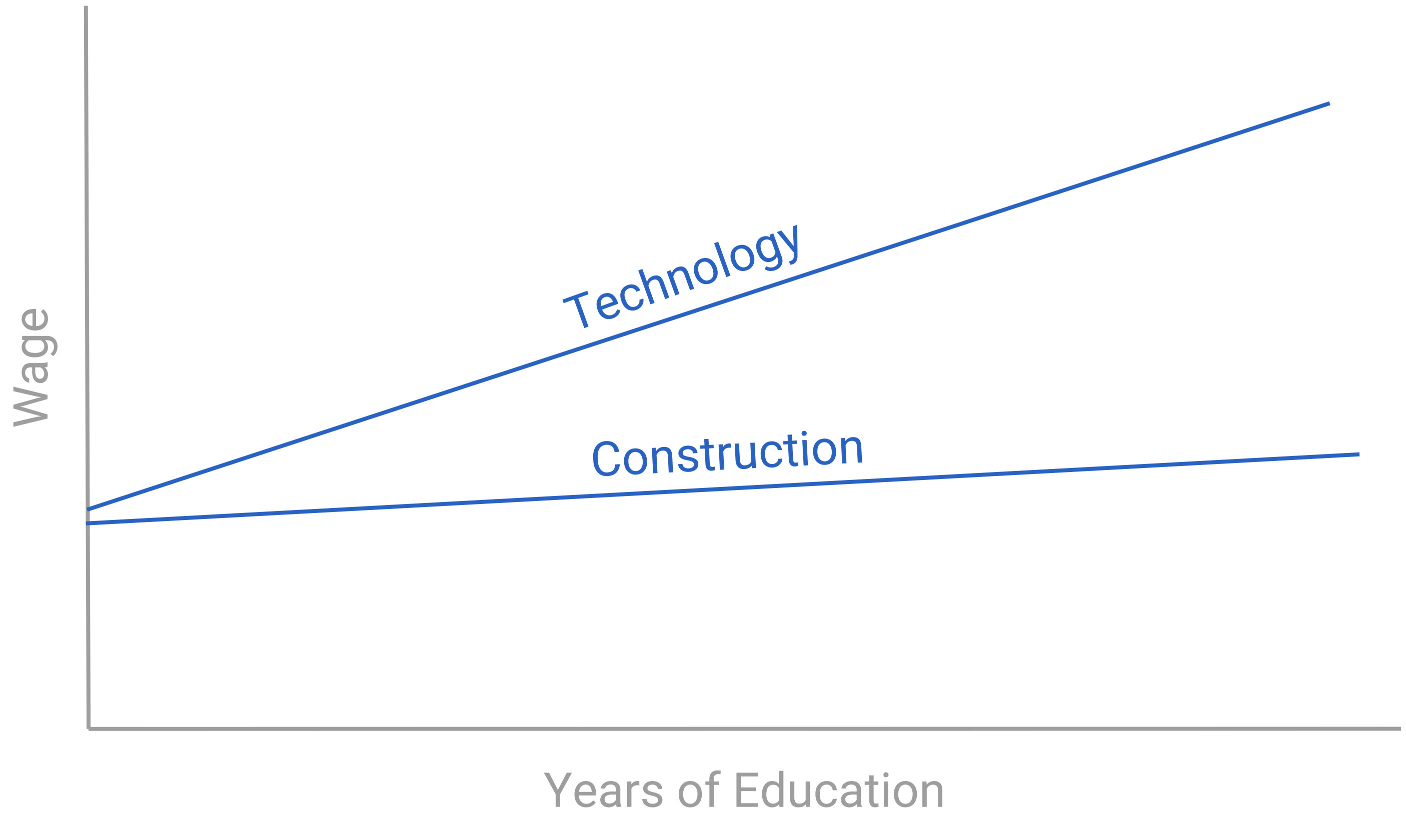
Reality is a lot more messy, shit interacts



No Interactions

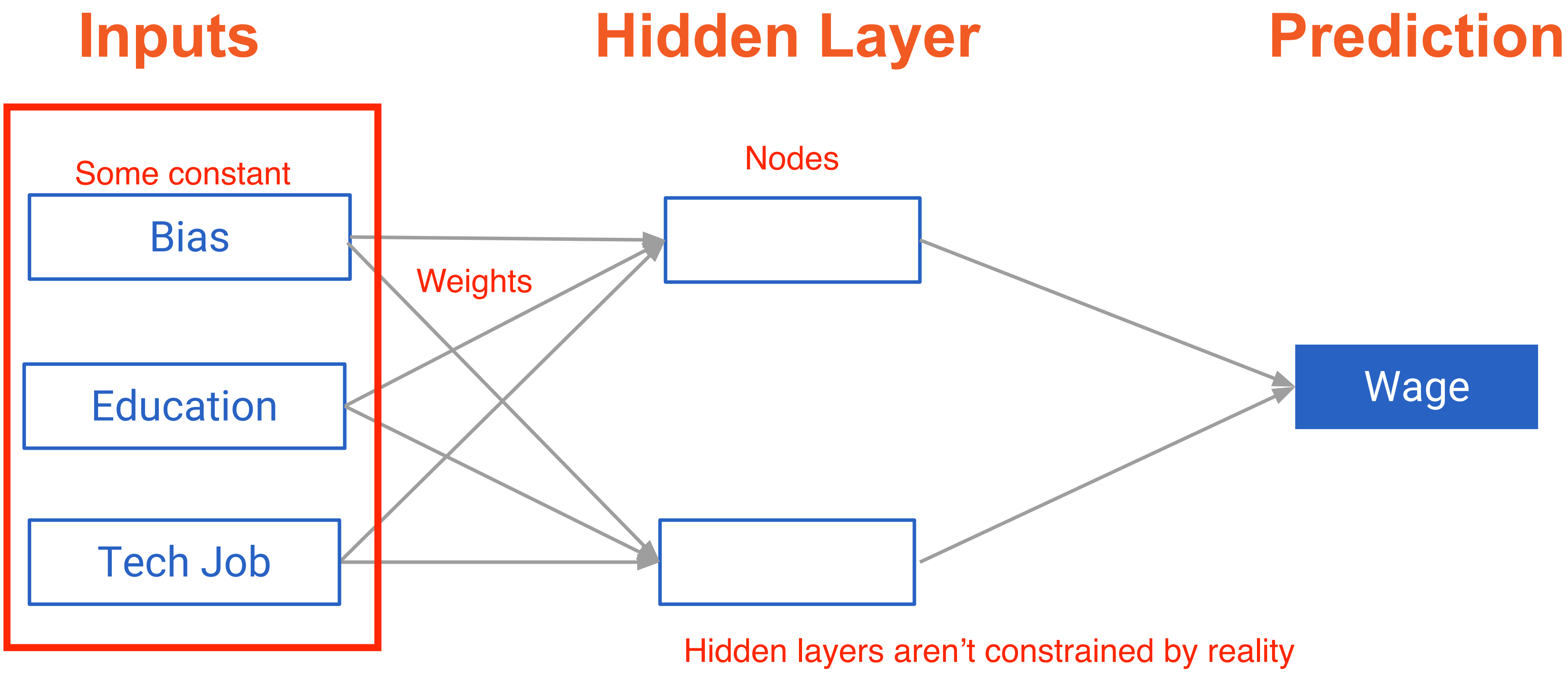


Accounting for Interactions

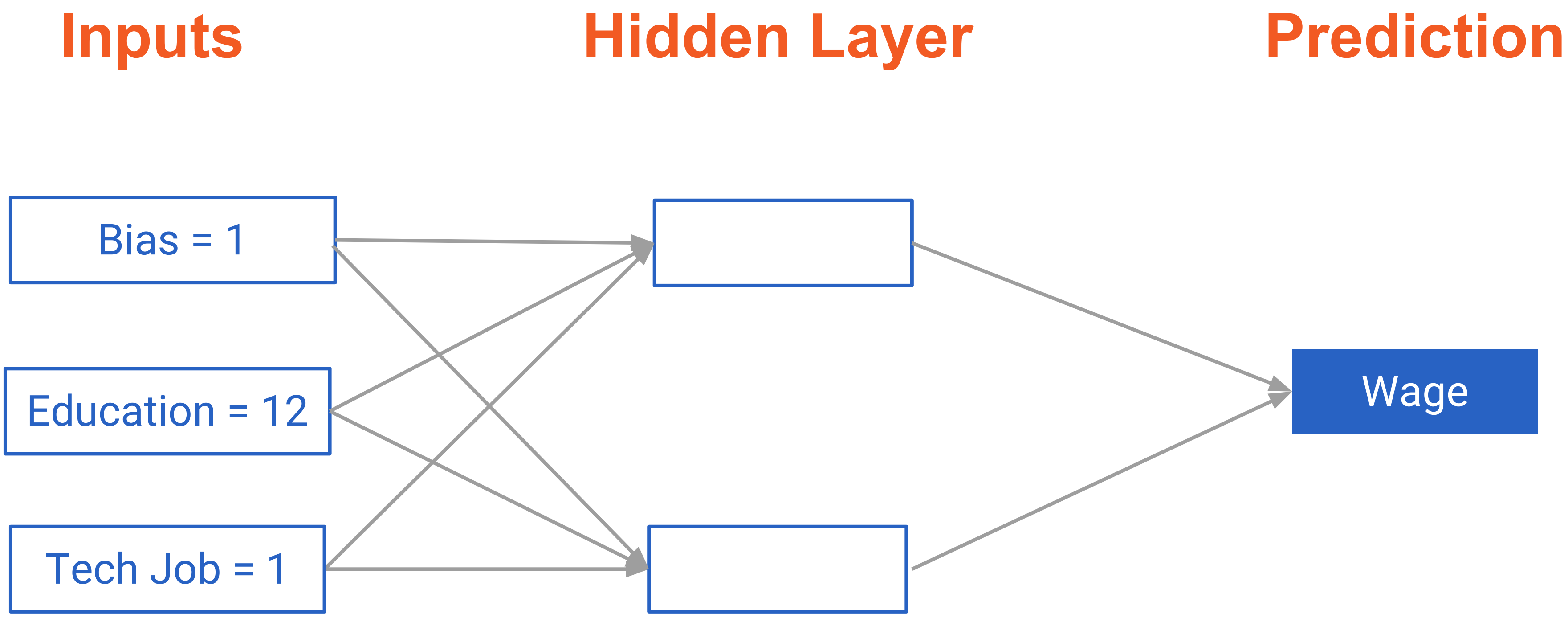


Forward Propagation

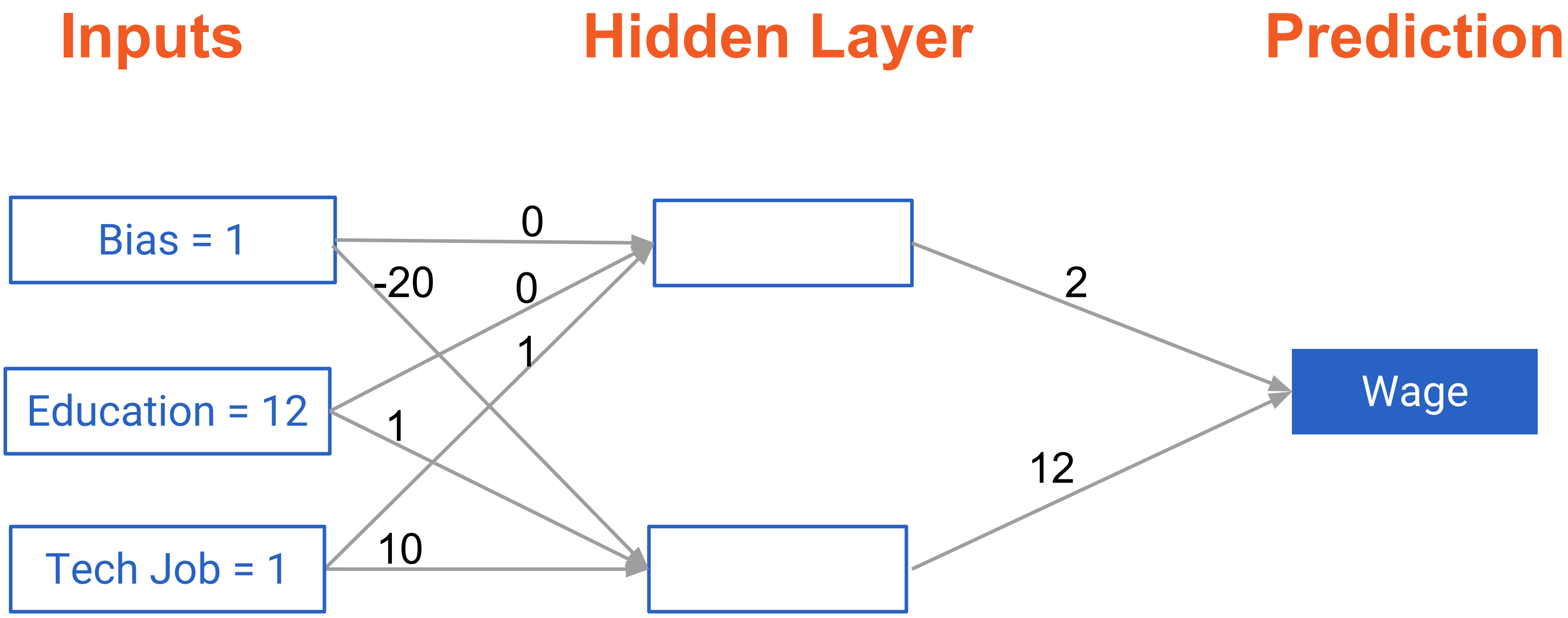
Toy Example



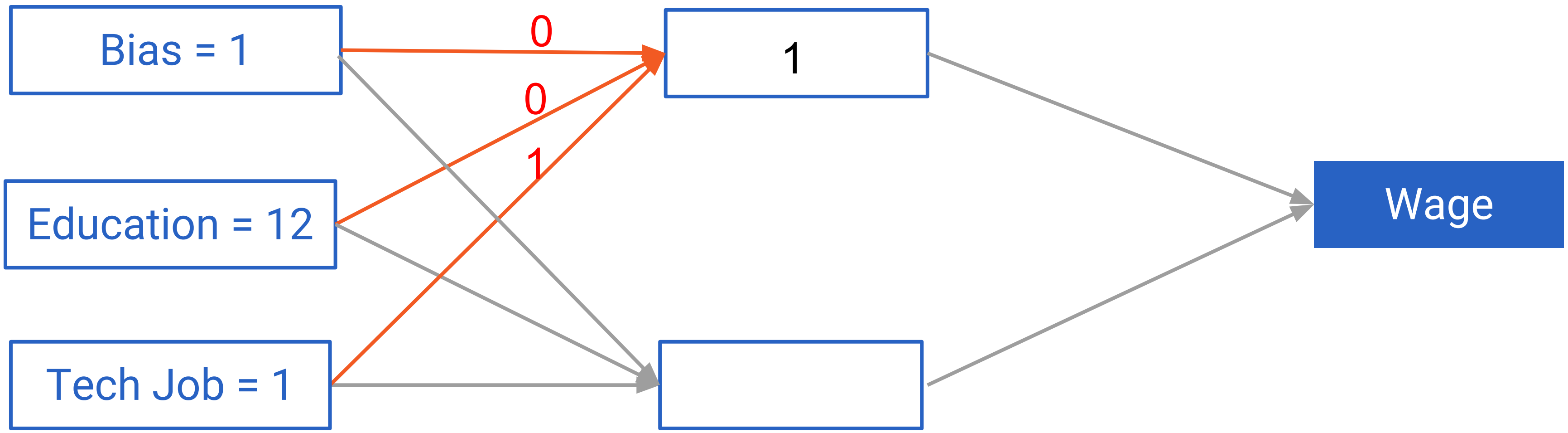
Forward Propagation



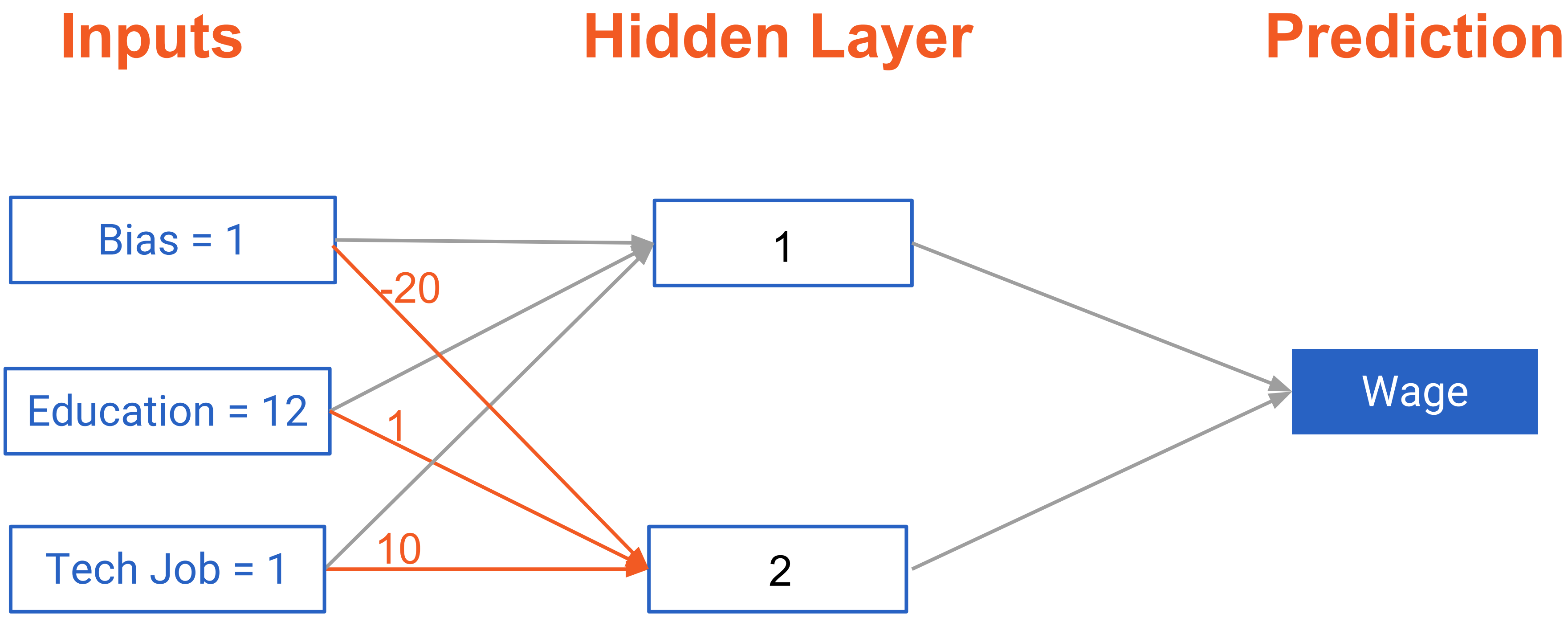
Forward Propagation



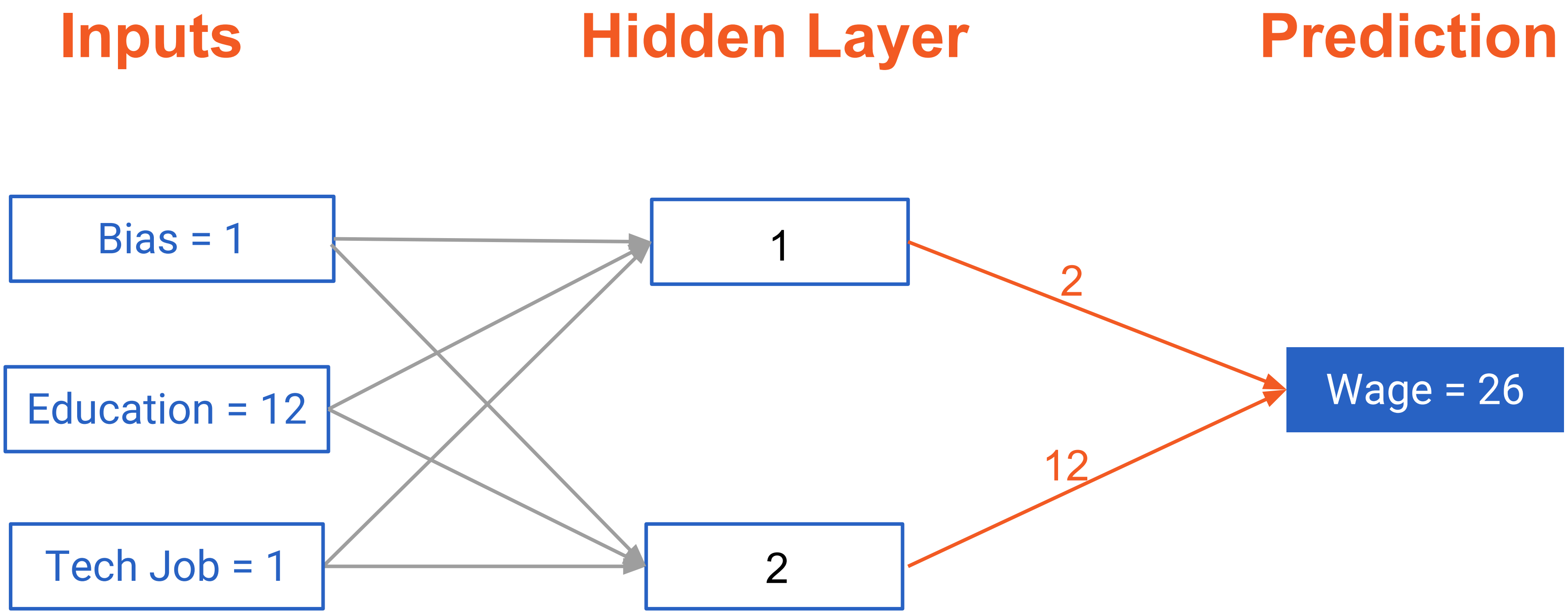
Forward Propagation



Forward Propagation



Forward Propagation

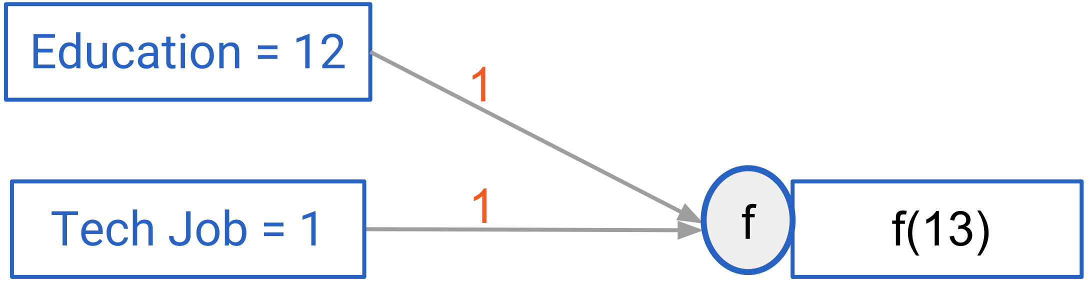


Activation Functions

Non-Linear Function Converting Node Input to Output

What

- Non-linear function converting node input to output



Activation Functions

Non-Linear Function Converting Node Input to Output

What

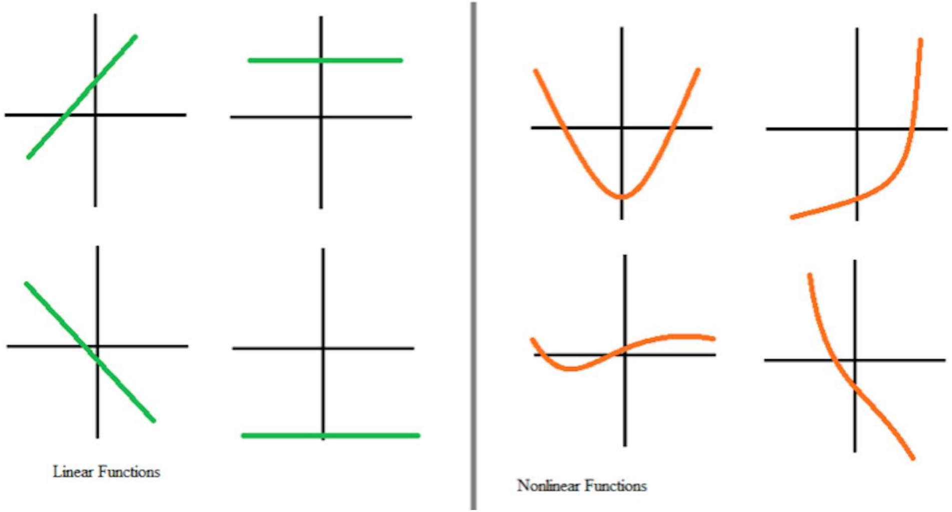
- Non-linear function converting node input to output

Why

- Account for non-linearities
- Improve ability to capture interactions

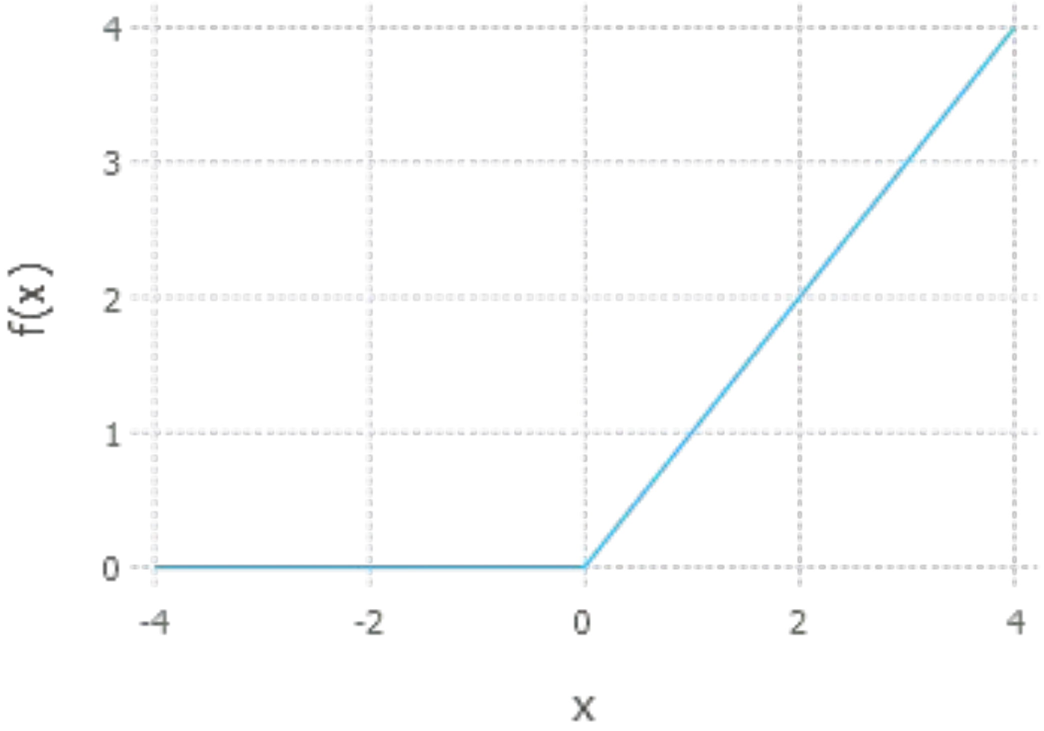
These aren't helpful

Not super critical….
just need the right weights



The ReLU Activation Function

Rectified Linear Activation Function



$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

Actually pretty powerful if there's a fuck ton of layers

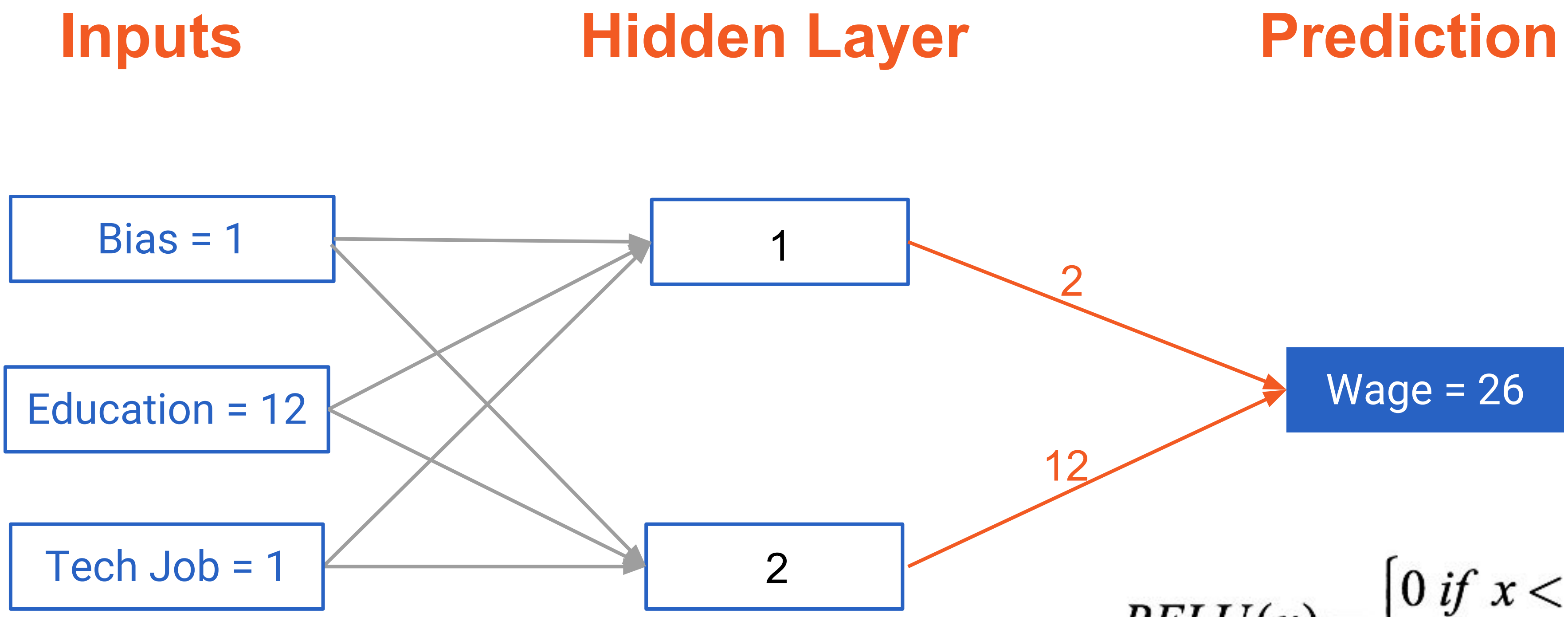
Return to Interactions

<u>Checking For Interactions</u>

- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

	Tech Job = 0	Tech Job = 1
Education = 12		
Education = 13		

Forward Propagation (with ReLU)



$$RELU(x) = \begin{cases} 0 \text{ } if \text{ } x < 0 \\ x \text{ } if \text{ } x >= 0 \end{cases}$$

Return to Interactions

Checking For Interactions

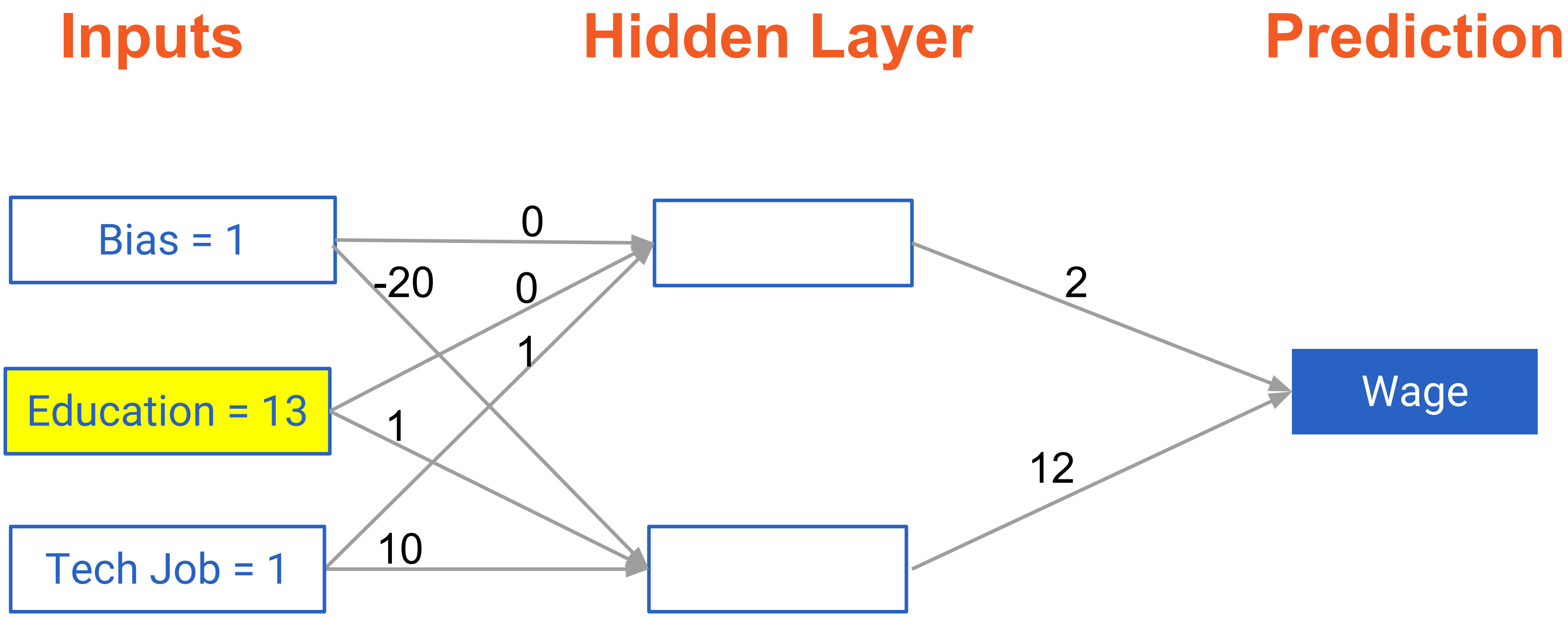
- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

Four people

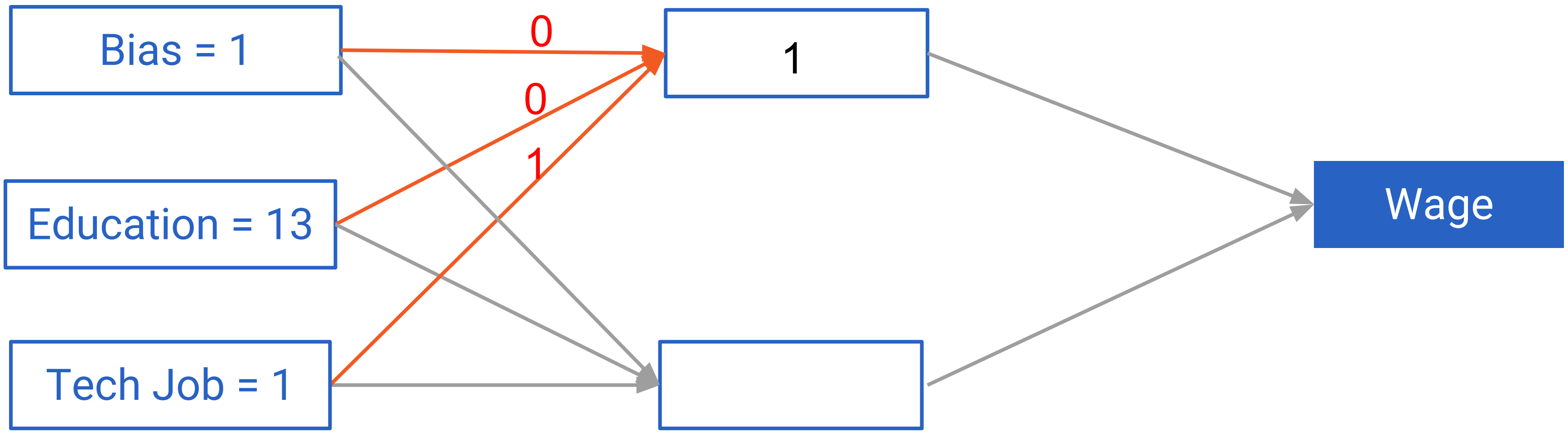
	Tech Job = 0	Tech Job = 1
Education = 12		**26**
Education = 13		

(If no interactions…
linear models would restrict on education)

Forward Propagation

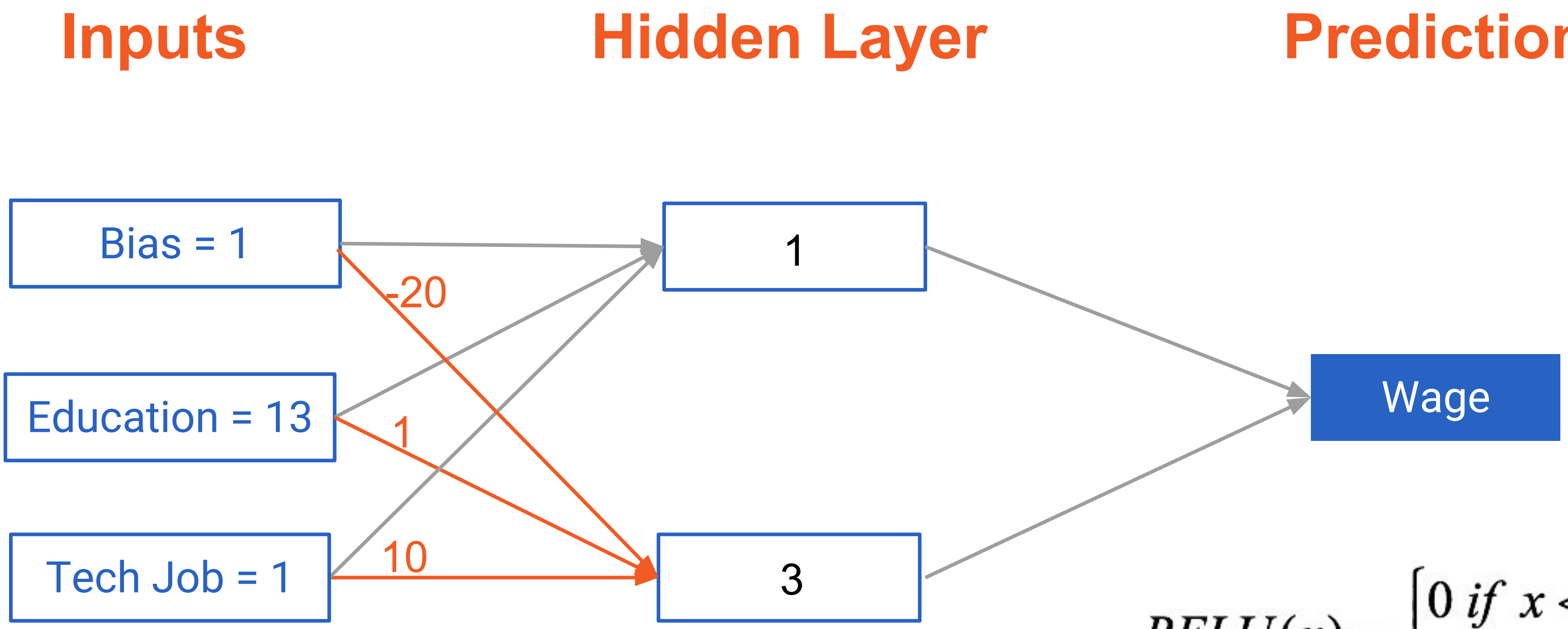


Forward Propagation



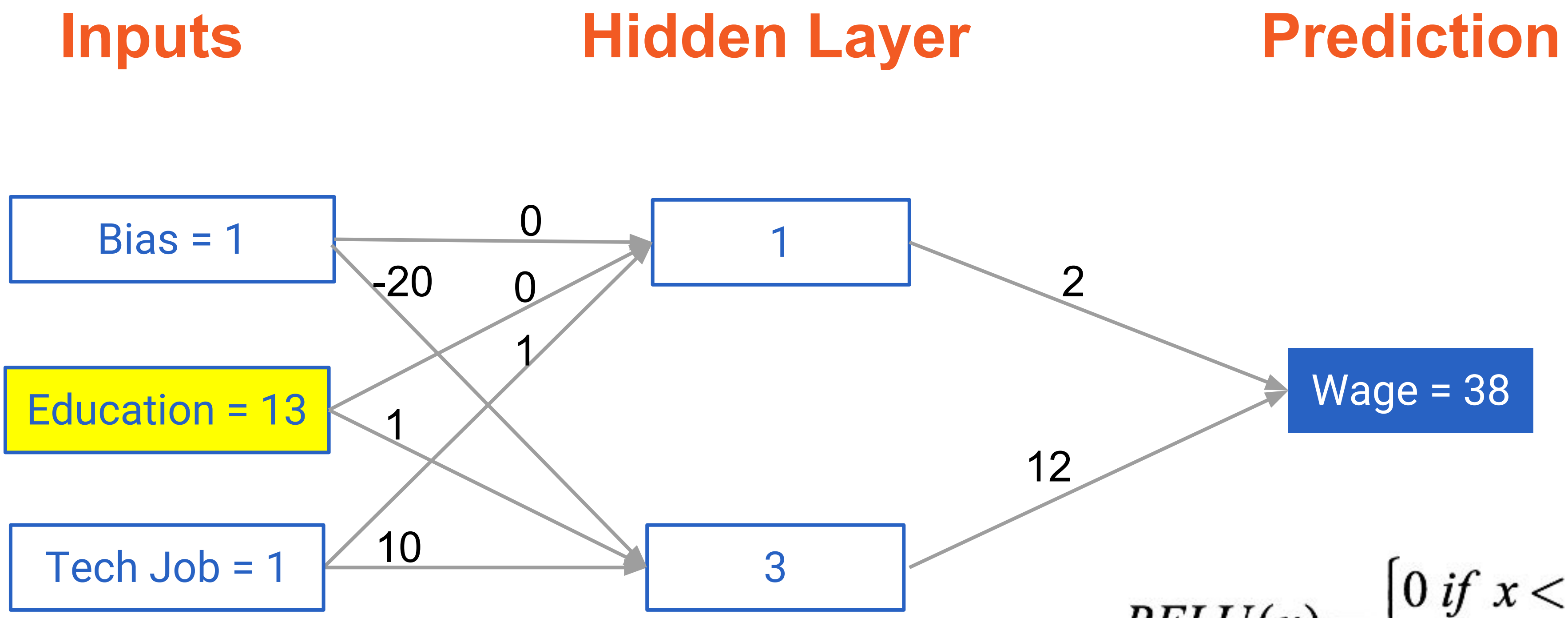
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation



$$RELU(x) = \begin{cases} 0 \text{ } if \text{ } x < 0 \\ x \text{ } if \text{ } x >= 0 \end{cases}$$

Forward Propagation



$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

Return to Interactions

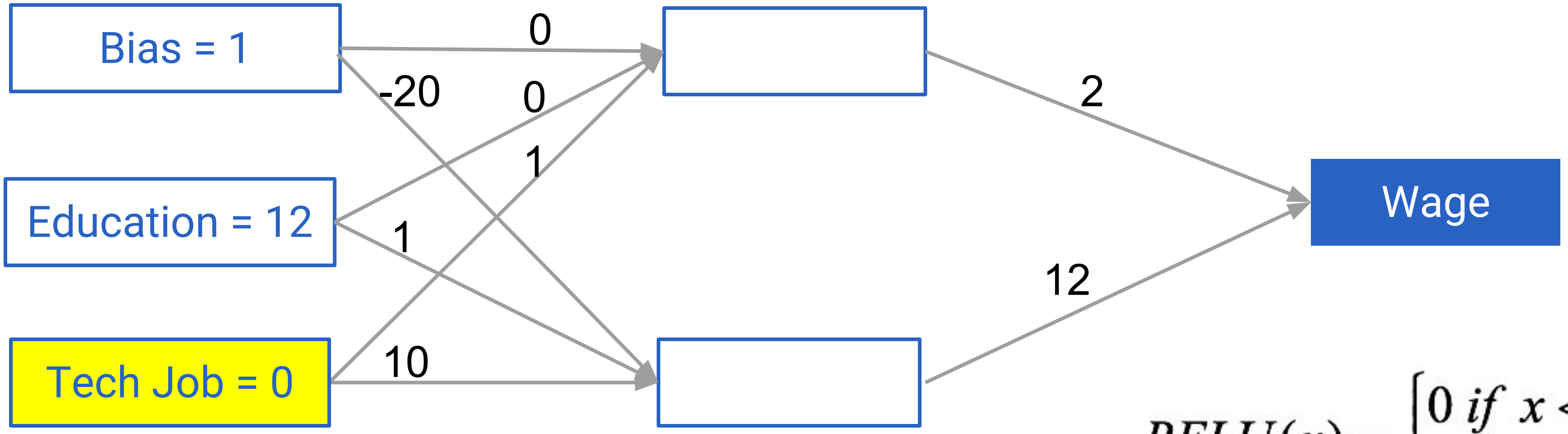
Checking For Interactions

- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

	Tech Job = 0	Tech Job = 1
Education = 12		**26**
Education = 13		**38**

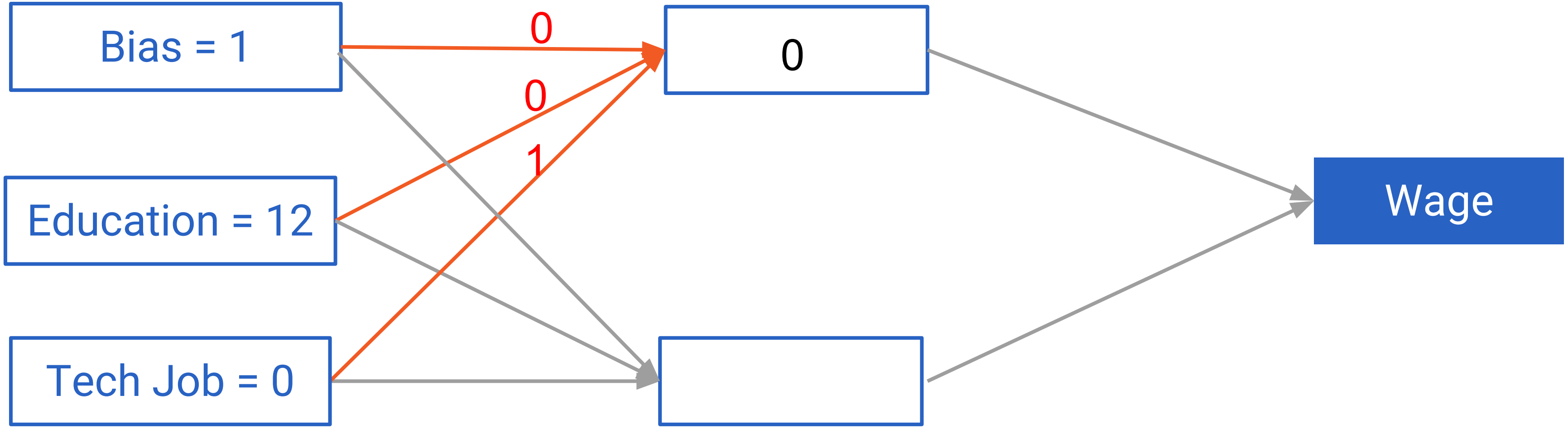
Forward Propagation (with ReLU)

Inputs **Hidden Layer** **Prediction**



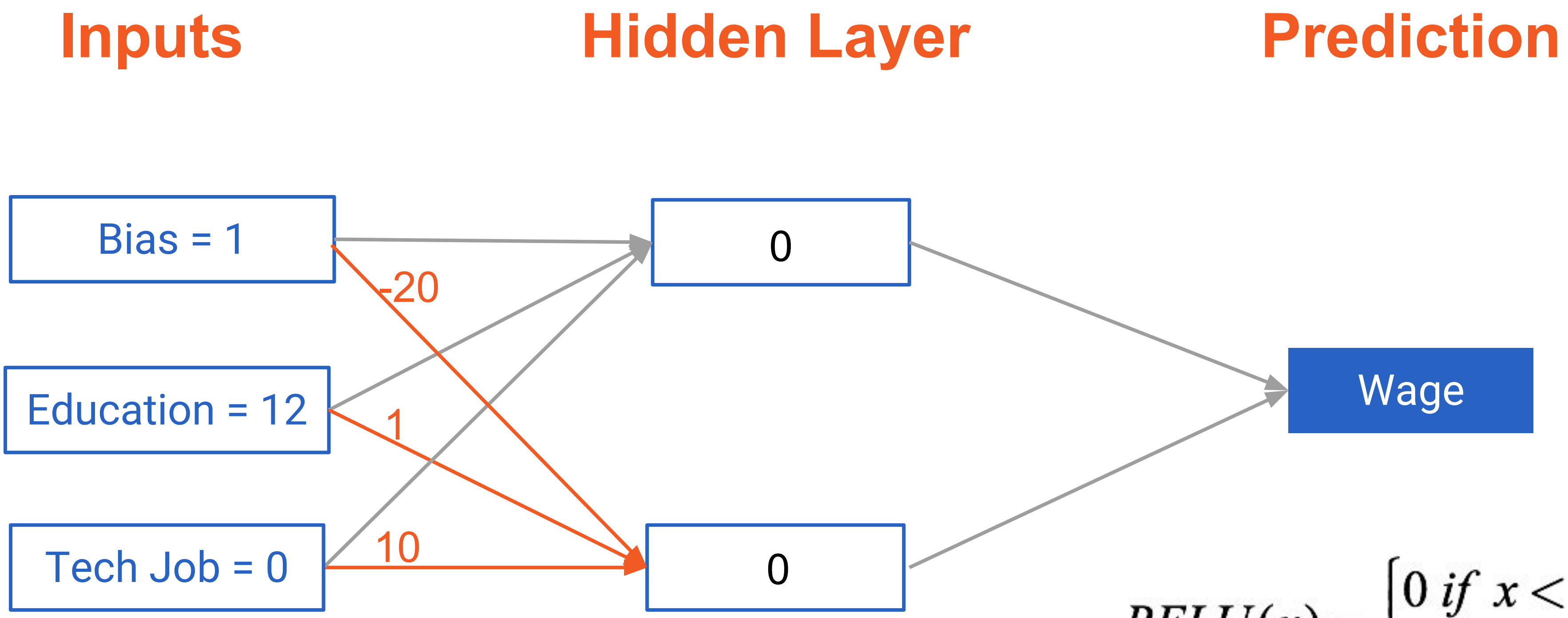
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



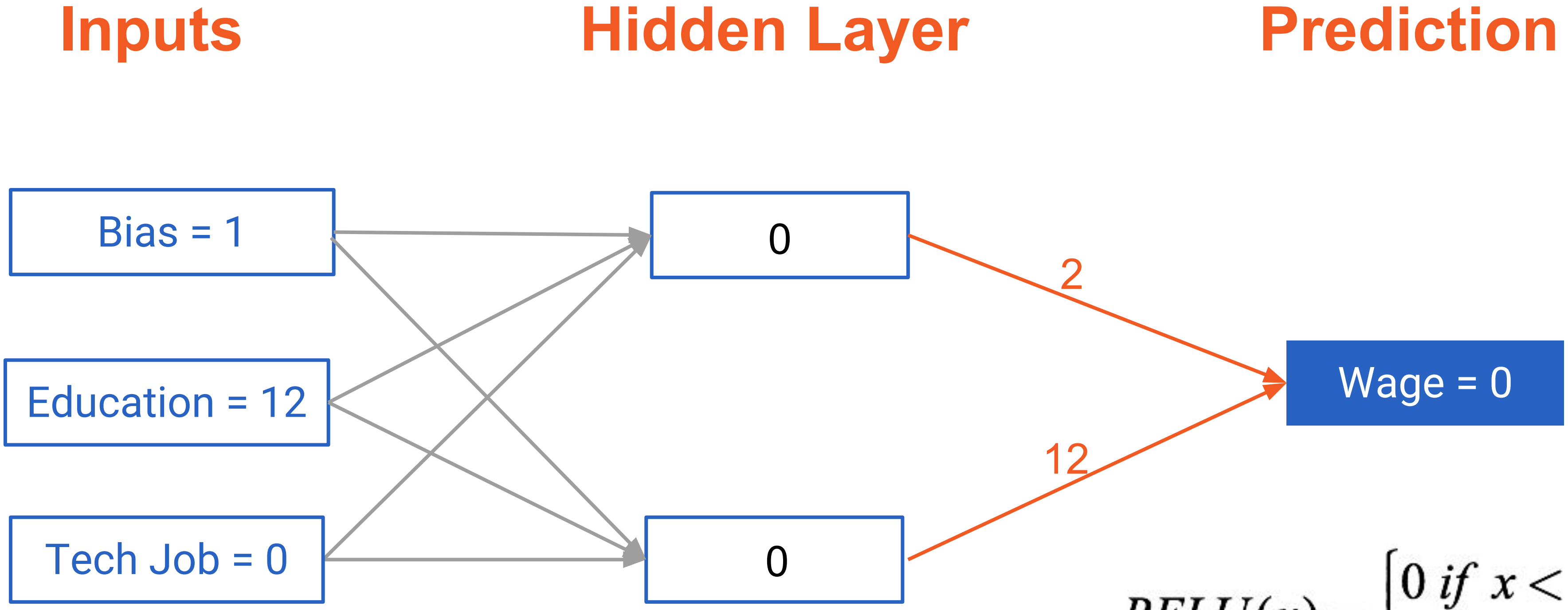
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

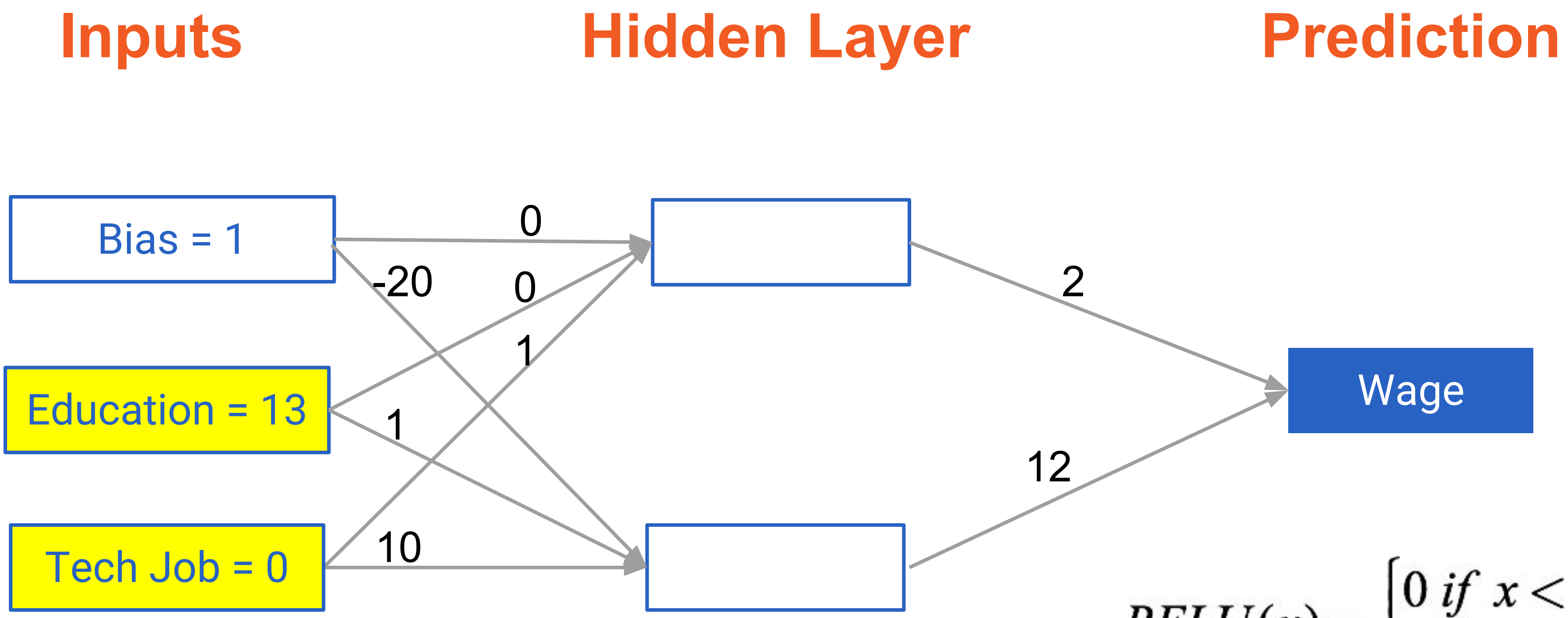
Return to Interactions

Checking For Interactions

- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

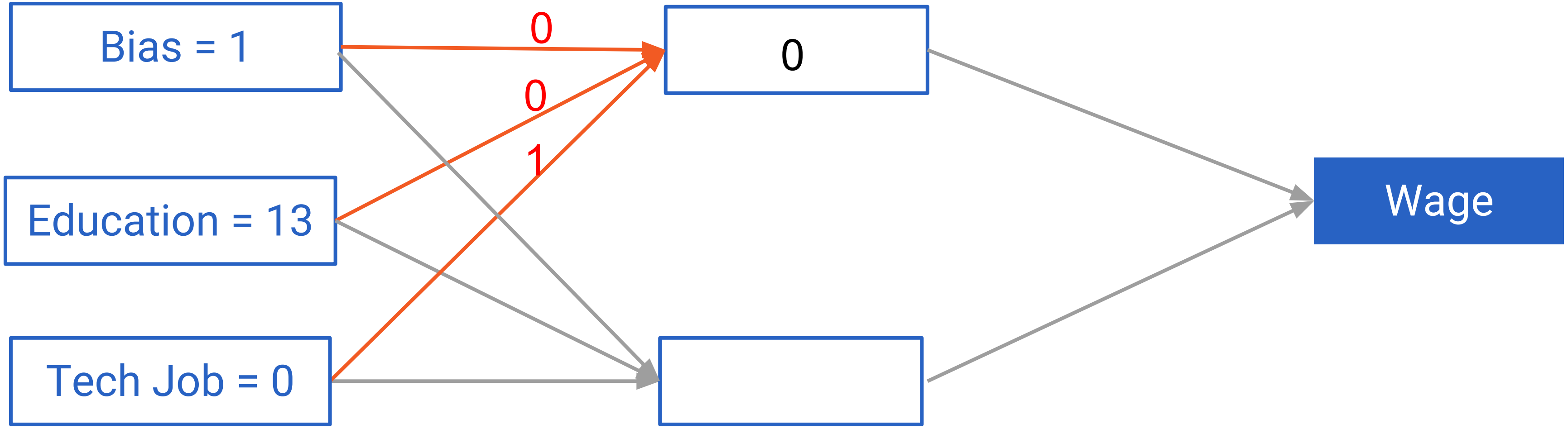
	Tech Job = 0	Tech Job = 1
Education = 12	**0**	**26**
Education = 13		**38**

Forward Propagation (with ReLU)



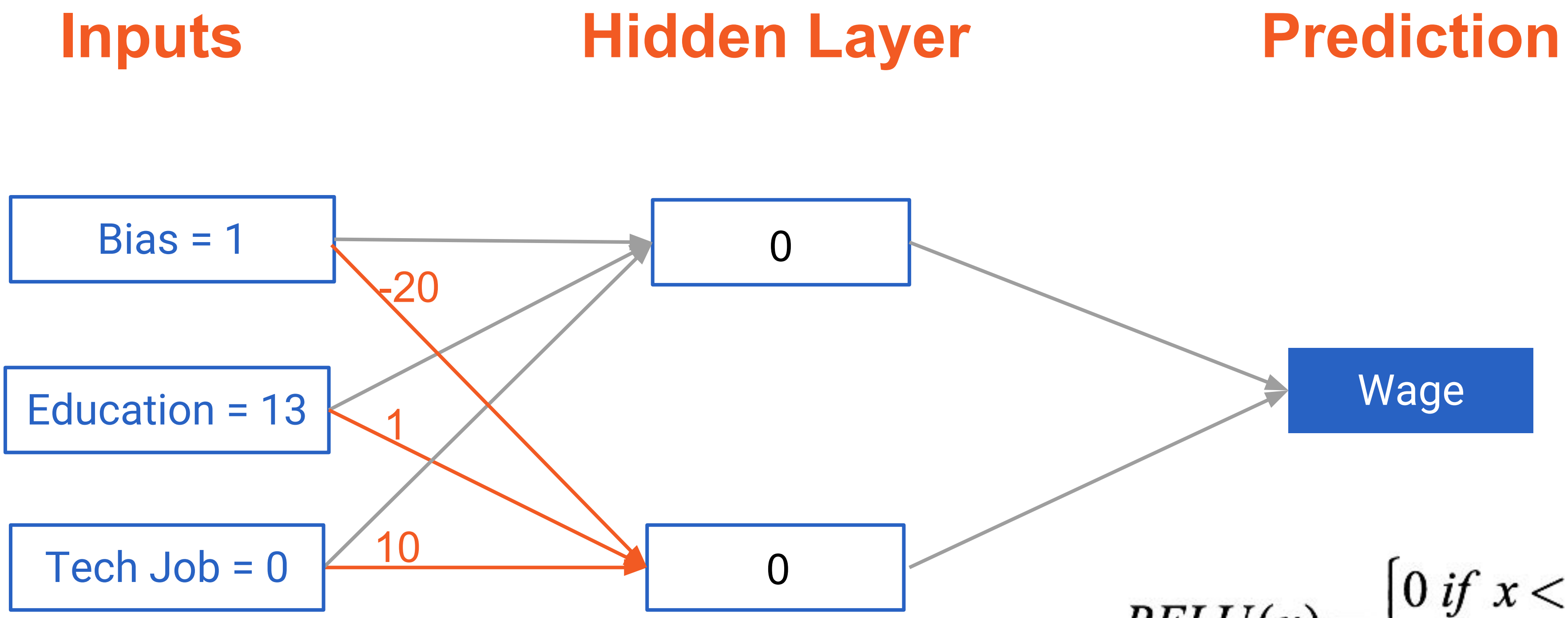
$$RELU(x) = \begin{cases} 0 \text{ if } x < 0 \\ x \text{ if } x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



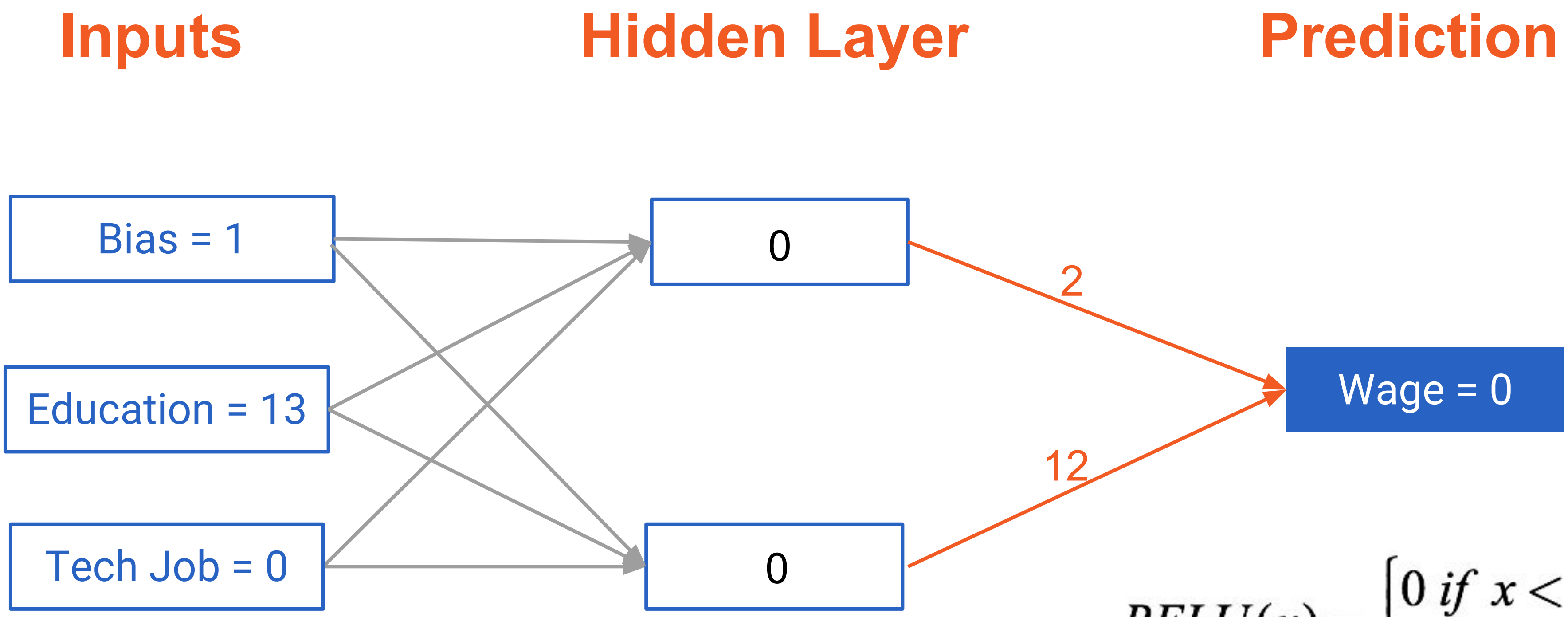
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Return to Interactions

Multiplication and addition is kinda all you need… can go more complex.

<u>Checking For Interactions</u>

- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

(Hopefully everyone is in tech…)

	Tech Job = 0	Tech Job = 1
Education = 12	**0**	**26**
Education = 13	**0**	**38**

The Takeaway

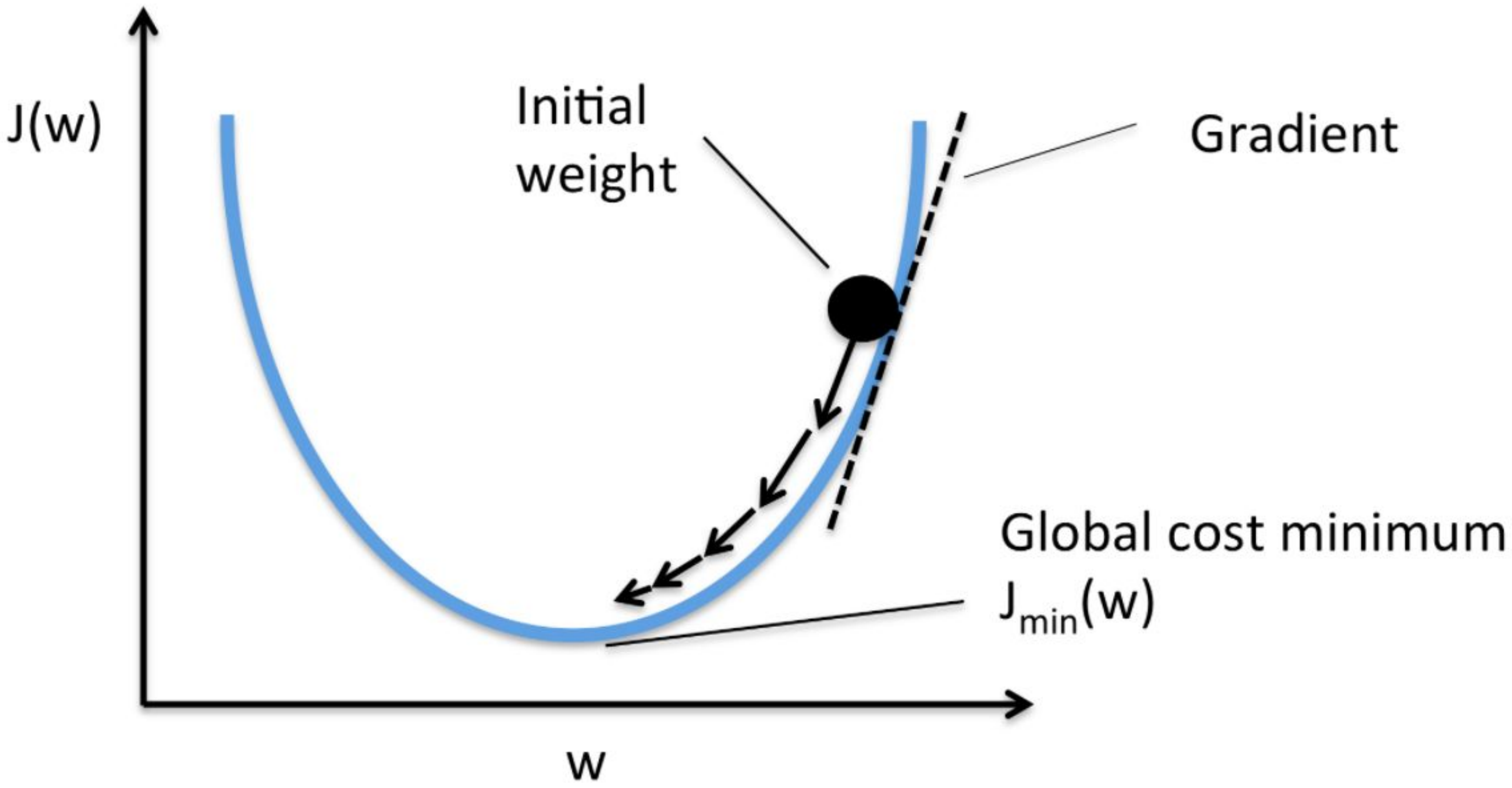
Checking For Interactions

- Neural network models capture interactions and non-linearities

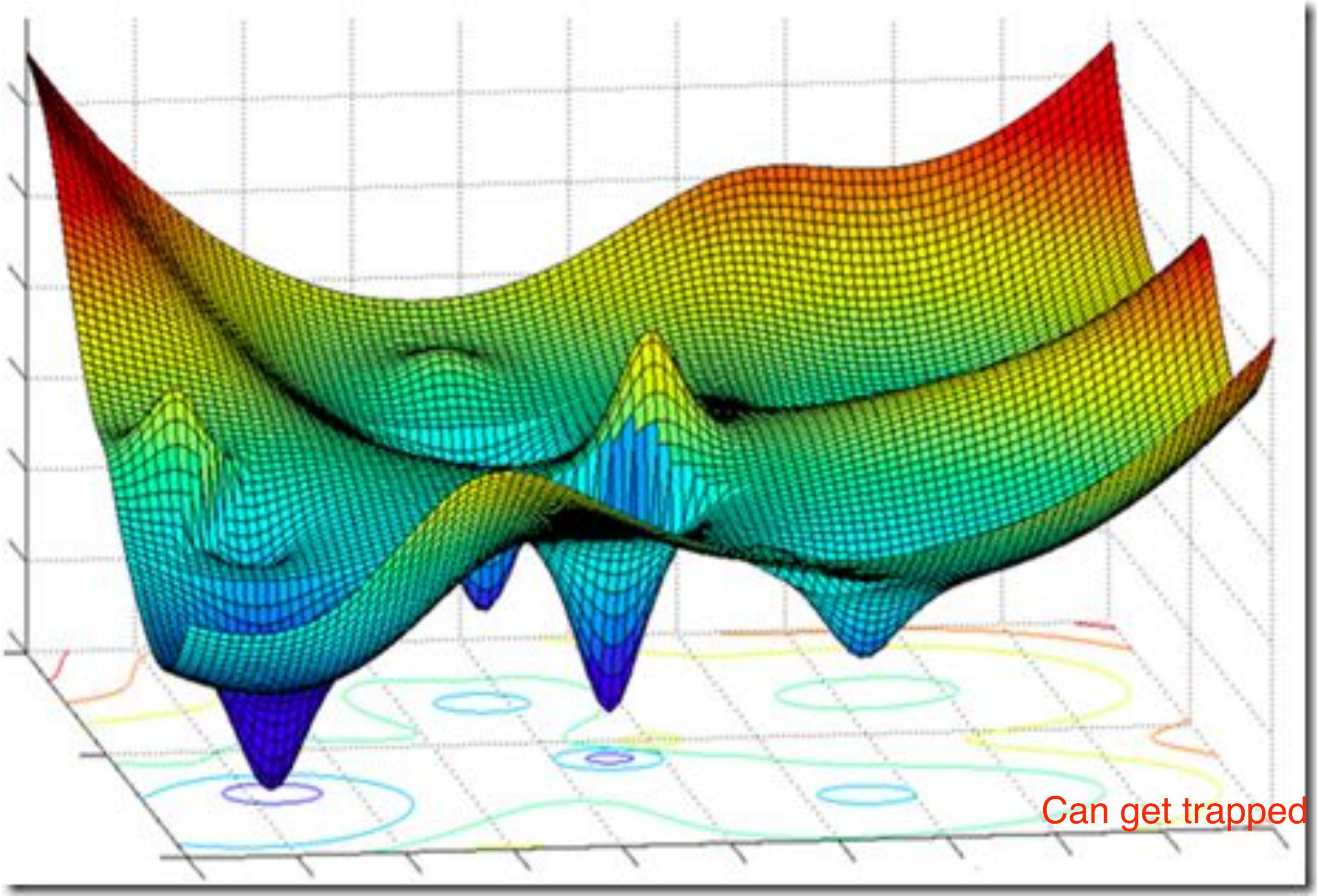
- Depending on the weights, they can still make bad predictions

Finding the right weights is the super hard part

Gradient Descent



Gradient Descent



Can get trapped in a local minimum

Gradient Descent

Repeatedly:

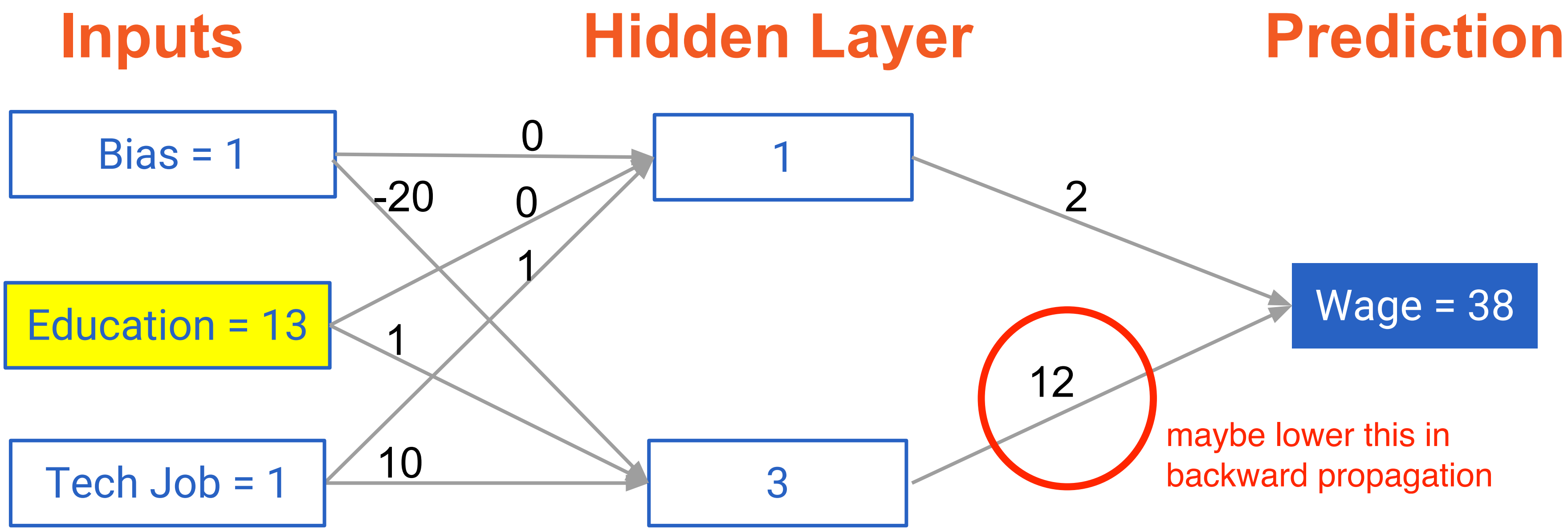
- Find derivative / slope of loss function with respect to each weight
- Take small step downhill (subtracting product of derivative and learning rate)

Back Propagation

the process to get…

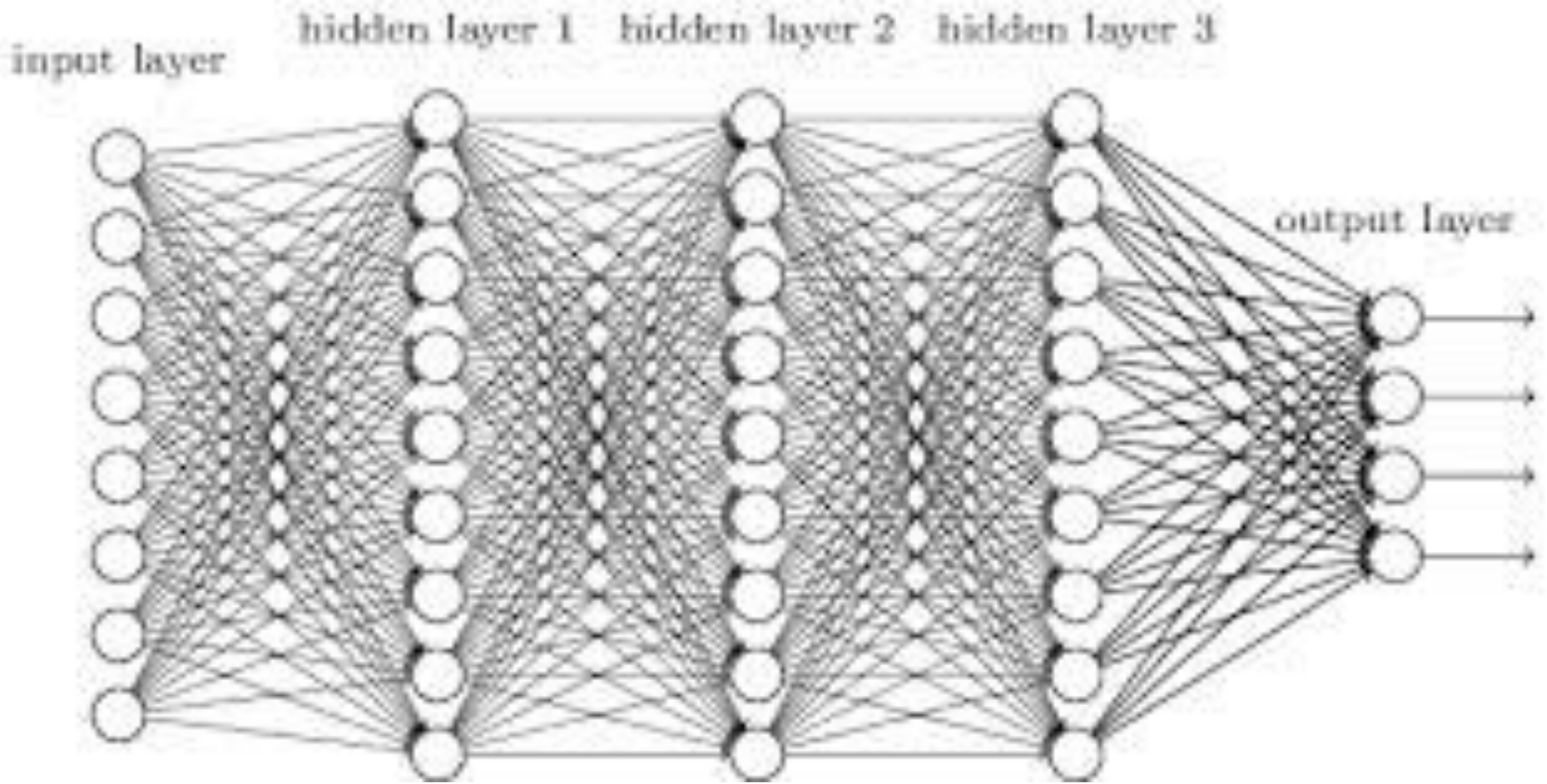
- Used to get derivatives needed to update weights
- Application of chain rule from calculus
- Used after forward propagation to find errors
- Not focusing on this math today

Backward Propagation



Actual Value = 30
Error = 8
Squared Error = 64

Deeper Networks



Topics

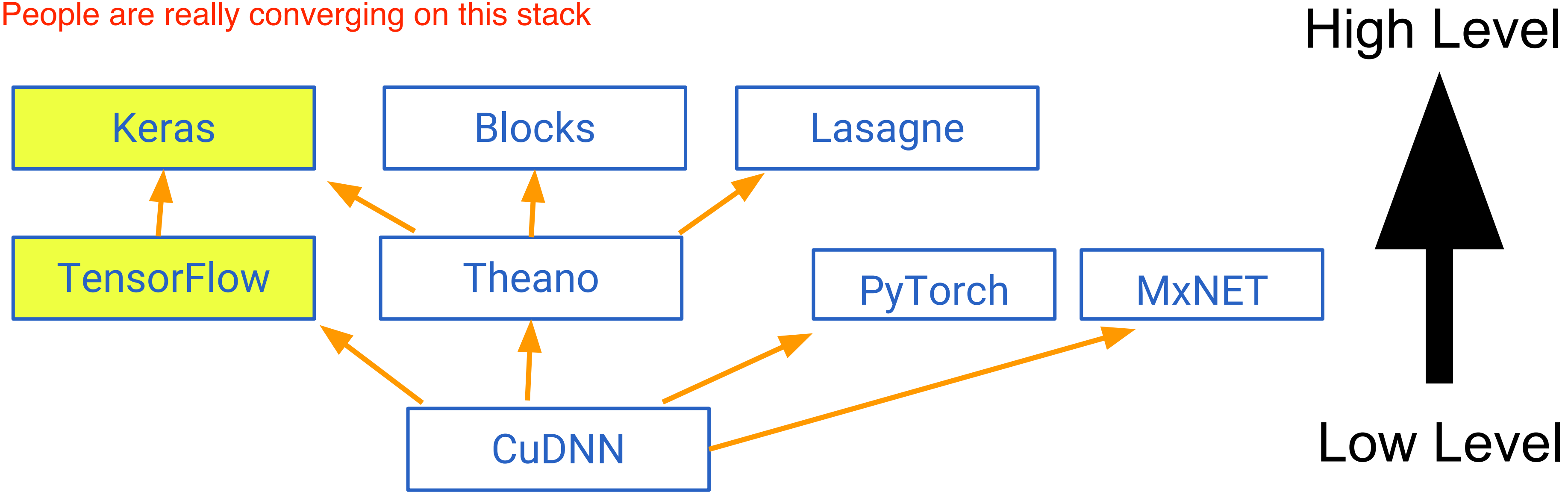
- ✓ Key Concepts
- ☐ The Deep Learning Landscape
- ☐ Your First Model in Keras
- ☐ Computer Vision and Convolutional Networks
- ☐ Transfer Learning

Where Deep Learning Shines

Good for unstructured (sound, video, text)…. but it’s funny because “unstructured is soooooo structured”, The ordering is super important, if you shuffled words, sentence loses meaning

Deep Learning Landscape

People are really converging on this stack



Many more libraries available than listed here

Topics

- ✓ Key Concepts
- ✓ The Deep Learning Landscape
- ◻ Your First Model in Keras
- ◻ Computer Vision and Convolutional Networks
- ◻ Transfer Learning

The Keras Workflow

Similar to XGBoost or sklearn

- Define Topology (nodes, layers)
- Compile Optimization steps for Gradient Descent
- Fit
- Predict

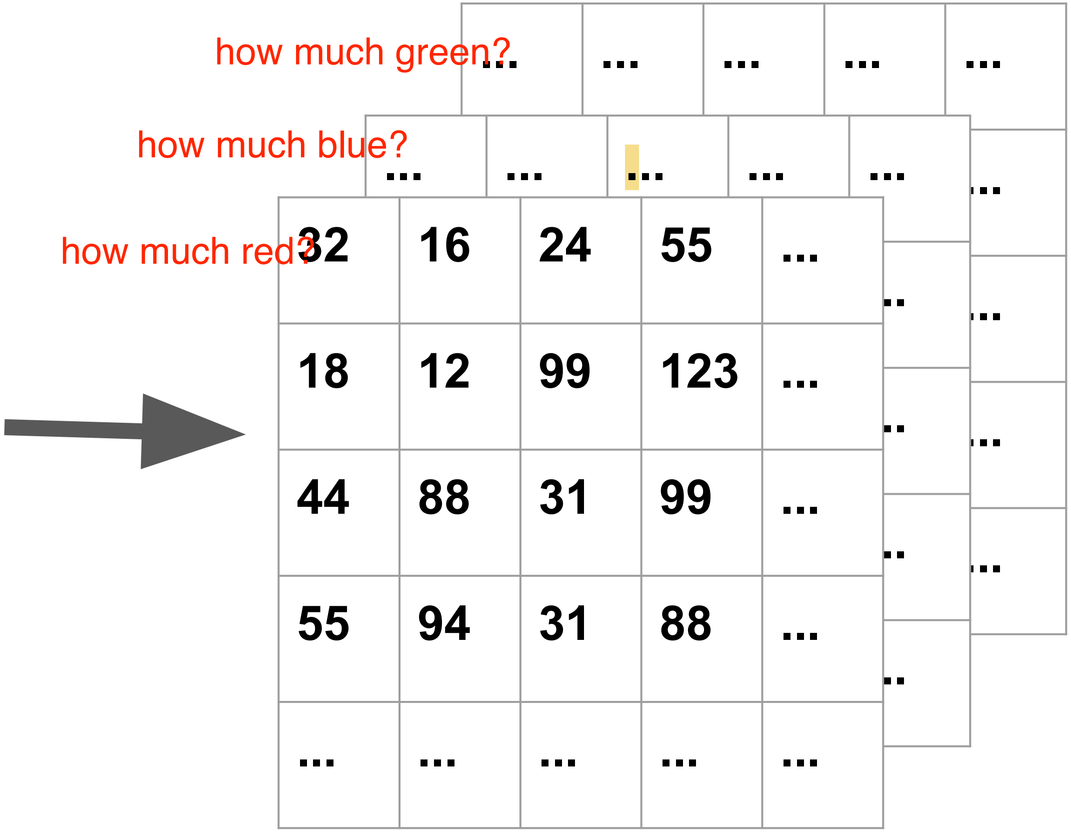
Topics

- ✓ Key Concepts
- ✓ The Deep Learning Landscape
- ✓ Your First Model in Keras
- ☐ Computer Vision and Convolutional Networks
- ☐ Transfer Learning

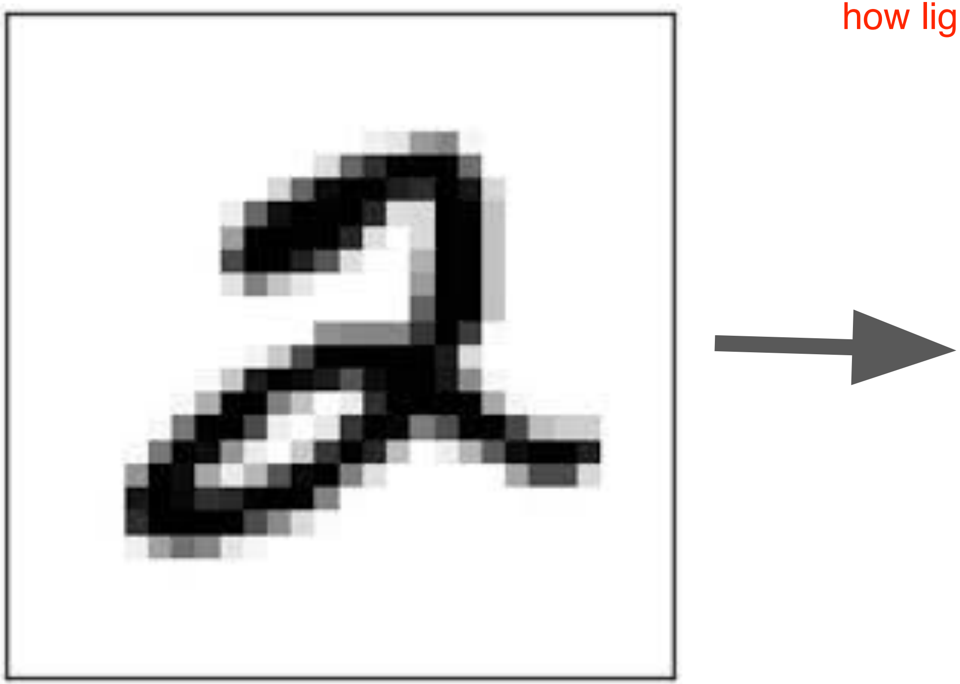
Applications

- Facial recognition
- Medical imaging and automated radiology
- Image tagging

How Are Images Represented



MNIST and Grayscale

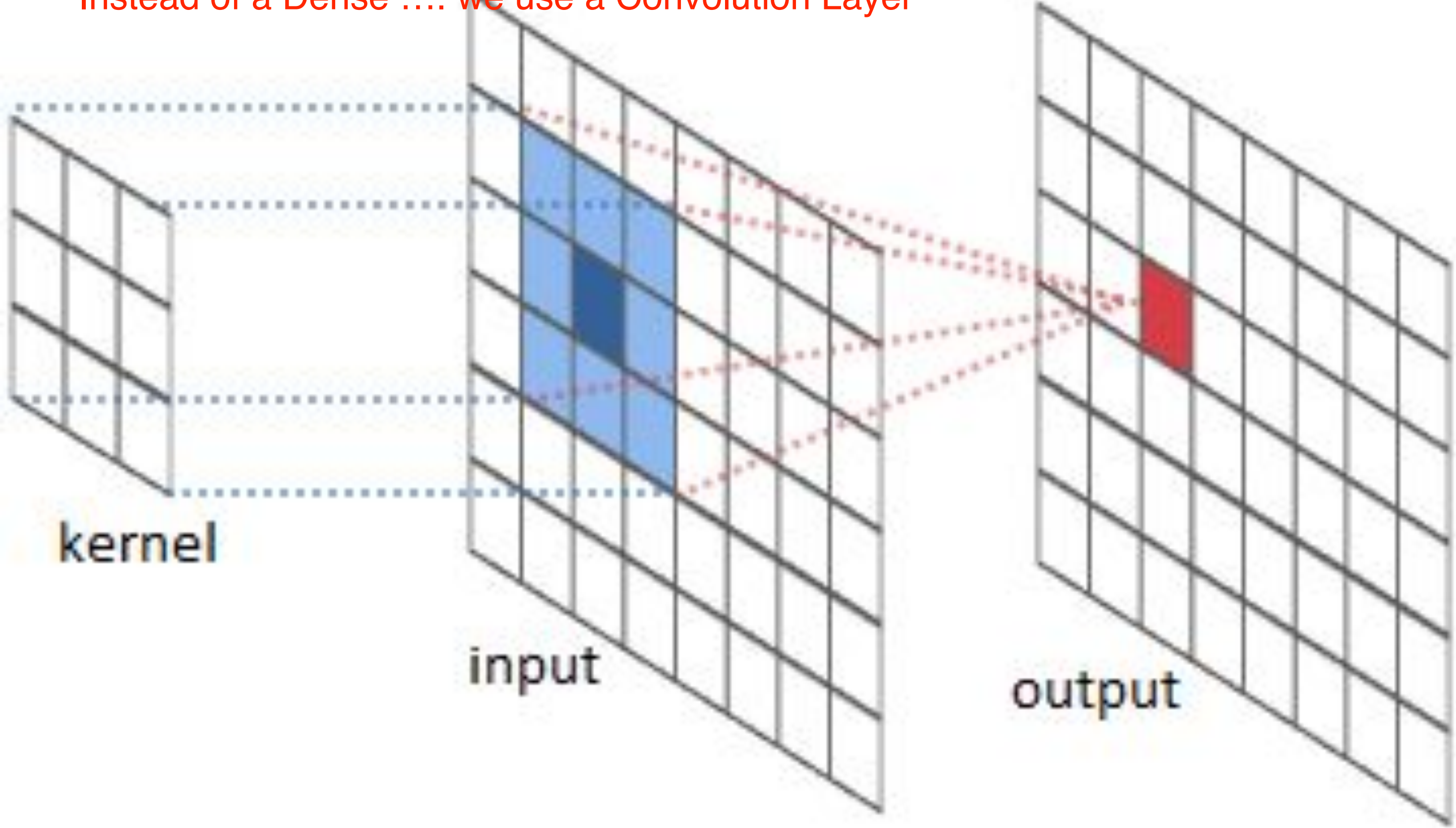


how light or dark?

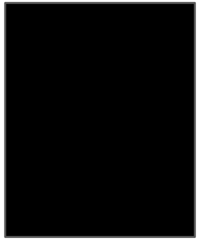
32	16	24	55	...
18	12	99	123	...
44	88	31	99	...
55	94	31	88	...
...	...	...	...	...

The Convolution

Instead of a Dense …. we use a Convolution Layer



The Convolution



Data

200	**200**	**...**	**...**	**...**
200	**200**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**

This specific convolution is
Kind of like a horizontal line detector

Convolution

1.5	1.5
-1.5	-1.5

$$= 200(1.5) + 200(1.5) - 200(1.5) - 200(1.5) = 0$$

The Convolution

Data

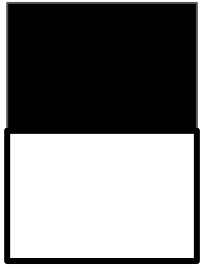
0	**0**	**...**	**...**	**...**
0	**0**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**

Convolution

1.5	1.5
-1.5	-1.5

$= 4(0)(1.5)$
$= 0$

The Convolution



Data

200	200	...	...	...
0	0	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Convolution

1.5	1.5
-1.5	-1.5

$= 2 * 1.5 * 200$
$= 600$

Convolutions for Everything

- Many patterns can be represented
- Filters in later layers capture more complex patterns
- Optimized to help prediction

HOW A DEEP NEURAL NETWORK SEES

Circles inside circles?

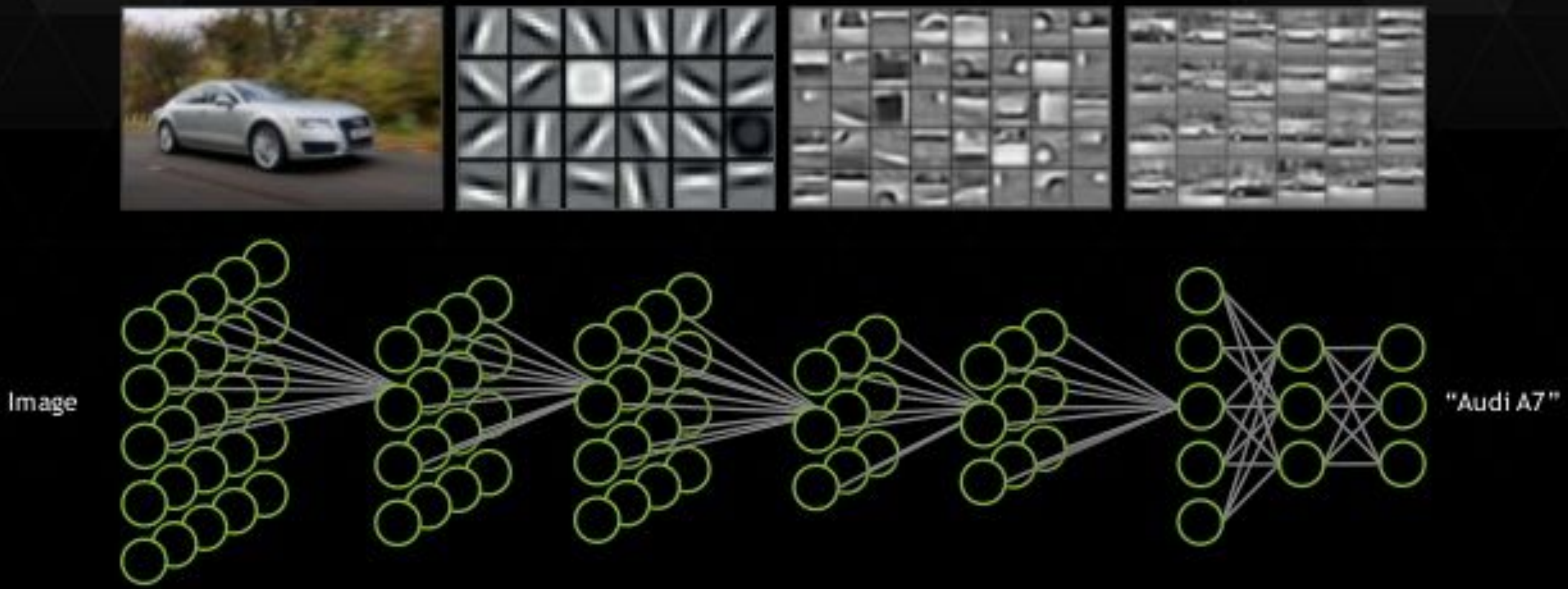
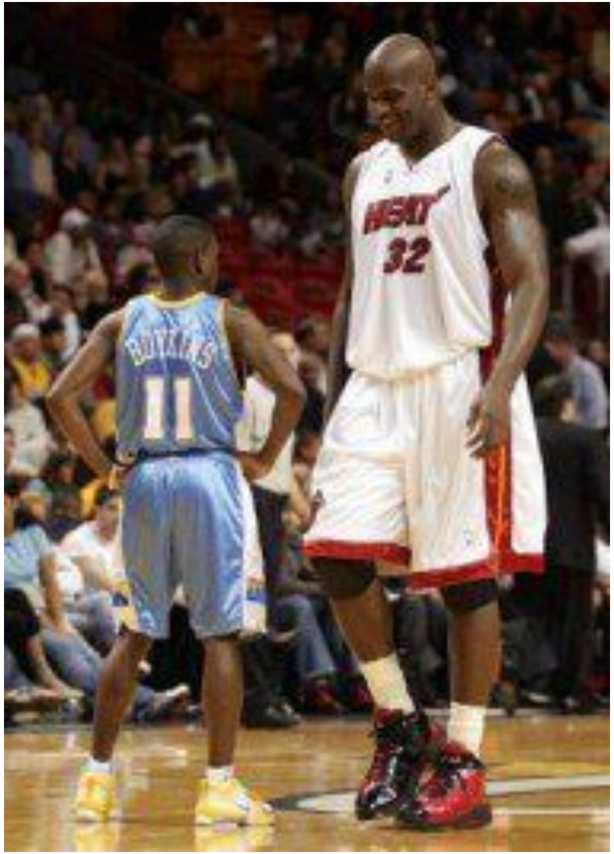


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

Topics

- ✓ Key Concepts
- ✓ The Deep Learning Landscape
- ✓ Your First Model in Keras
- ✓ Computer Vision and Convolutional Networks
- ☐ Transfer Learning

What About Small Data Sets



HOW A DEEP NEURAL NETWORK SEES

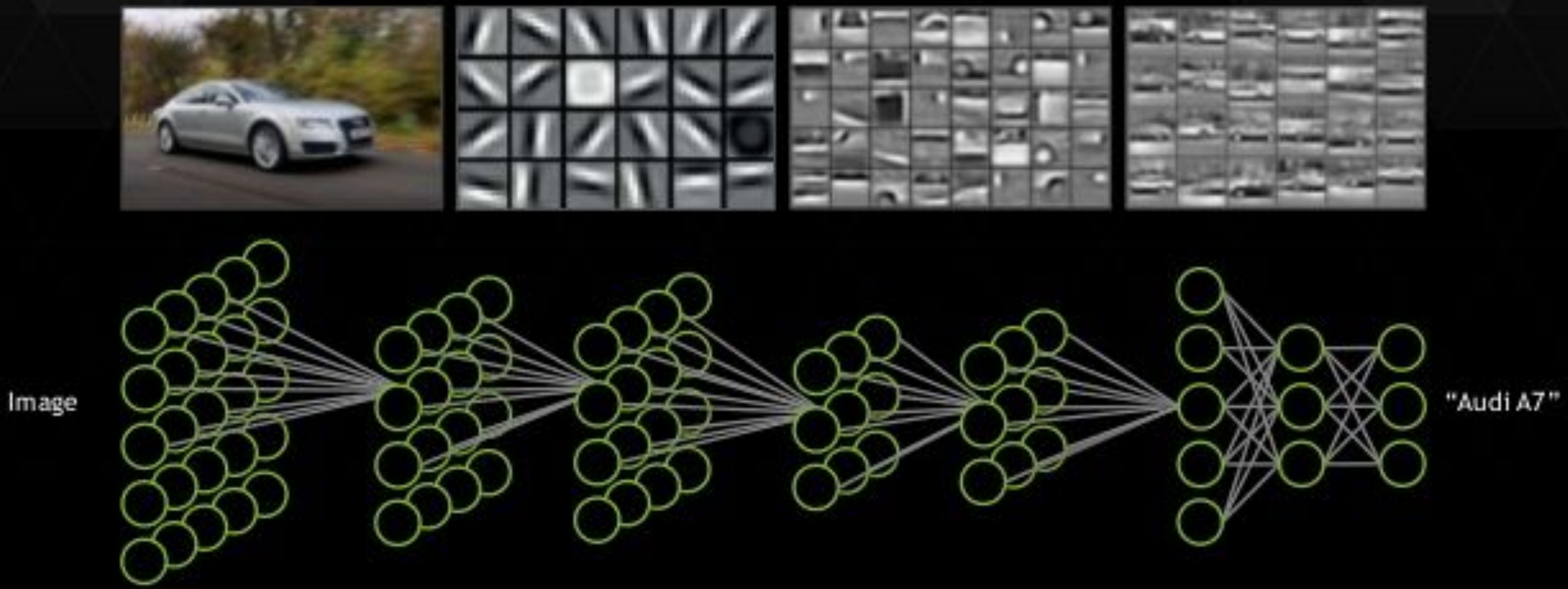


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

Topics

- ✓ Key Concepts
- ✓ The Deep Learning Landscape
- ✓ Your First Model in Keras
- ✓ Computer Vision and Convolutional Networks
- ✓ Transfer Learning