Introduction

To Deep Learning

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

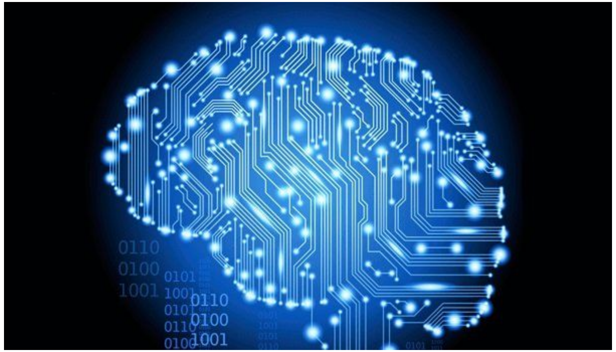
Quick Survey

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



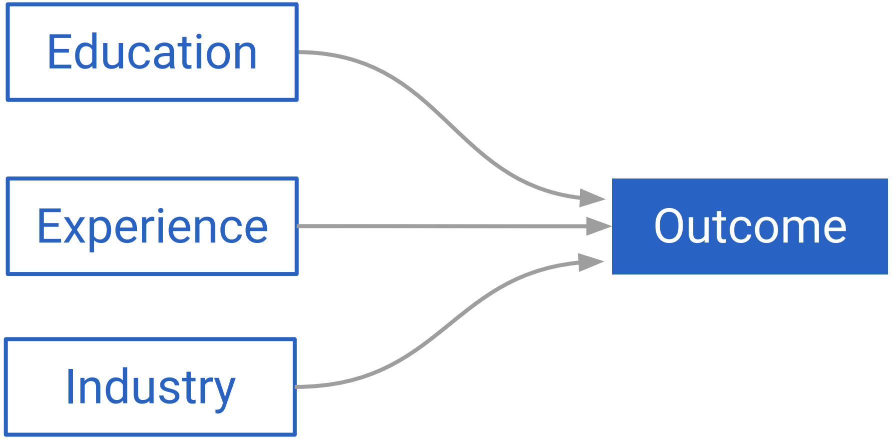
Key Concepts

The Problem With Linear (GLM) Models

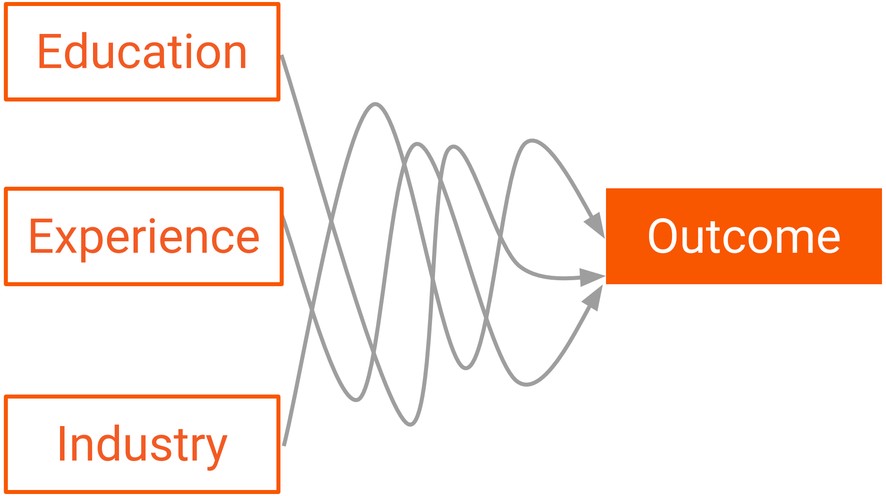
```
= ___ + ___ * years of education
      + ___ * years of experience
      + ___ * geographic indicators
      + ___ * industry indicators
      + …
```

Why We Need Machine Learning

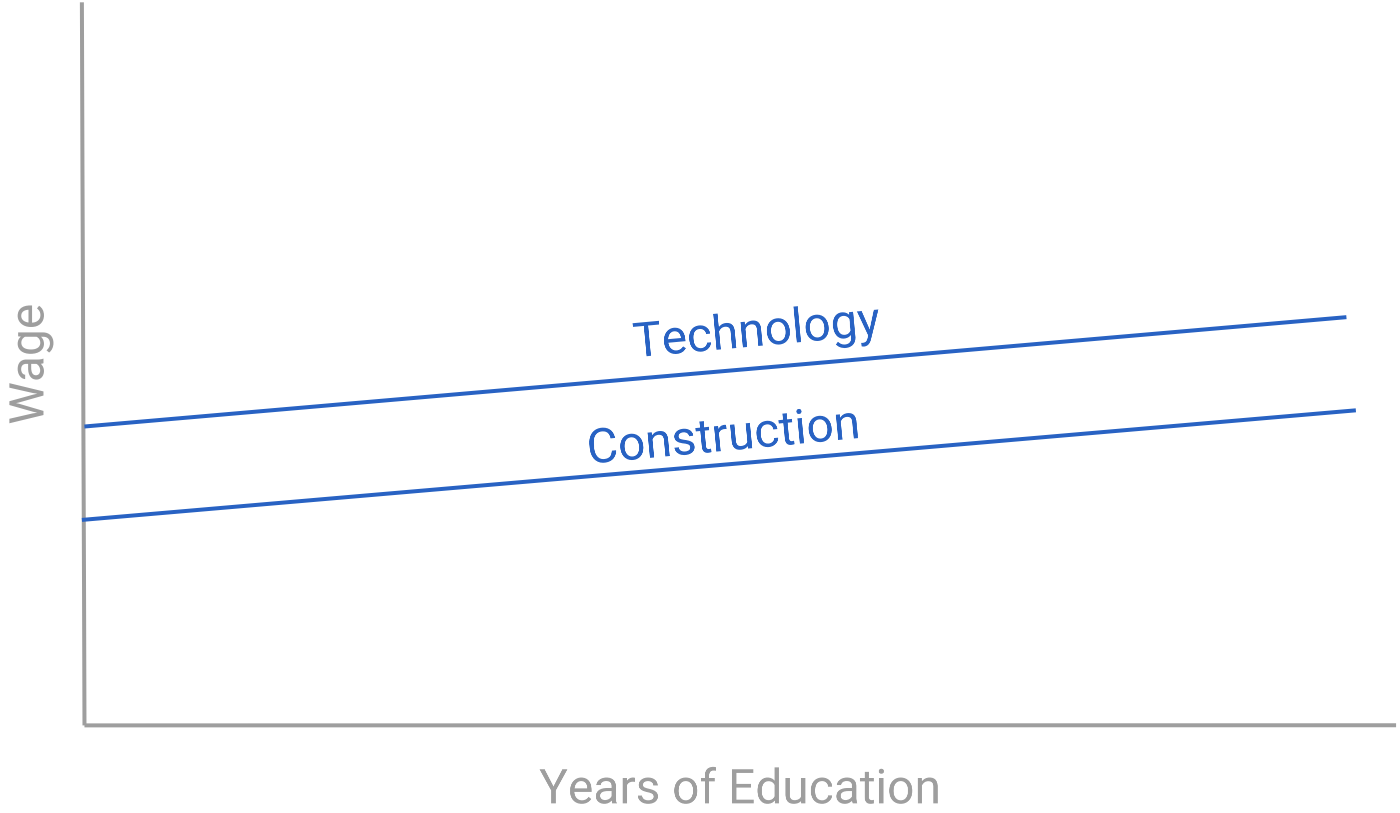
What the Linear Model Captures



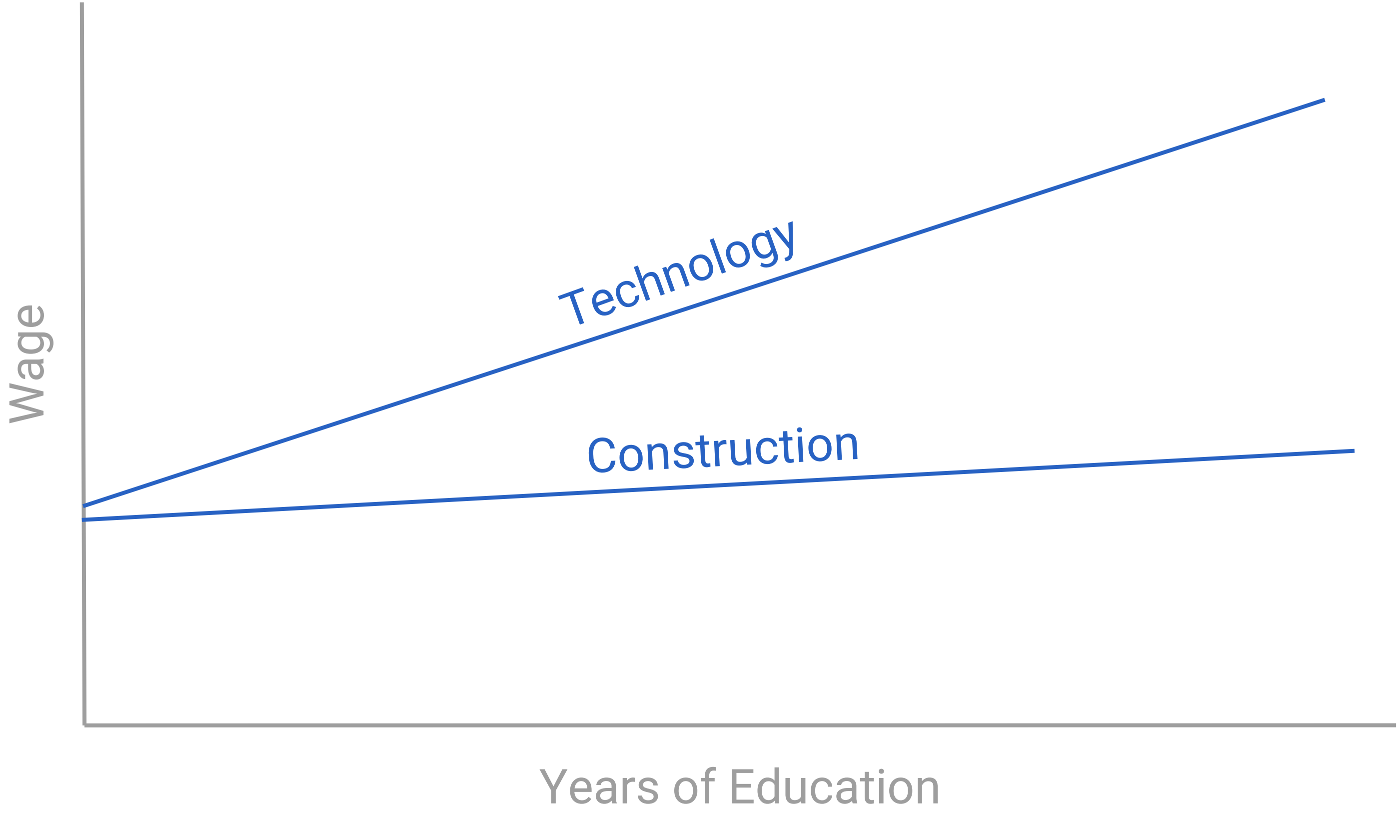
Reality



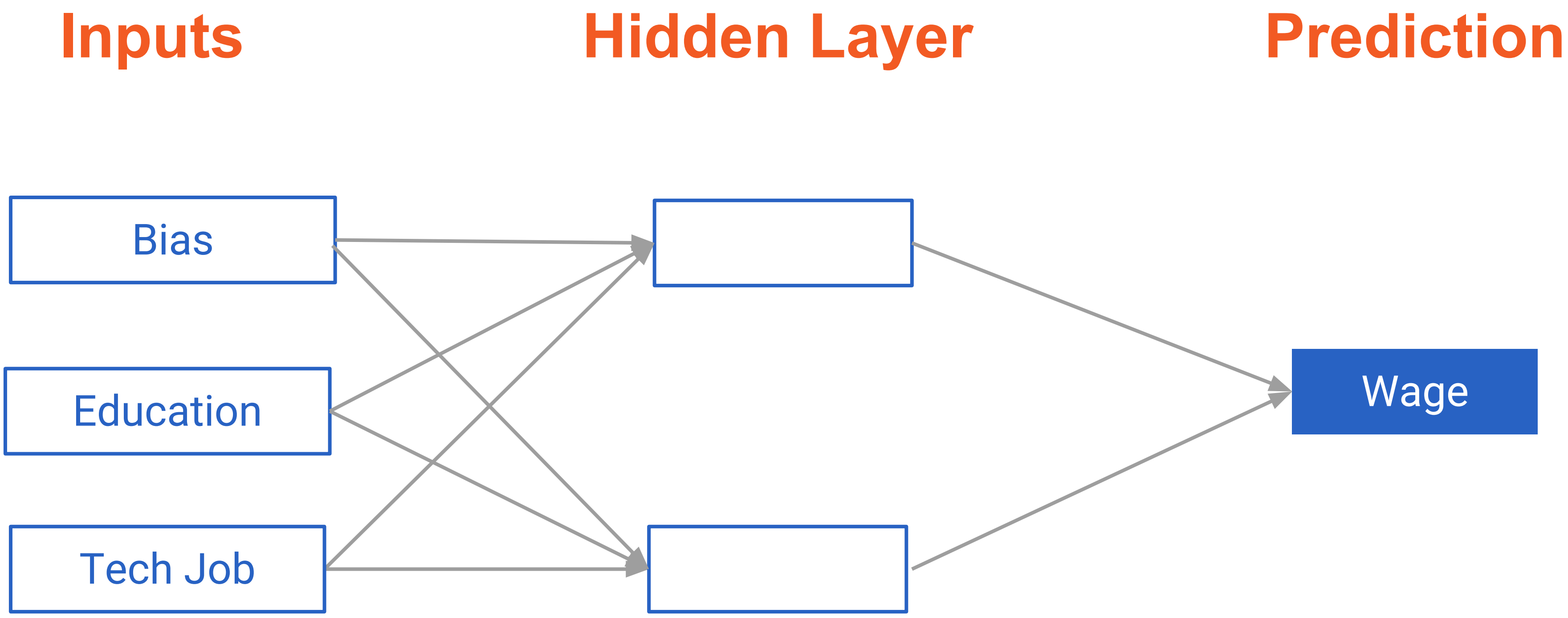
No Interactions



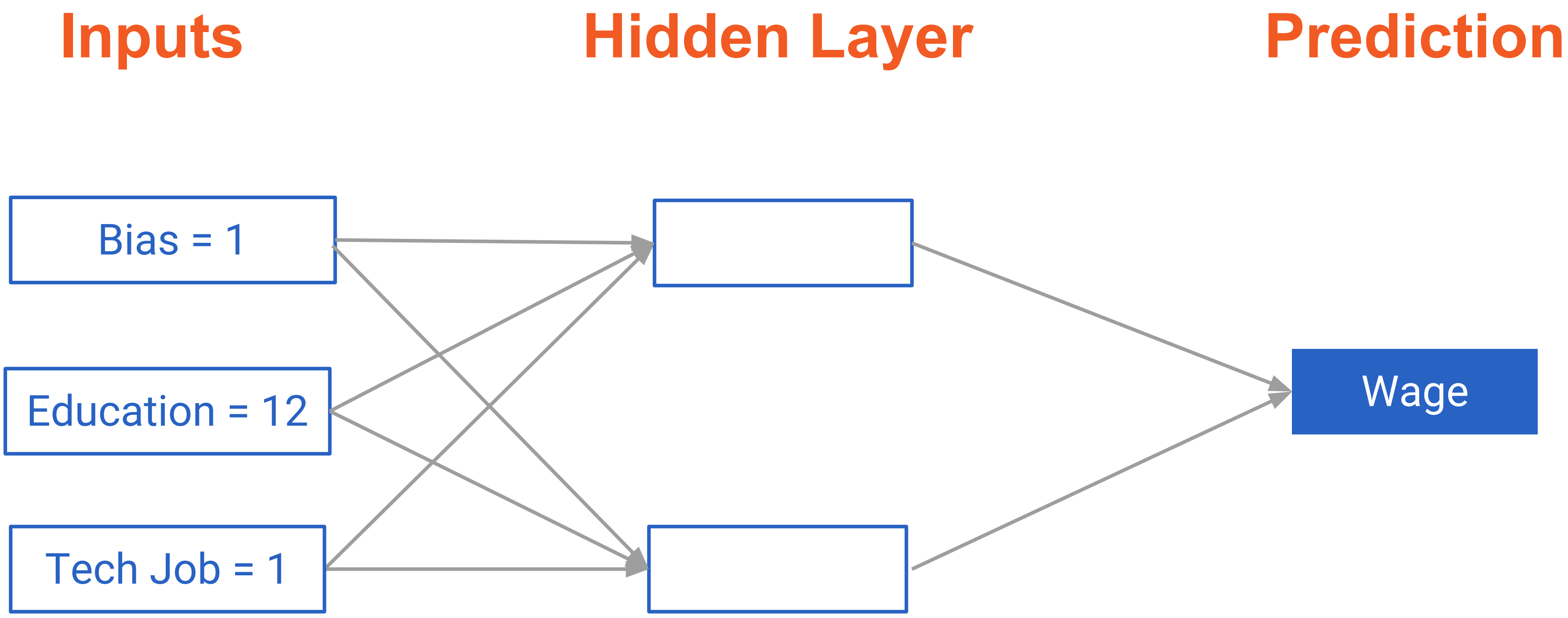
Accounting for Interactions



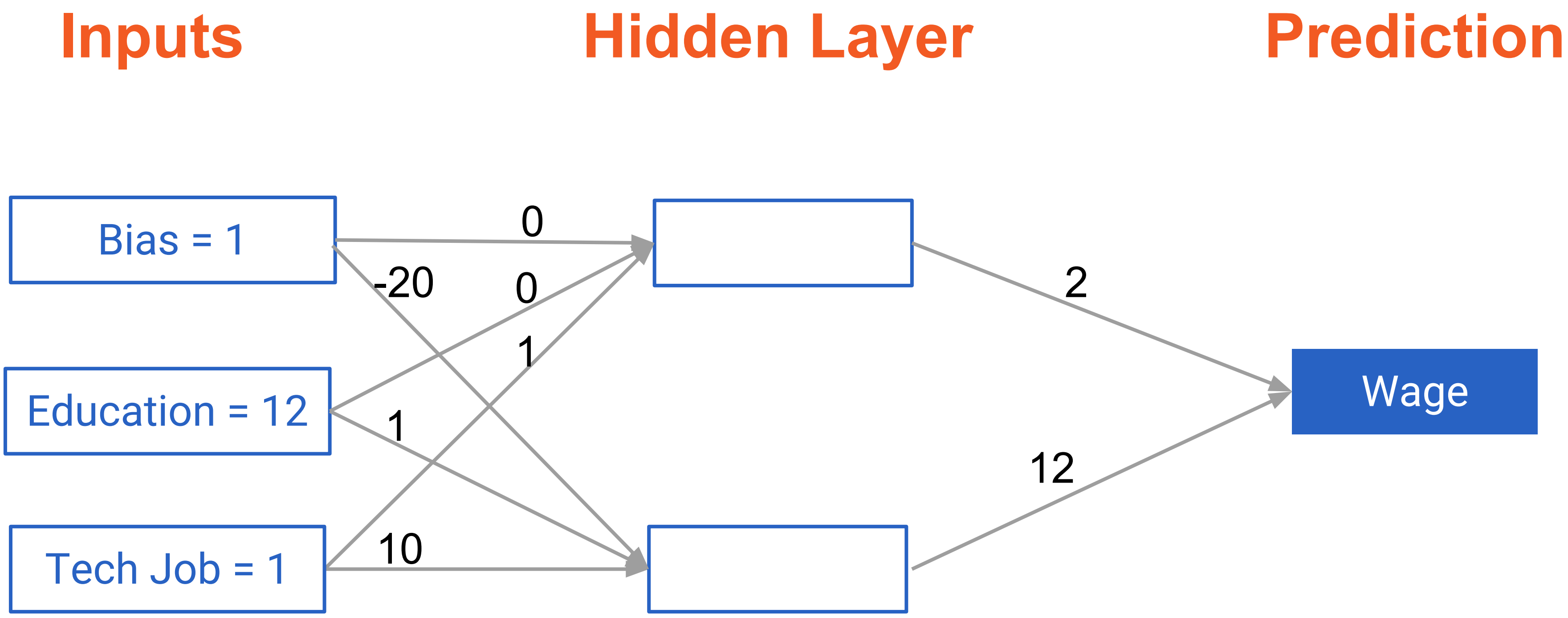
Forward Propagation



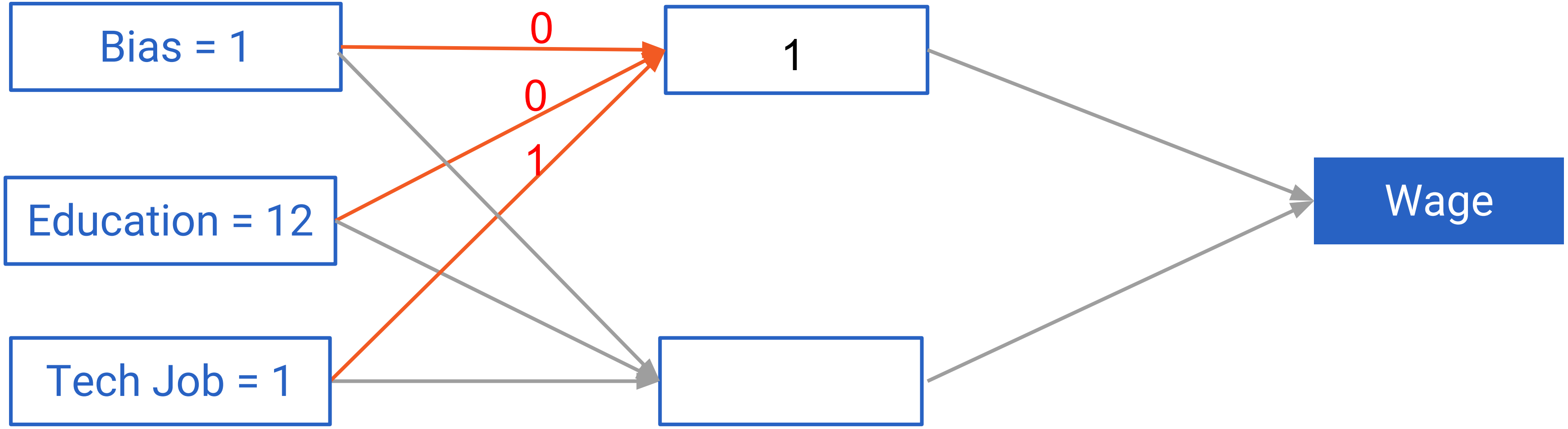
Forward Propagation



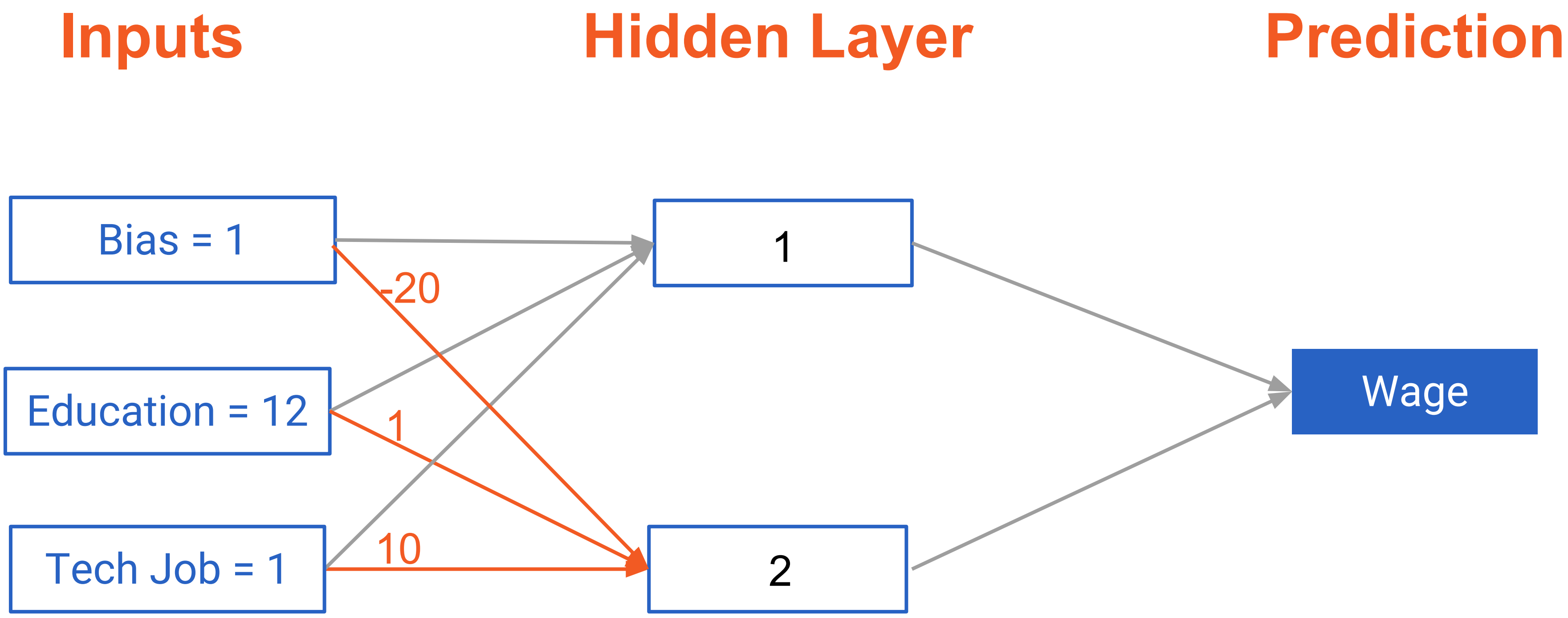
Forward Propagation



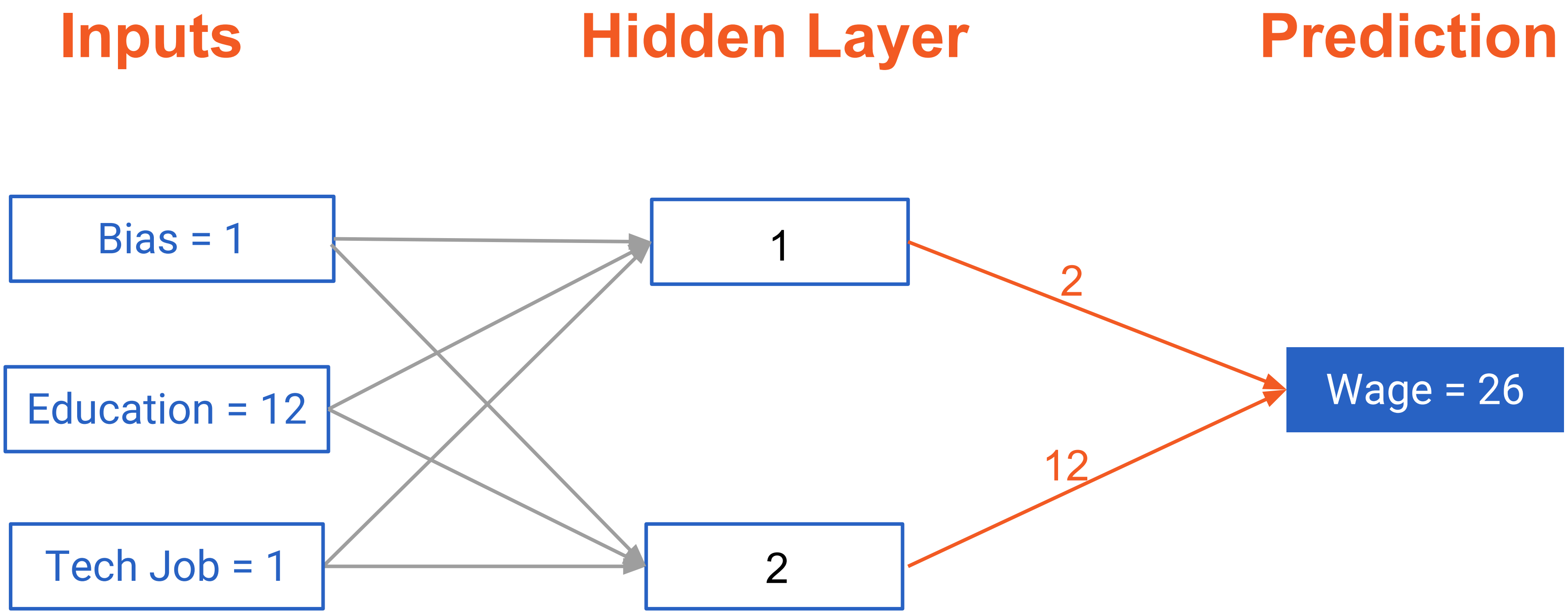
Forward Propagation



Forward Propagation



Forward Propagation

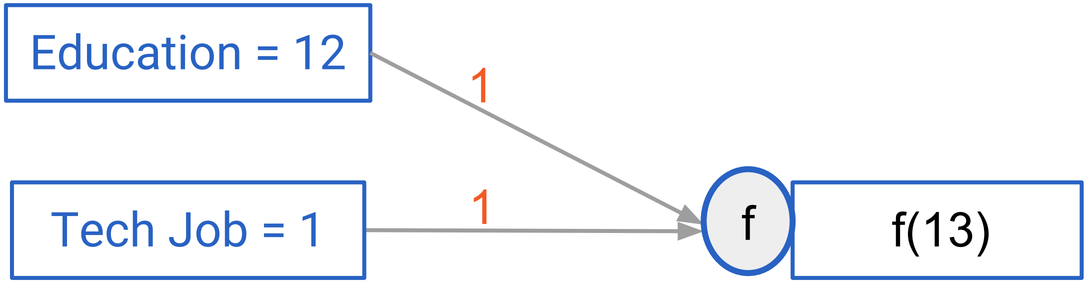


Activation Functions

Non-Linear Function Converting Node Input to Output

<u>What</u>

- Non-linear function converting node input to output



Activation Functions

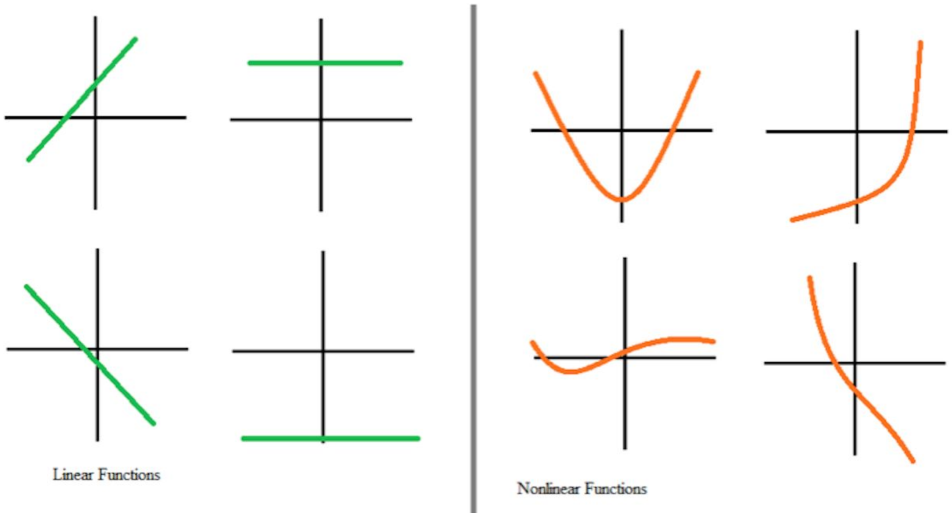
Non-Linear Function Converting Node Input to Output

What

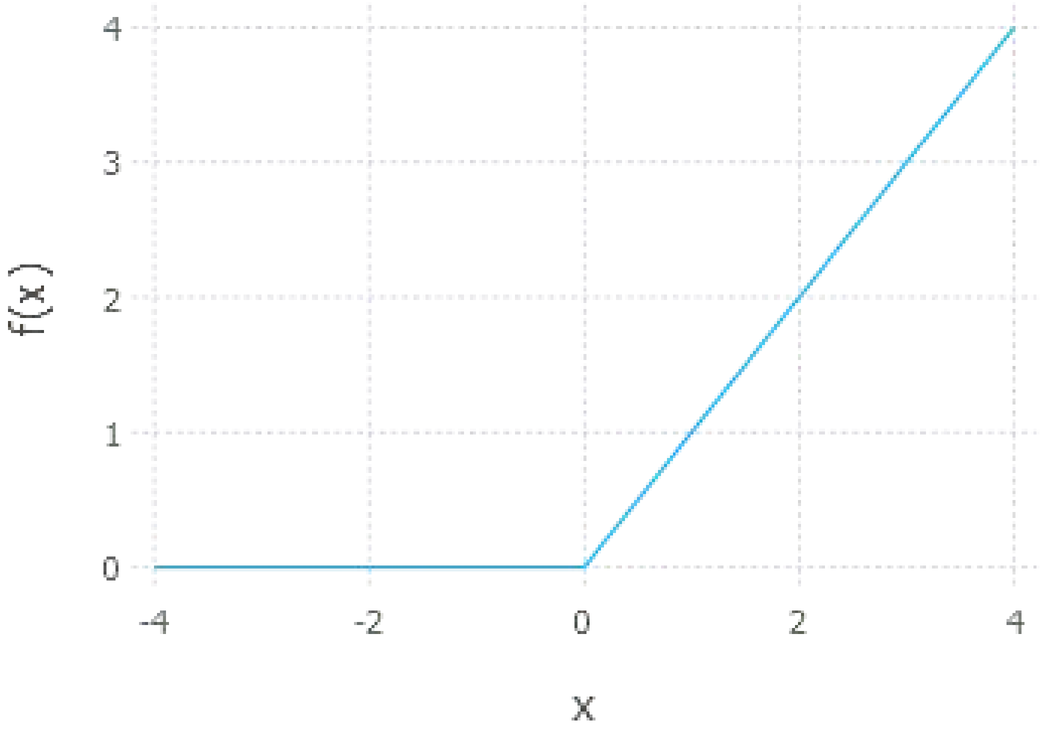
- Non-linear function converting node input to output

Why

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



The ReLU Activation Function



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

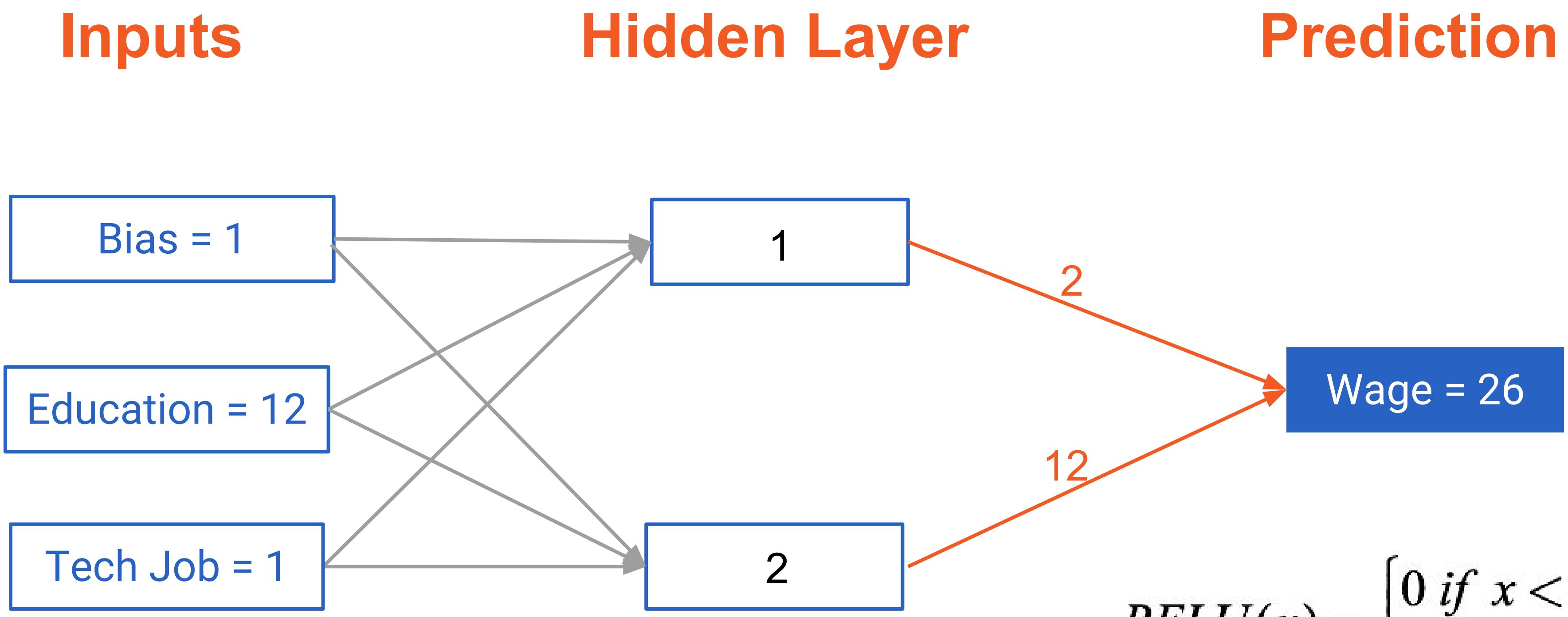
Return to Interactions

<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)



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

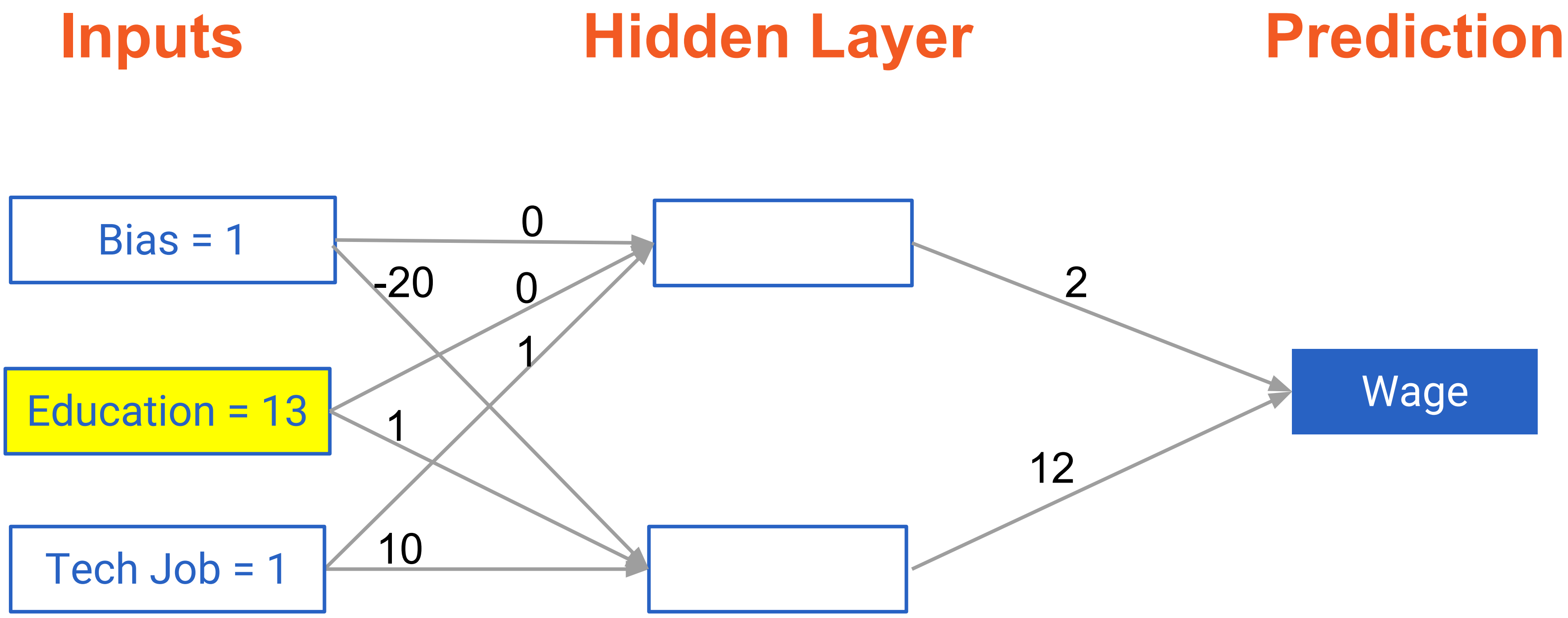
Return to Interactions

<u>Checking For Interactions</u>

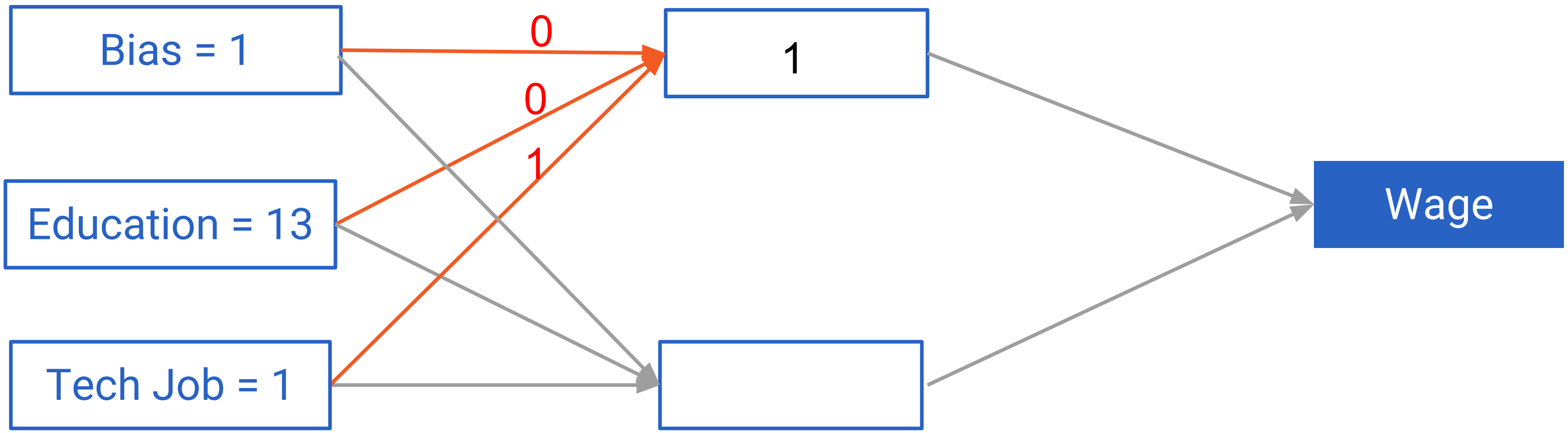
- Make prediction for two education values for tech worker
- Make prediction for same two education values for non-tech worker
- See if increase in wage differs

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

Forward Propagation



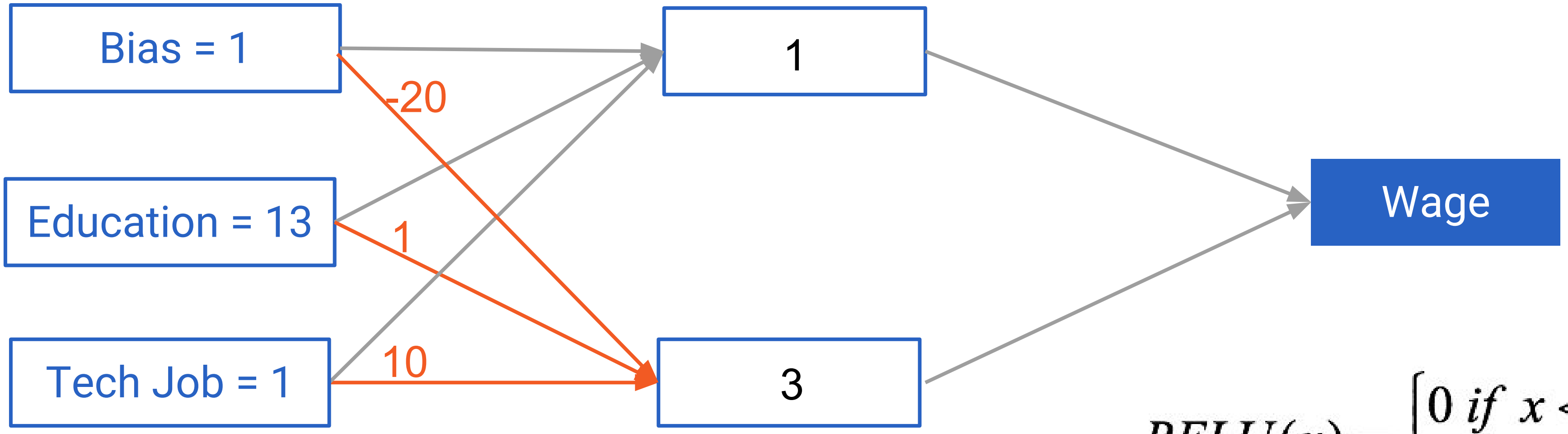
Forward Propagation



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

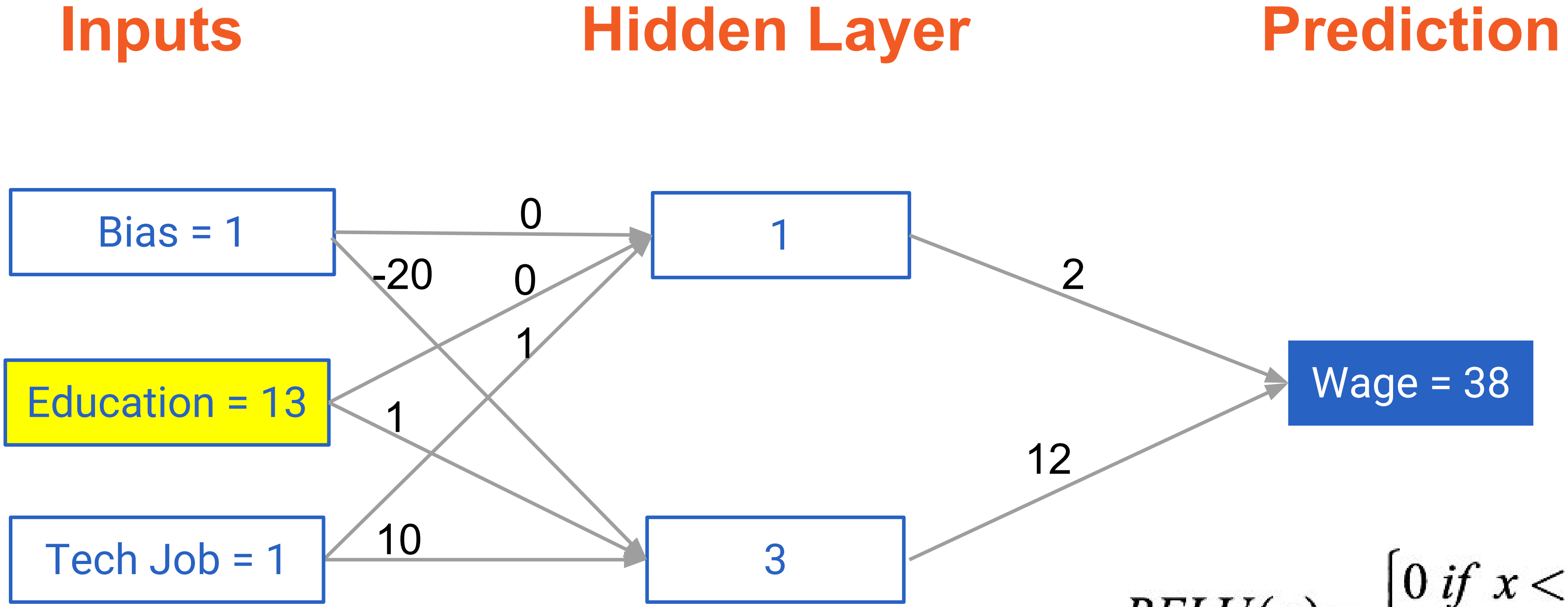
Forward Propagation

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



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

Forward Propagation



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

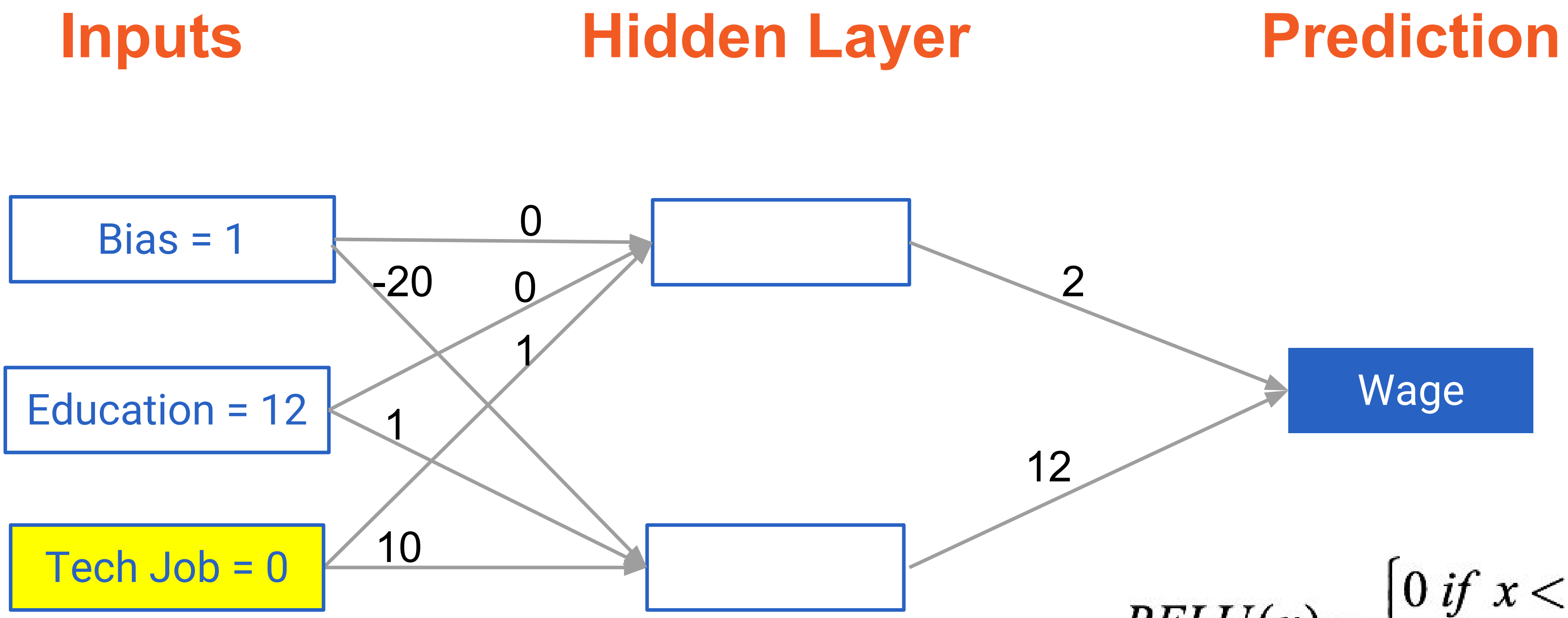
Return to Interactions

<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

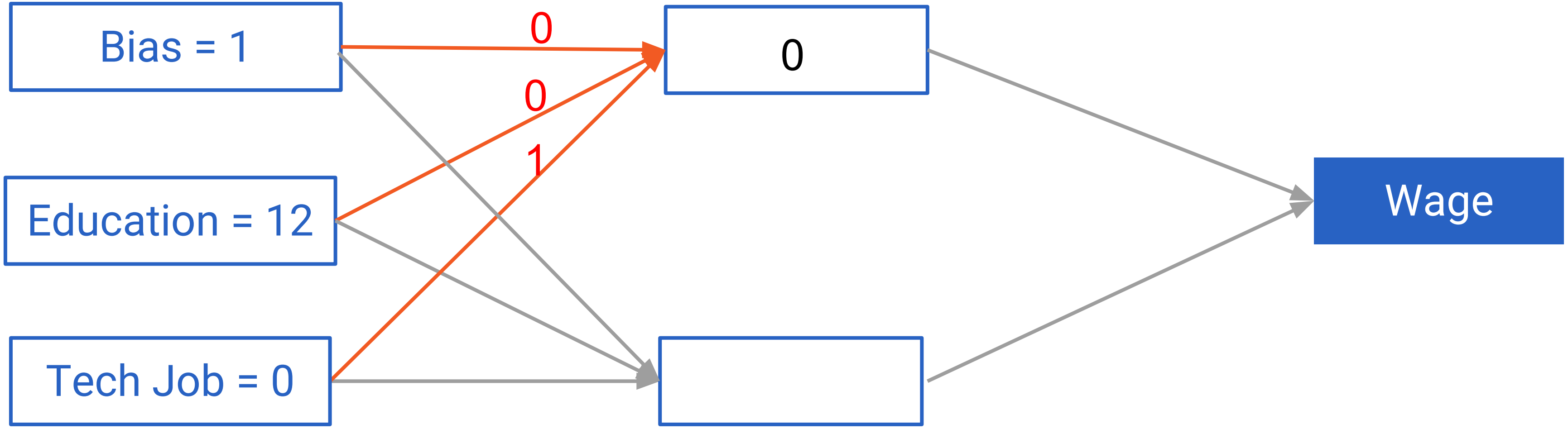
	Tech Job = 0	Tech Job = 1
Education = 12		**26**
Education = 13		**38**

Forward Propagation (with ReLU)



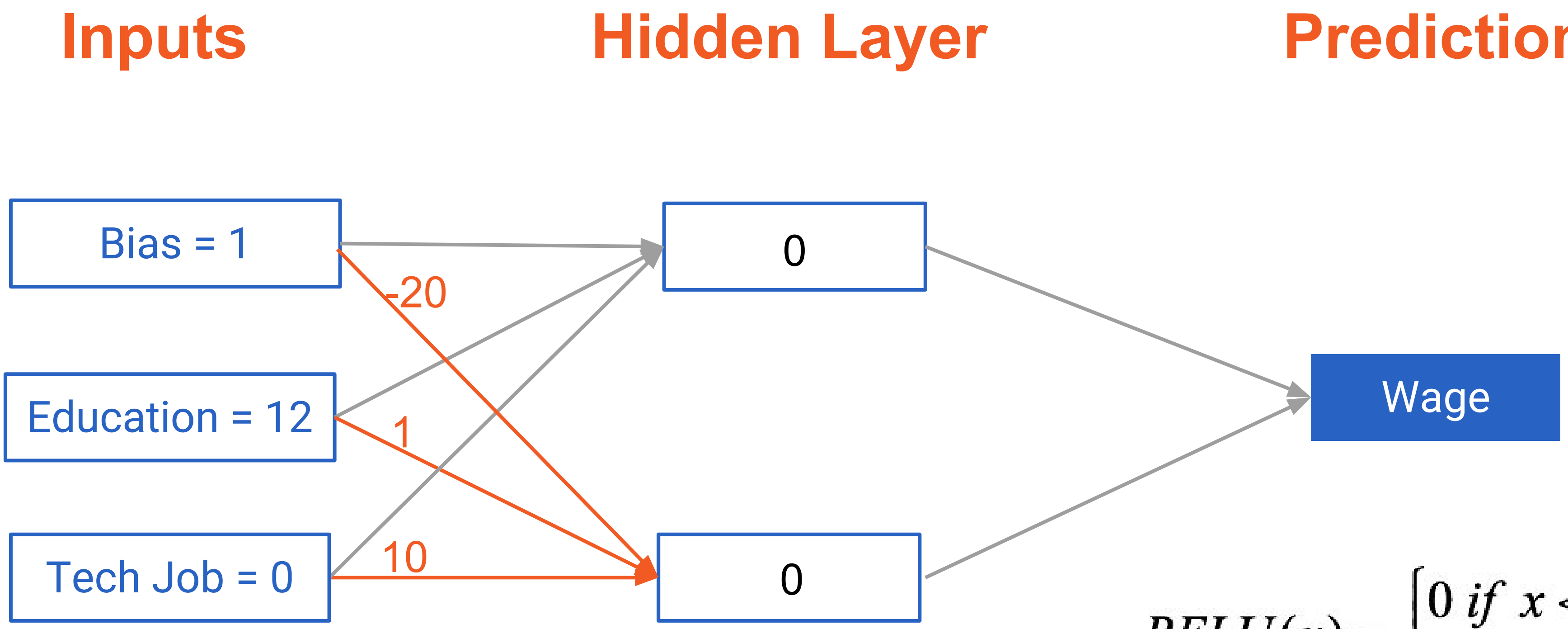
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



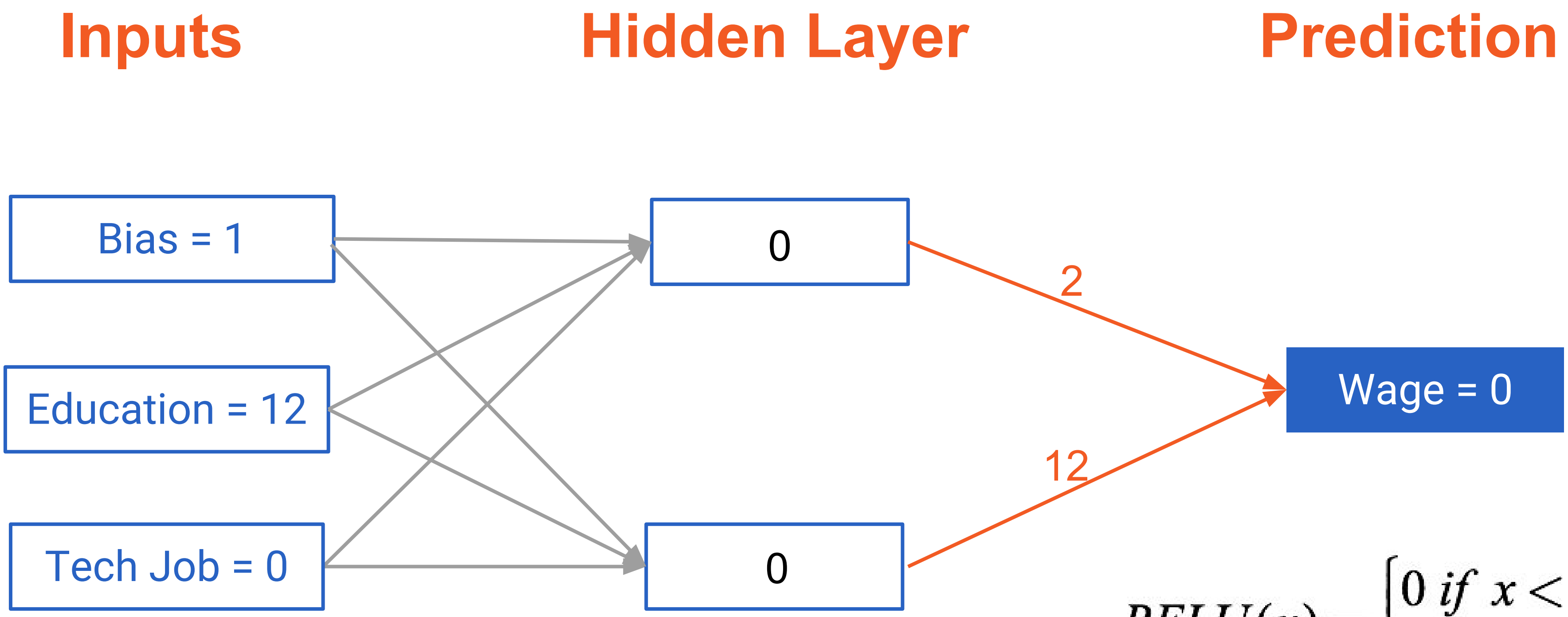
$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



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

Forward Propagation (with ReLU)



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

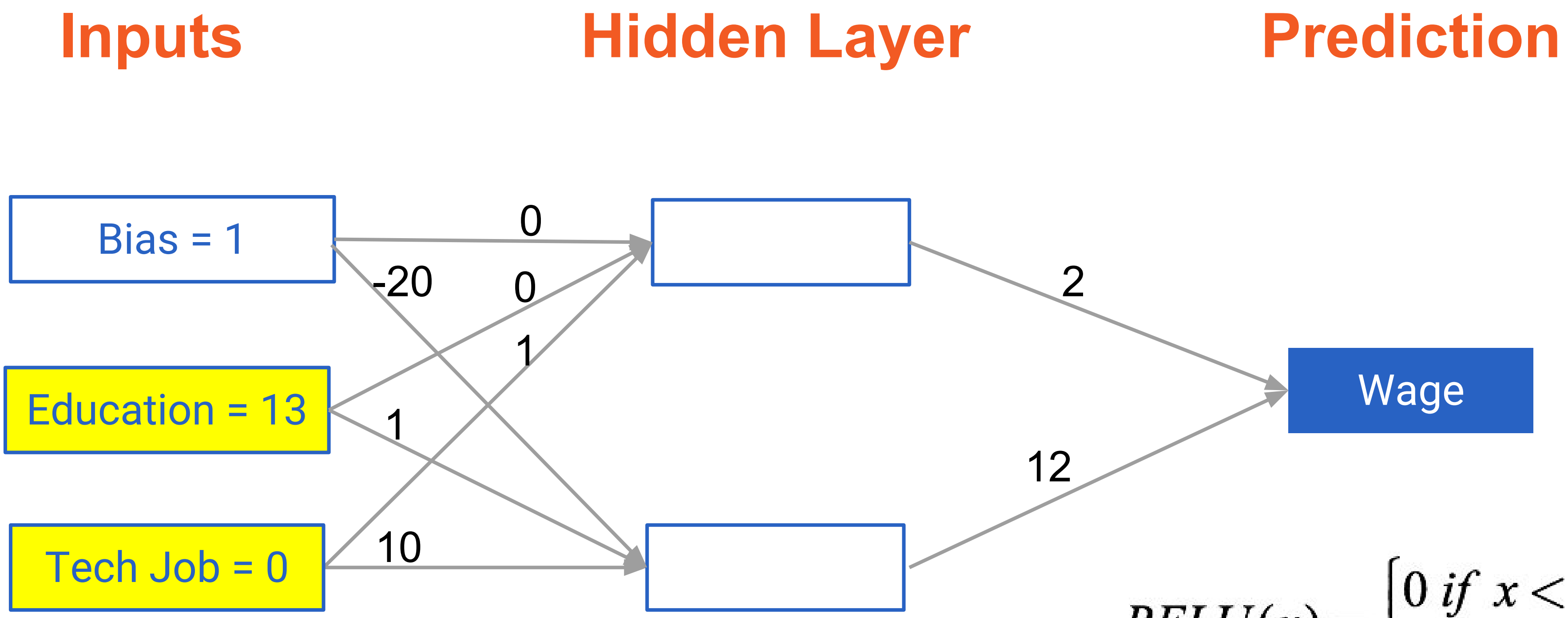
Return to Interactions

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

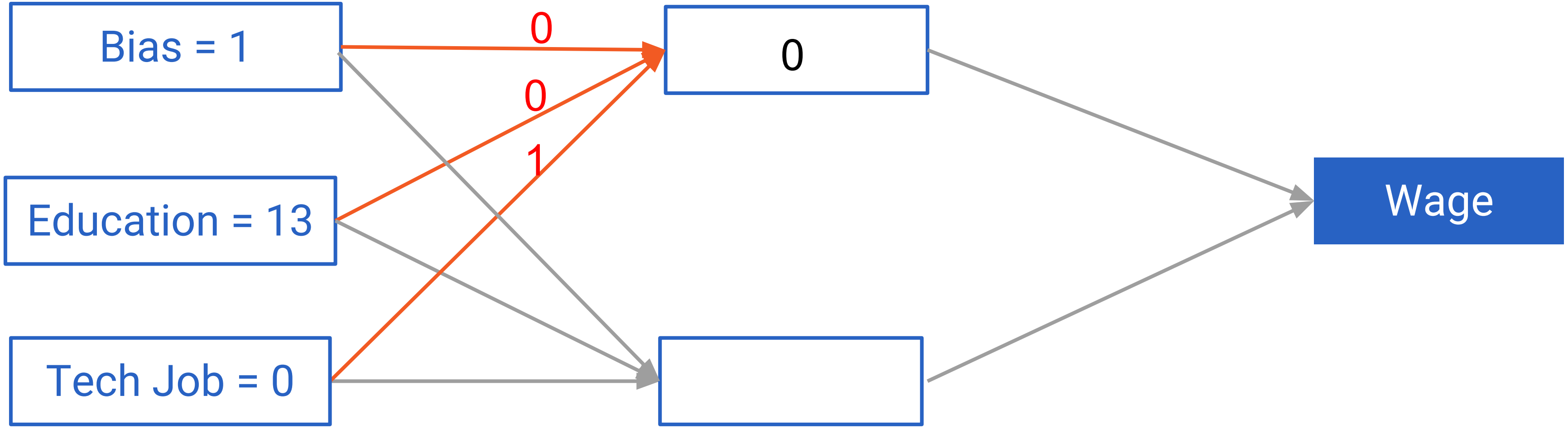
	Tech Job = 0	Tech Job = 1
Education = 12	**0**	**26**
Education = 13		**38**

Forward Propagation (with ReLU)



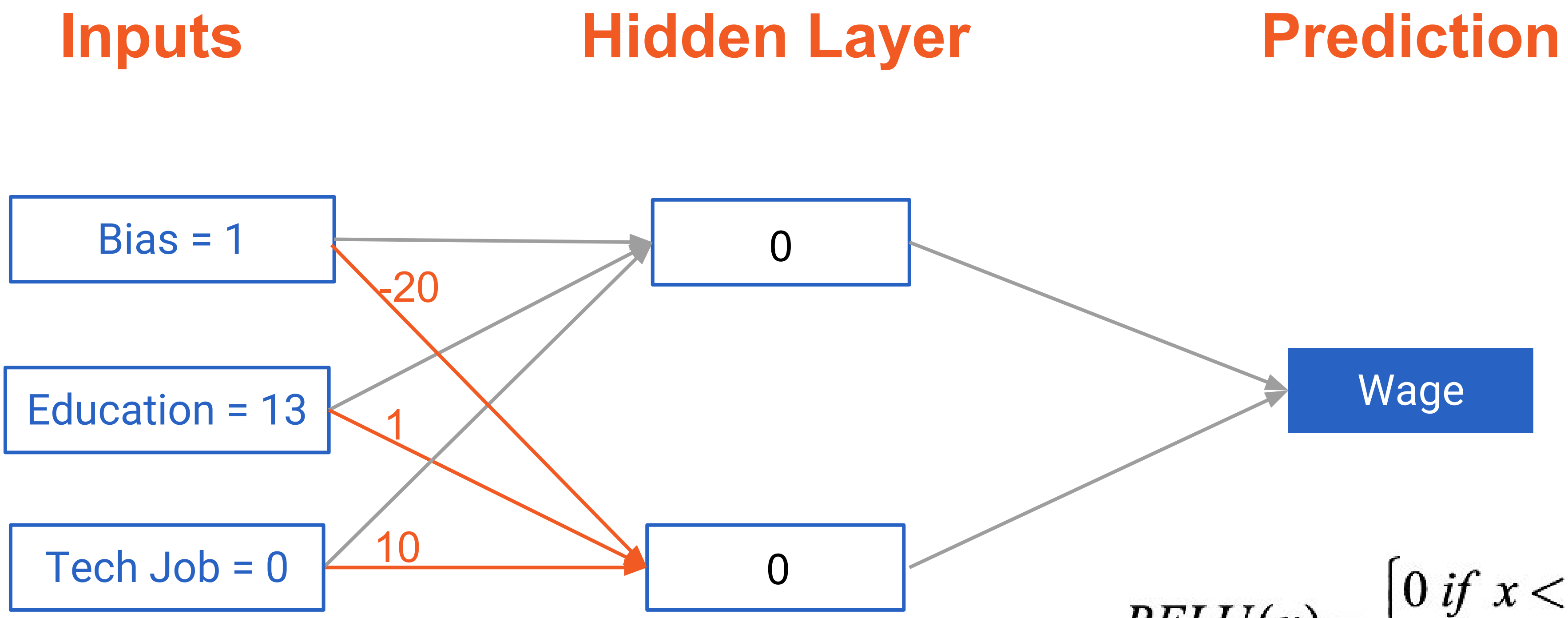
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

Forward Propagation (with ReLU)



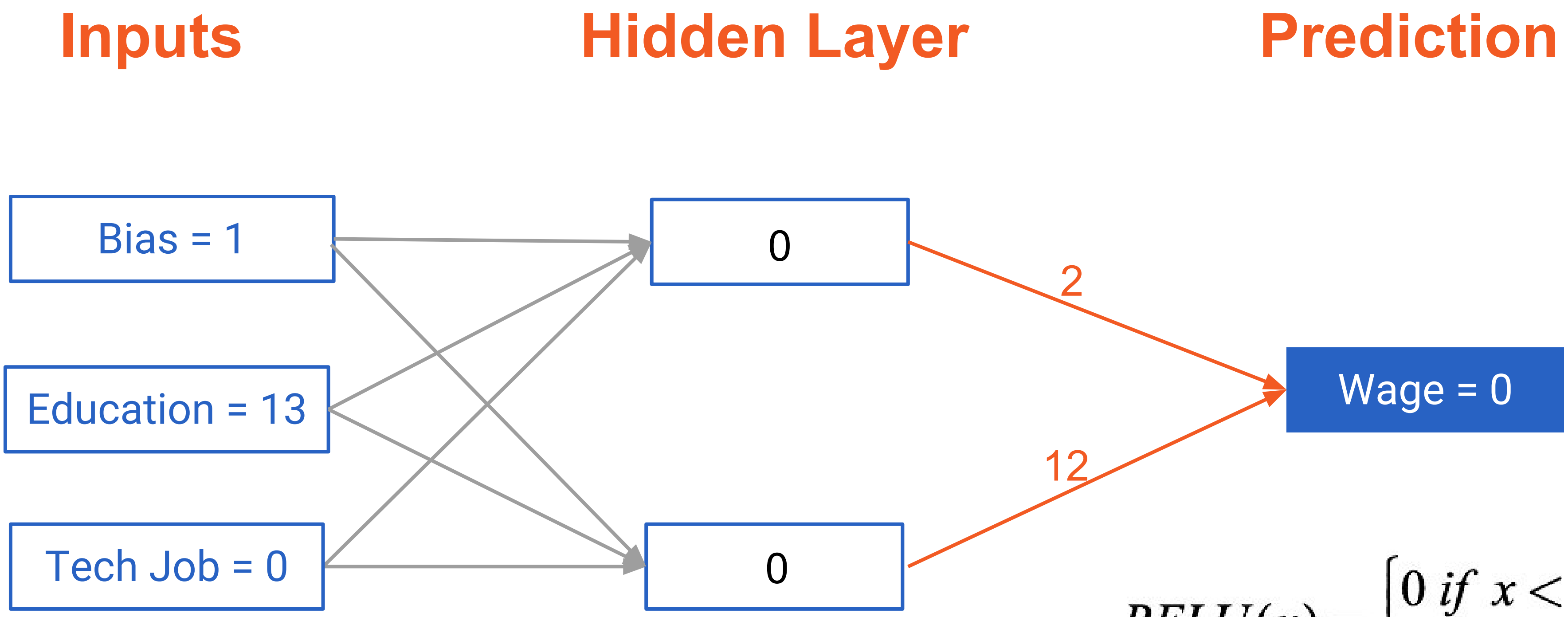
$$RELU(x) = \begin{cases} 0 \ if \ x < 0 \\ x \ if \ x >= 0 \end{cases}$$

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	**0**	**38**

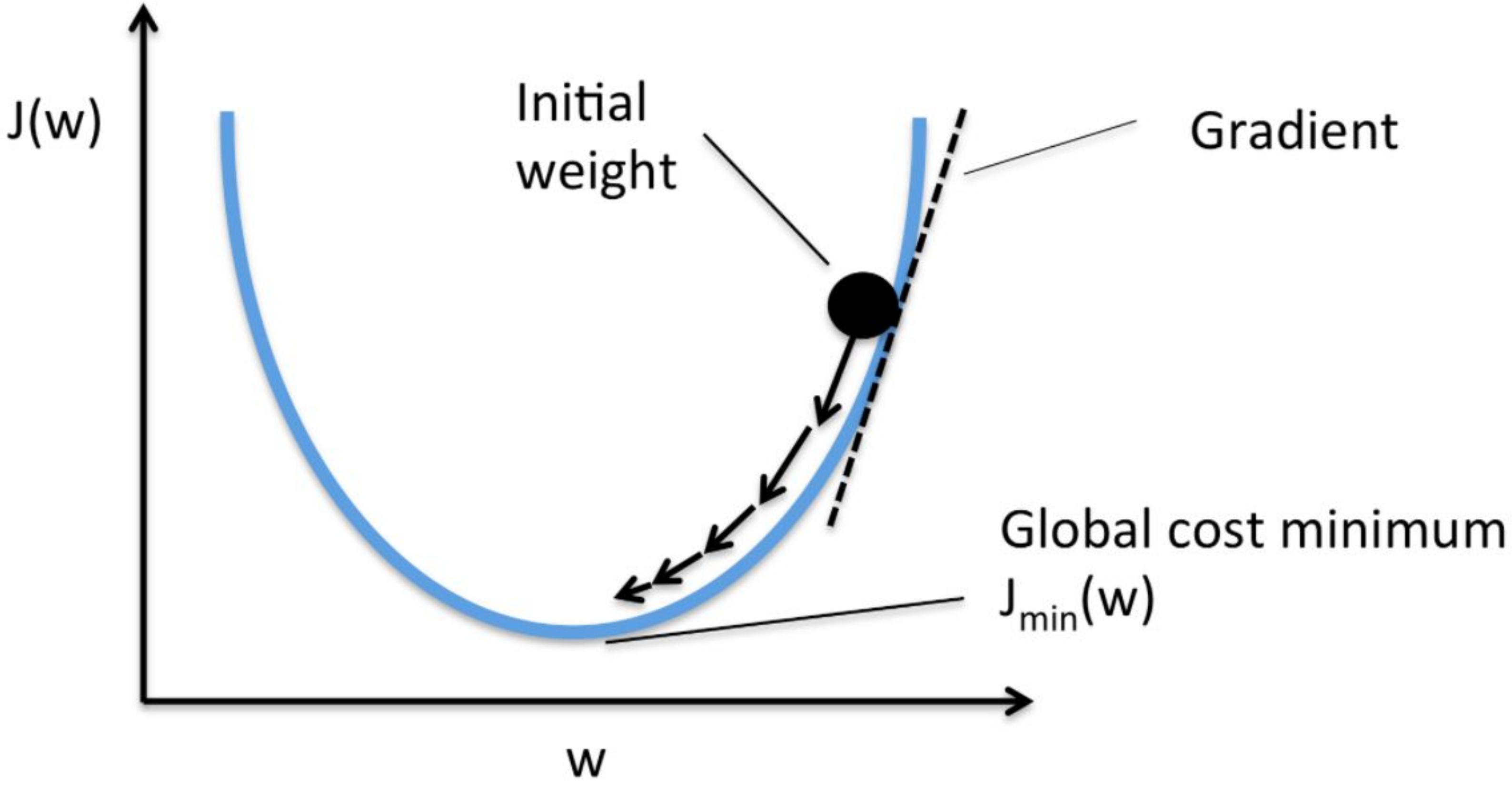
The Takeaway

Checking For Interactions

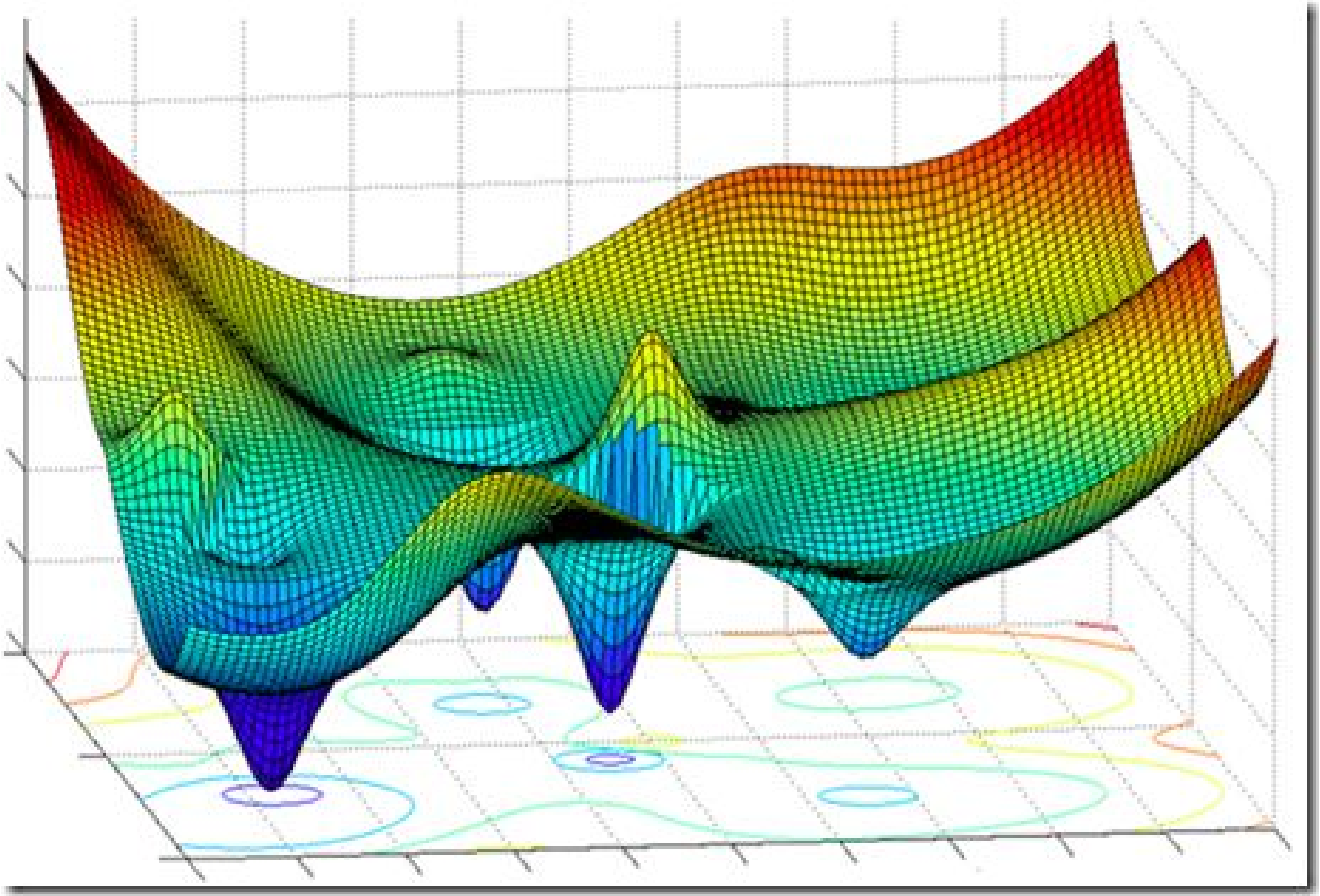
- Neural network models capture interactions and non-linearities

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

Gradient Descent



Gradient Descent



Gradient Descent

Repeatedly:

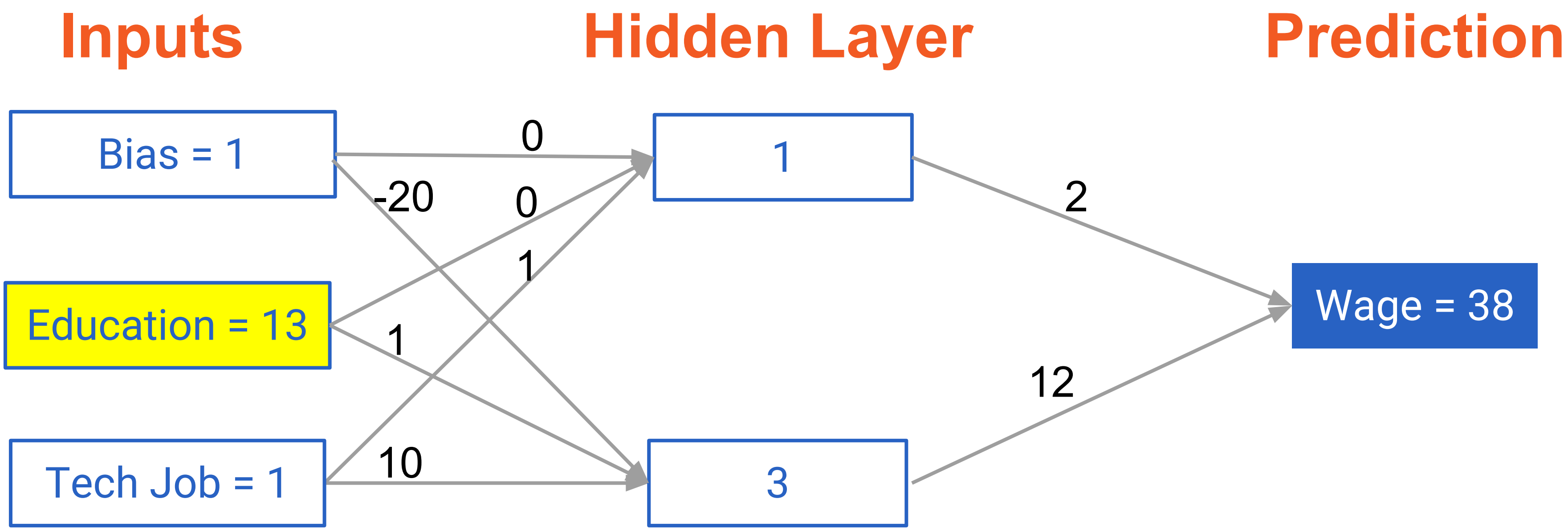
- Find derivative / slope of loss function with respect to each weight

- Take small step downhill (subtracting product of derivative and learning rate)

Back Propagation

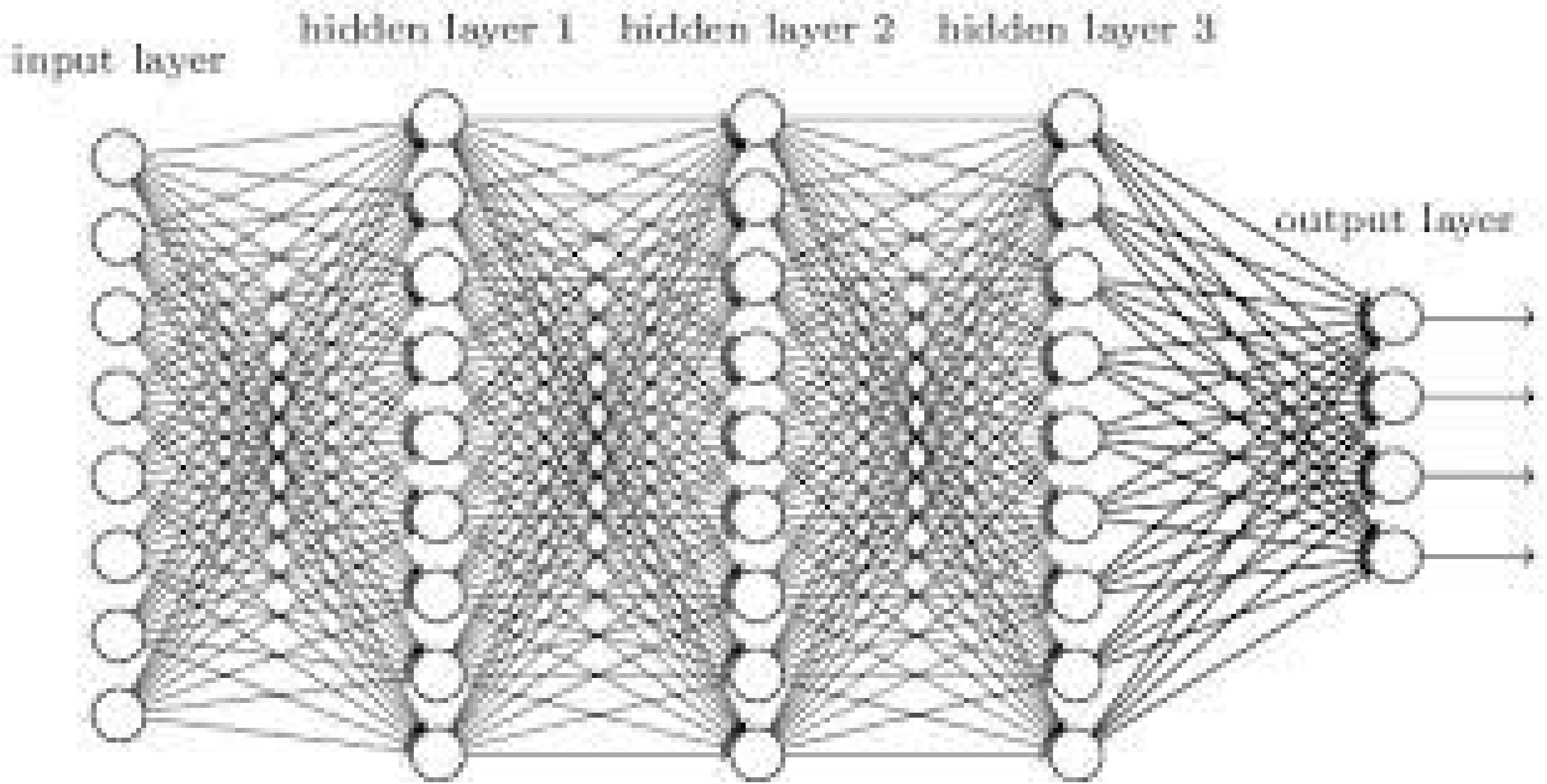
- Used to get derivatives needed to update weights
- Application of chain rule from calculus
- Used after forward propagation to find errors
- Not focusing on this math today

Backward Propagation



Actual Value = 30
Error = 8
Squared Error = 64

Deeper Networks

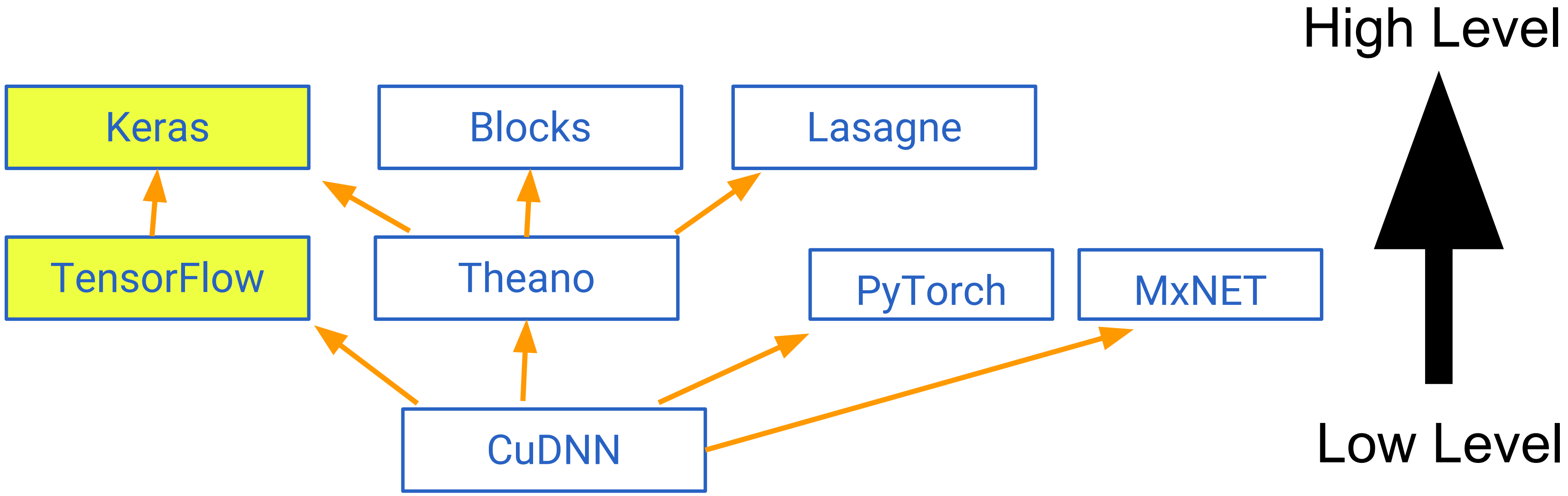


Topics

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

Where Deep Learning Shines

Deep Learning Landscape



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

- Define
- Compile
- Fit
- Predict

Topics

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

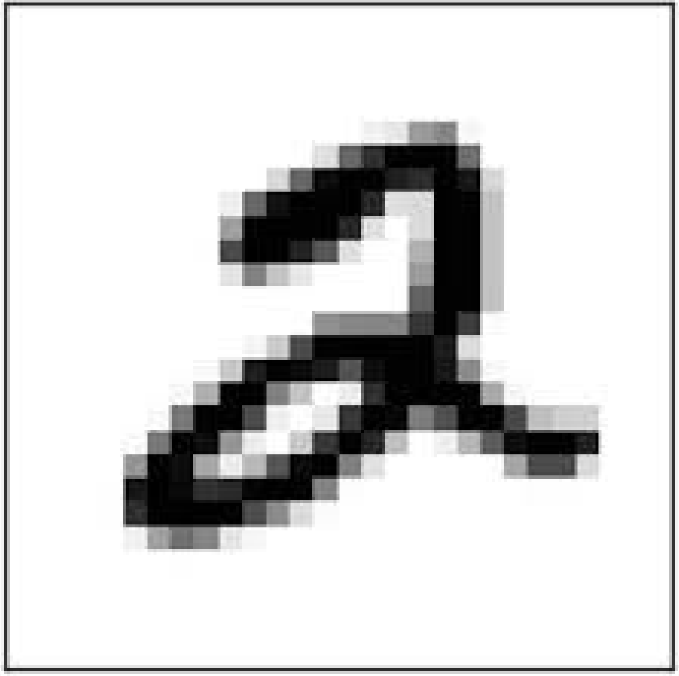
Applications

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

How Are Images Represented

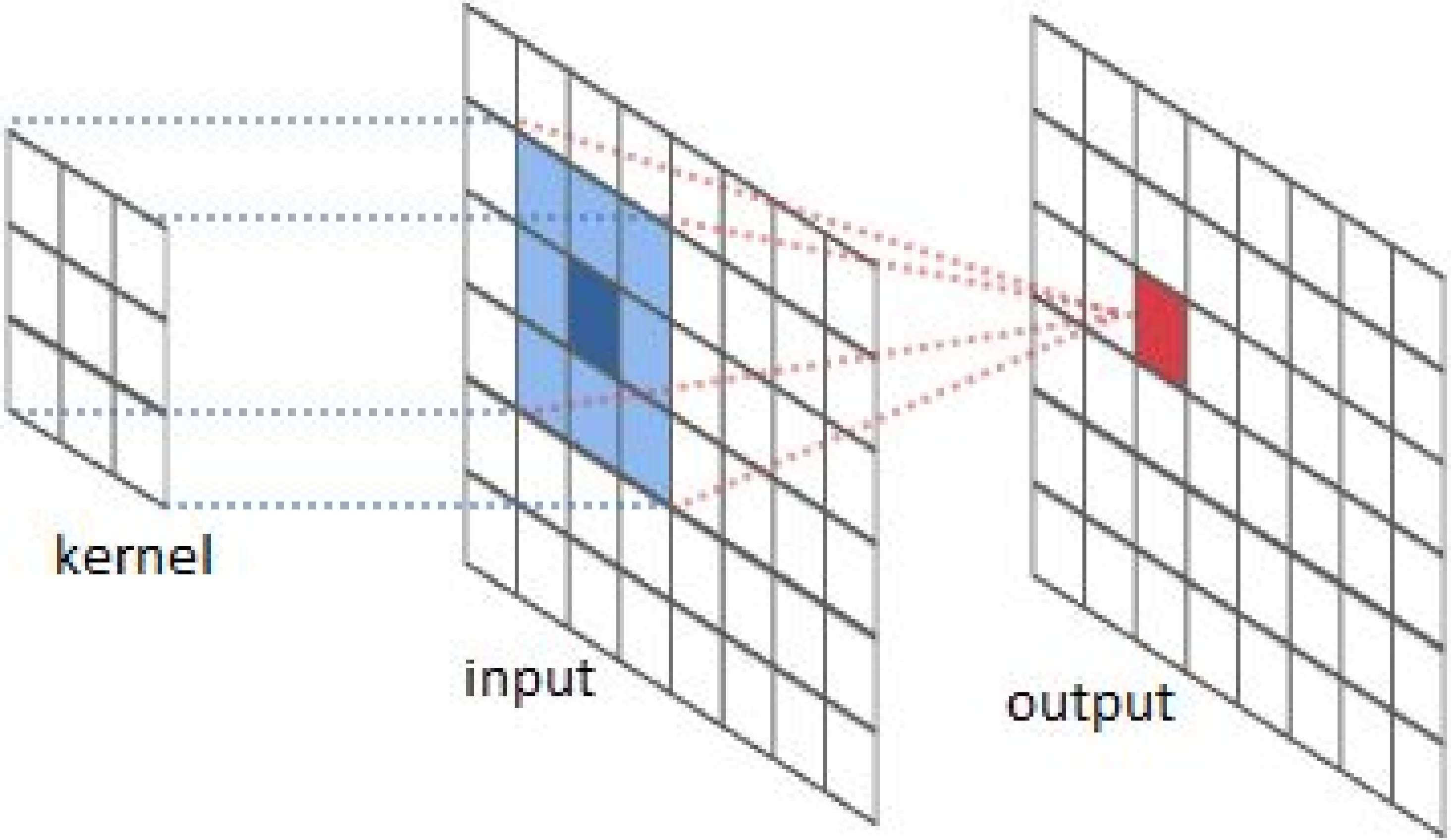
32	16	24	55	...
18	12	99	123	...
44	88	31	99	...
55	94	31	88	...
...	...	...	...	...

MNIST and Grayscale

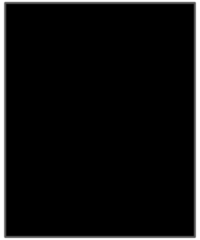


32	16	24	55	...
18	12	99	123	...
44	88	31	99	...
55	94	31	88	...
...	...	...	...	...

The Convolution



The Convolution



Data

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

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

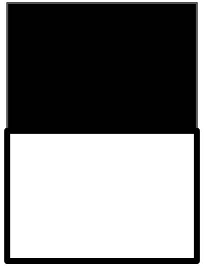
0	**0**	**...**	**...**	**...**
0	**0**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**
...	**...**	**...**	**...**	**...**

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

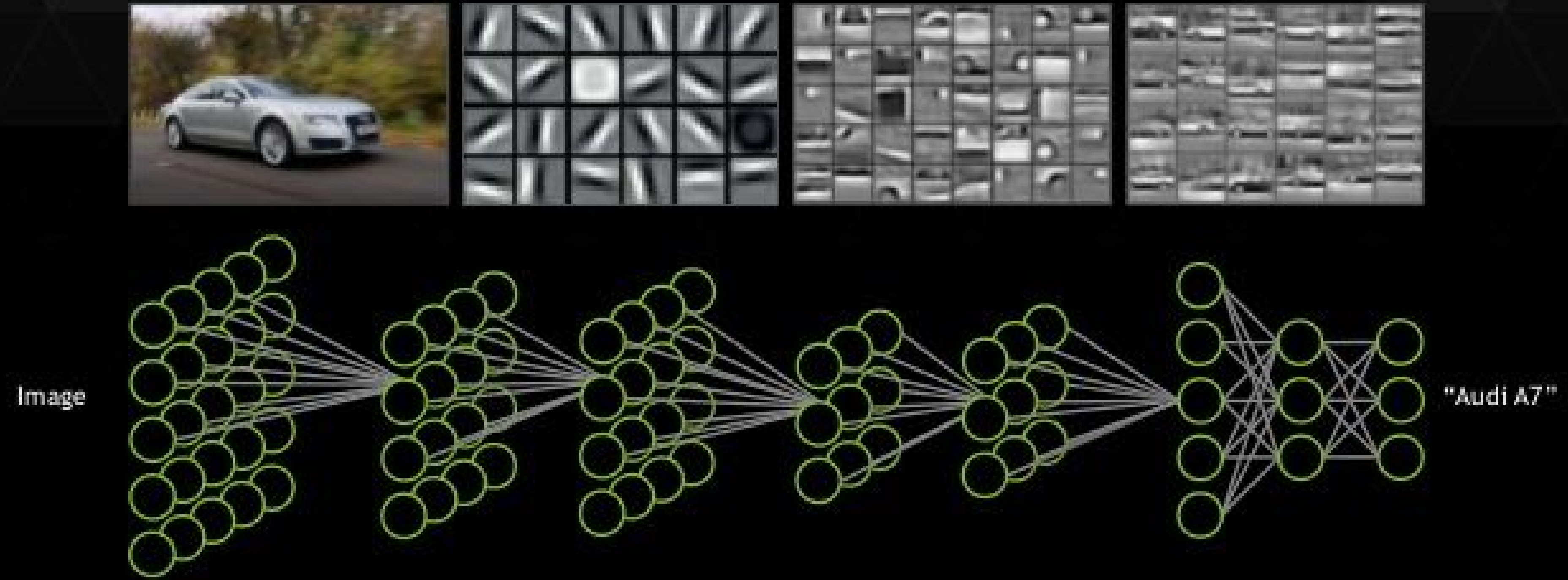
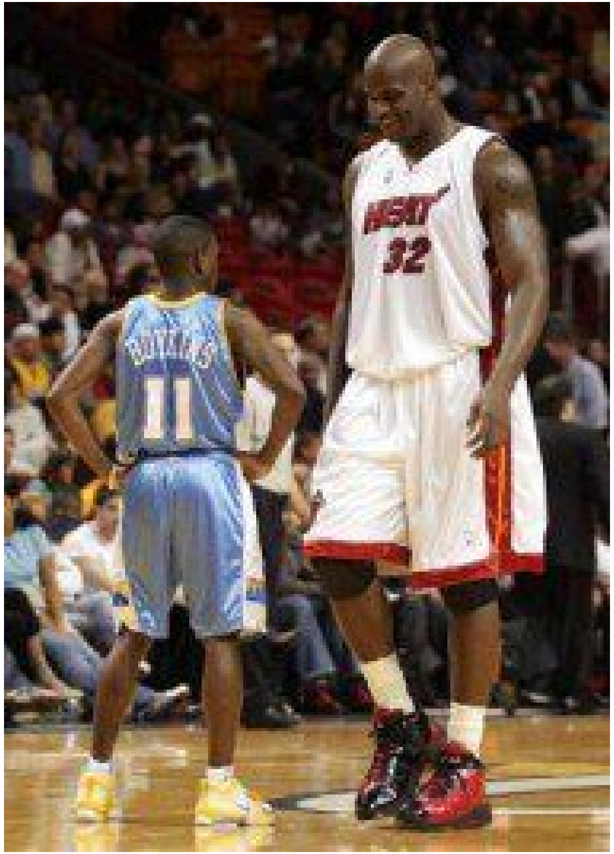


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011. Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

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

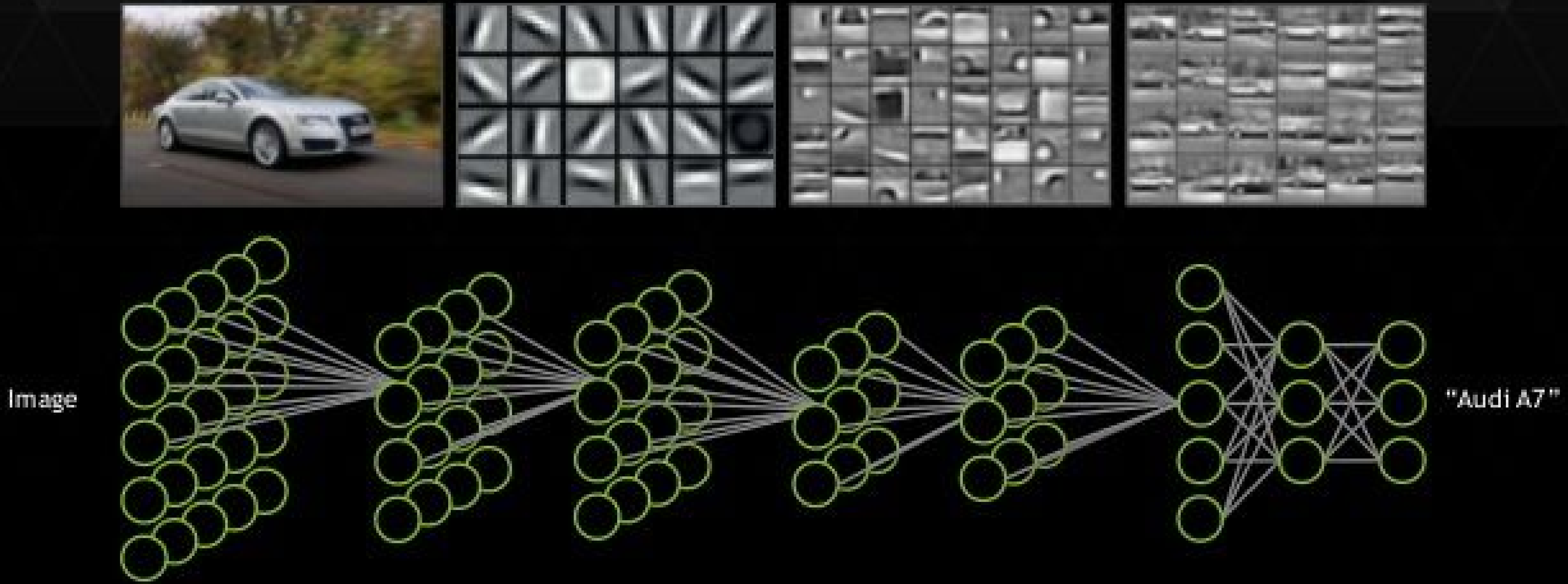


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

Topics

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