A Small Intro in Deeplearning Fast-Start 2018

What are they ?

What can they do ?

How to create them ?

Self Driving Car !

Willem Hendriks
willem.hendriks@nl.ibm.com

IBM

David Robinson

Chief Data Scientist at DataCamp, works in R and Python.

Email
Twitter
Github
Stack Overflow

Subscribe

Your email

Subscribe to this blog

Recommended Blogs

- DataCamp
- R Bloggers
- RStudio Blog
- R4Stats
- Simply Statistics

What's the difference between data science, machine learning, and artificial intelligence?

When I introduce myself as a data scientist, I often get questions like "What's the difference between that and machine learning?" or "Does that mean you work on artificial intelligence?" I've responded enough times that my answer easily qualifies for my "rule of three":

> **David Robinson**
> @drob
>
> When you've written the same code 3 times, write a function
>
> When you've given the same in-person advice 3 times, write a blog post
>
> 3:22 AM - Nov 9, 2017
>
> 46 1,385 3,950

The fields do have a great deal of overlap, and there's enough hype around each of them that the choice can feel like a matter of marketing. But **they're not interchangeable**: most professionals in these fields have an intuitive understanding of how particular work could be classified as data science, machine learning, or artificial intelligence, even if it's difficult to put into words.

So in this post, I'm proposing an *oversimplified* definition of the difference between the three fields:

- **Data science** produces **insights**
- **Machine learning** produces **predictions**
- **Artificial intelligence** produces **actions**

To be clear, this isn't a *sufficient* qualification: not everything that fits each definition is a part of that field. (A fortune teller makes predictions, but we'd never say that they're machine learning!) These also aren't a good way of determining someone's role or [illegible] of focus and experience. (This is true

Home > Data Science > Machine Learning

Neural Networks and Deep Learning

About this course: If you want to break into cutting-edge AI, this course will help you do so. Deep learning engineers are highly sought after, and mastering deep learning will give you numerous new career opportunities. Deep learning is also a new "superpower" that will let you build AI systems that just weren't possible a few years ago.

In this course, you will learn the foundations of deep learning. When you finish this class, you will:
- Understand the major technology trends driving Deep Learning
- Be able to build, train and apply fully connected deep neural networks
- Know how to implement efficient (vectorized) neural networks
- Understand the key parameters in a neural network's architecture

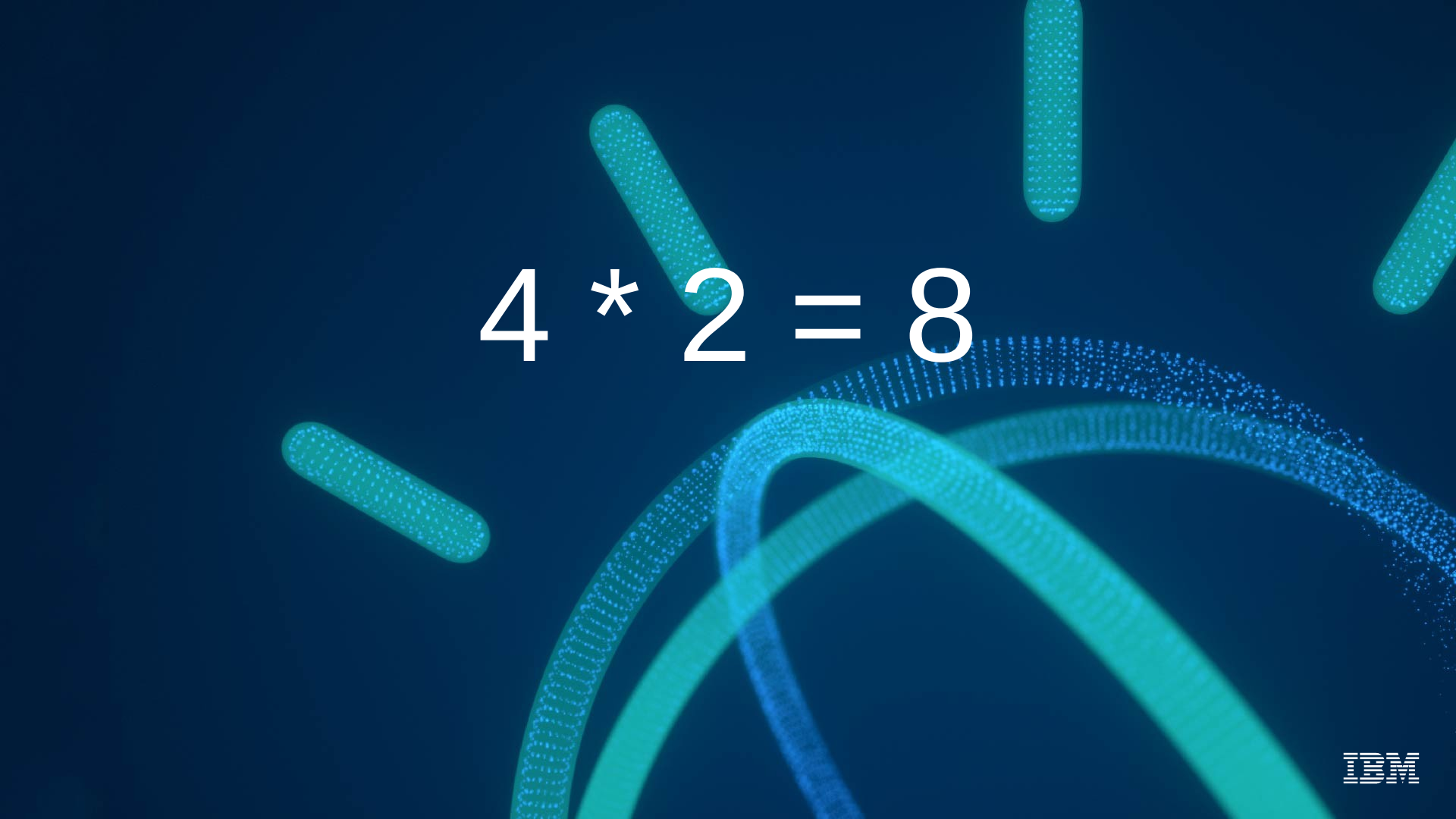
This course also teaches you how Deep Learning actually works, rather than presenting only a cursory or surface-level description. So after completing it, you will be able to apply deep learning to a your own applications. If you are looking for a job in AI, after this course you will also be able to answer basic interview questions.

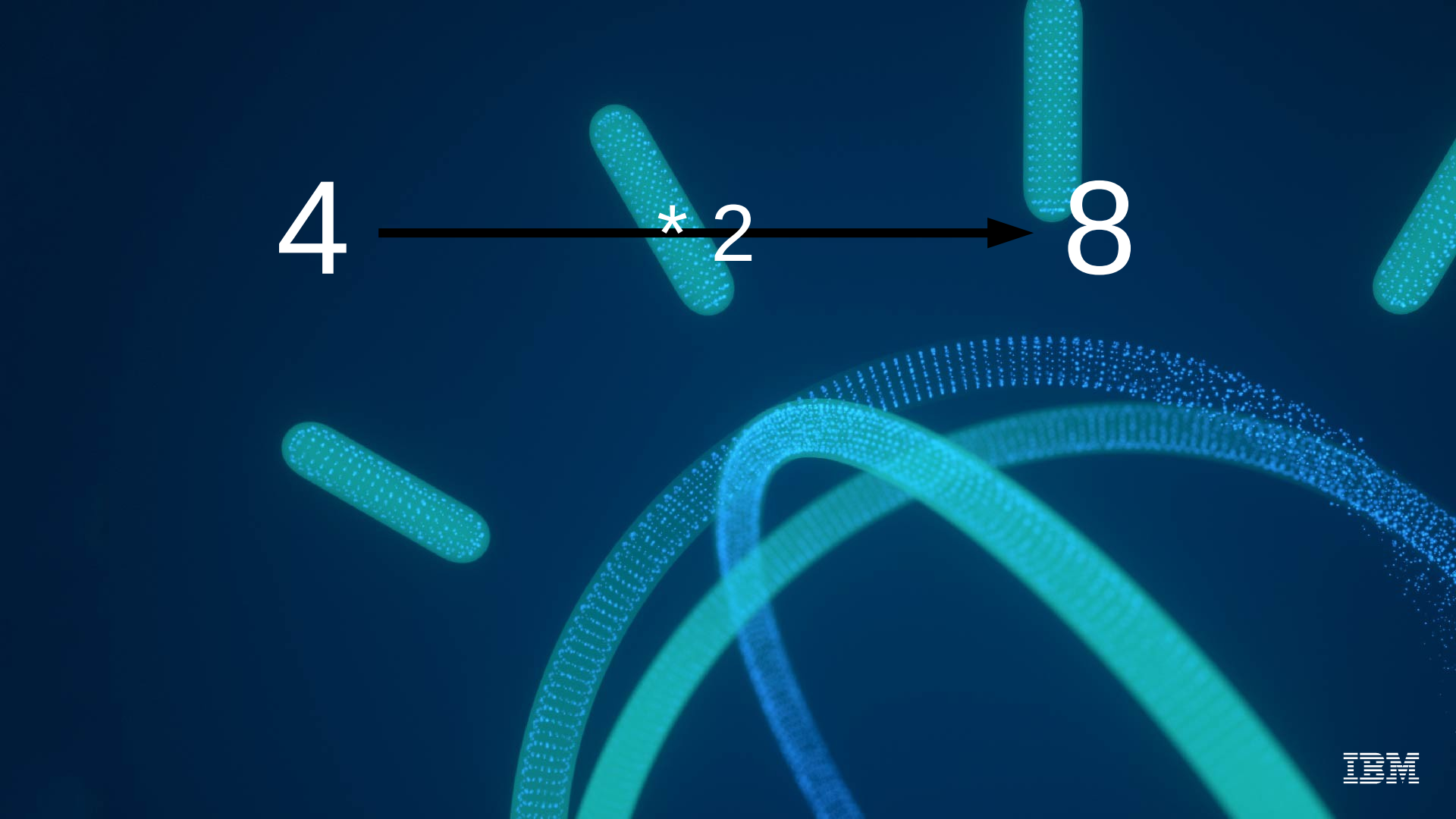
This is the first course of the Deep Learning Specialization.

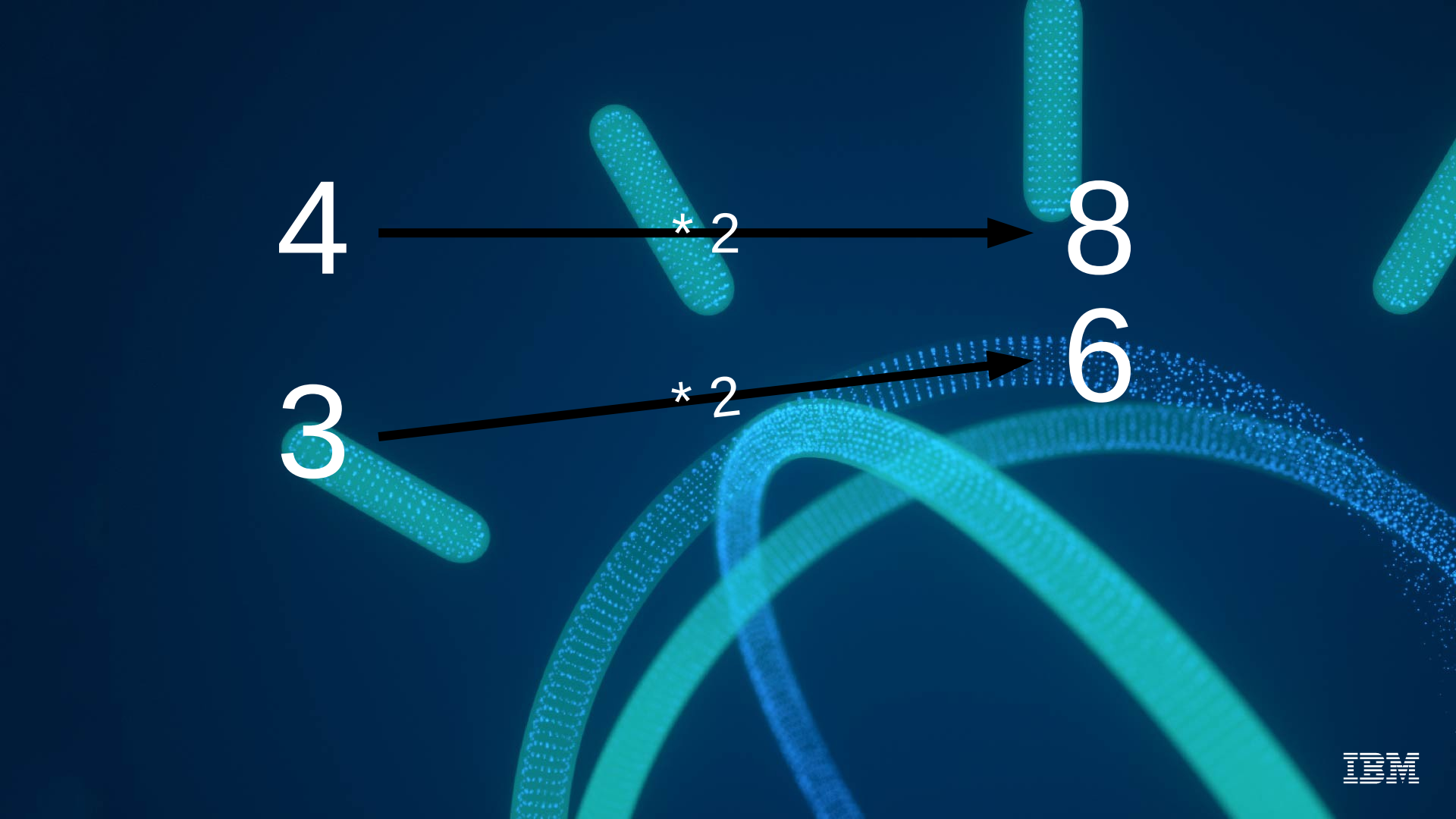
Show less

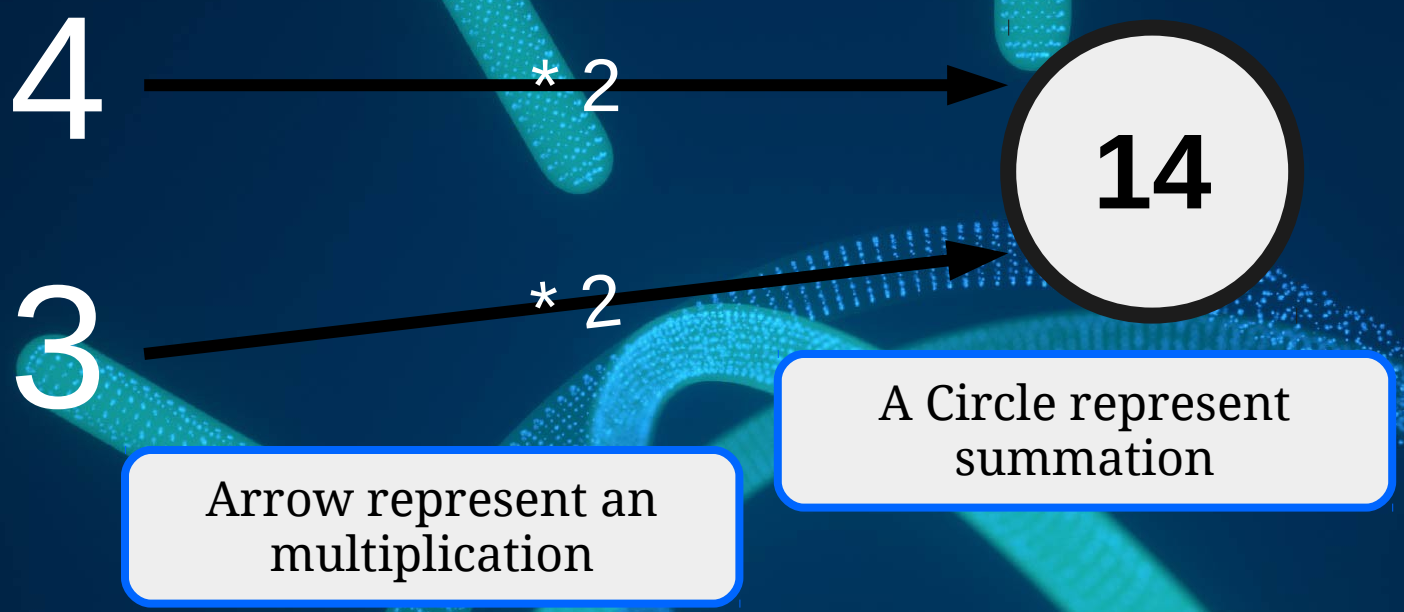
Who is this class for: Prerequisites: Expected: - Programming: Basic Python programming skills, with the capability to work effectively with data structures. Recommended: - Mathematics: Matrix vector operations and notation. - Machine Learning: Understanding how to frame a machine learning problem, including how data is represented will be beneficial. If you have taken my Machine Learning course here, you have much more than

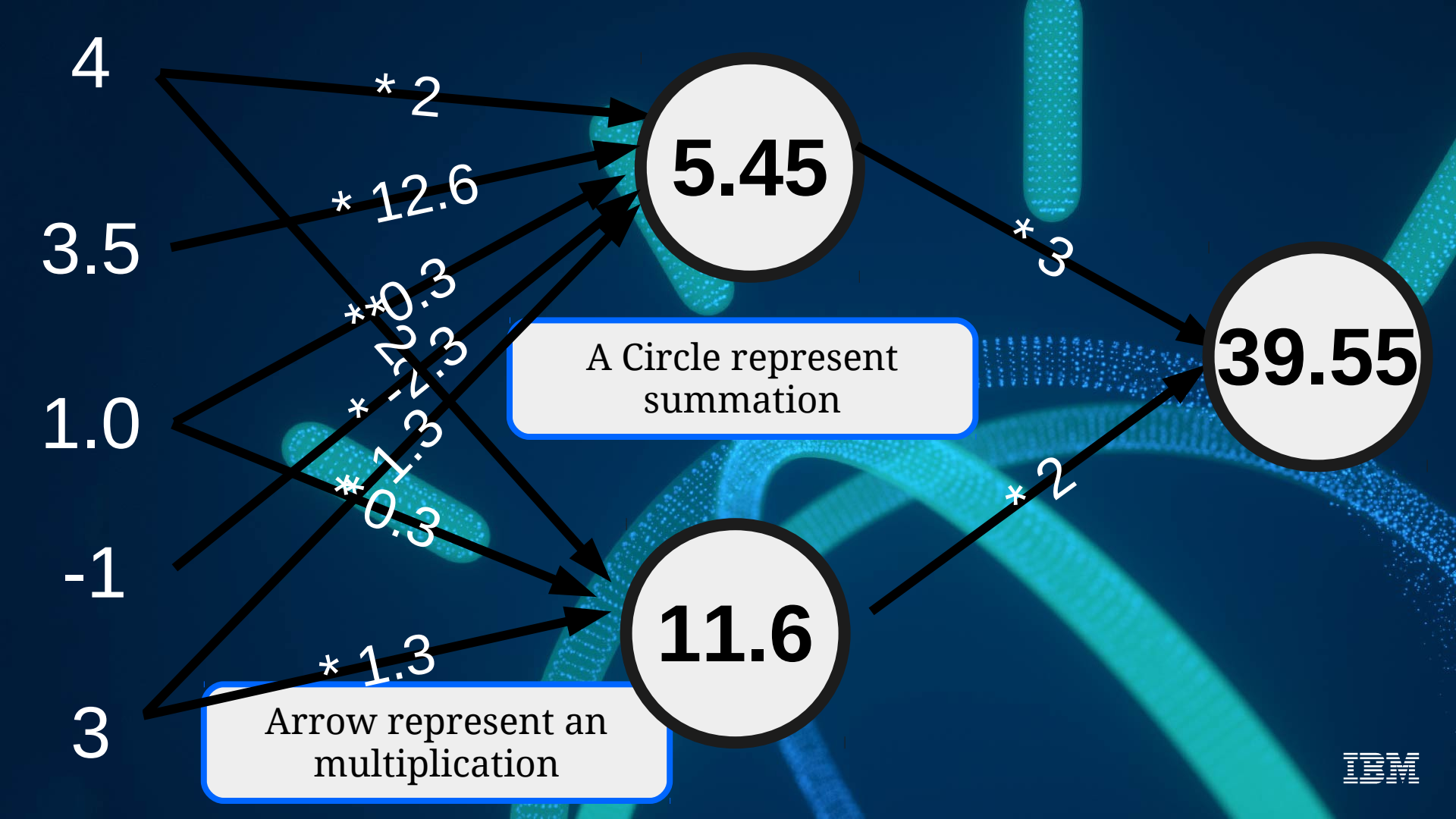
IBM


4 * 2 = 8


4
* 2
8


4
* 2
8
3
* 2
6


4
* 2
14
3
* 2
Arrow represent an multiplication
A Circle represent summation


4
3.5
1.0
-1
3
* 2
* 12.6
**0.3
* -2.3
* 1.3
*0.3
* 1.3
5.45
11.6
* 3
* 2
39.55
A Circle represent summation
Arrow represent an multiplication
IBM

After the summation,

we can apply a function

5.45

Vanilla – keep it as it is

5.45

-1.5

Weather in Madrid - Never below 0.0

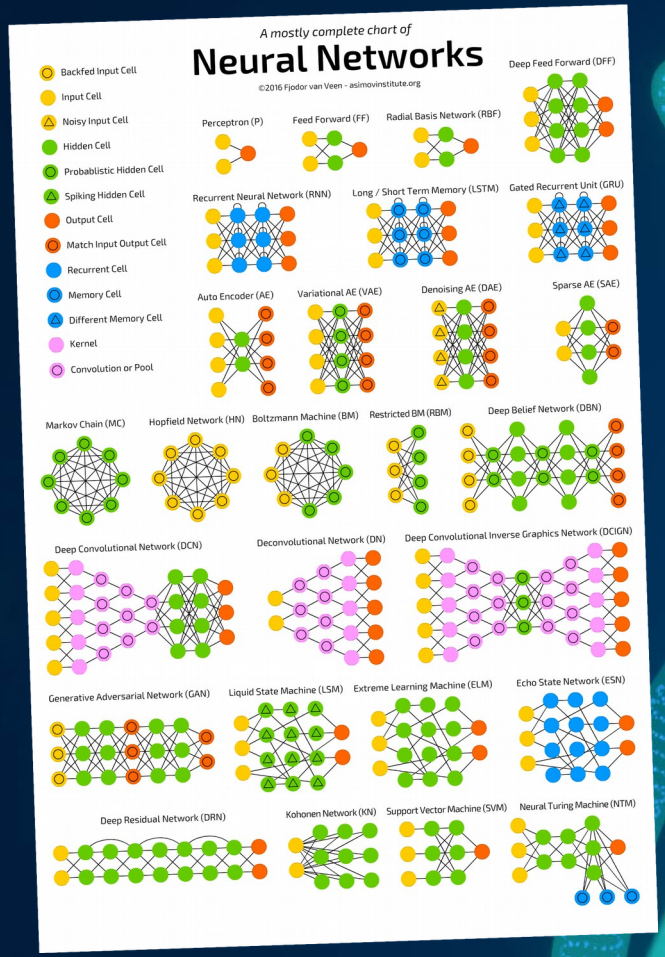
0

34.4

Thor! Smack extreme numbers to between 0.0 and 1.0

0.9

IBM



Many Different Design.

Each day, people experiment with new designs

Each week, papers appear with test results

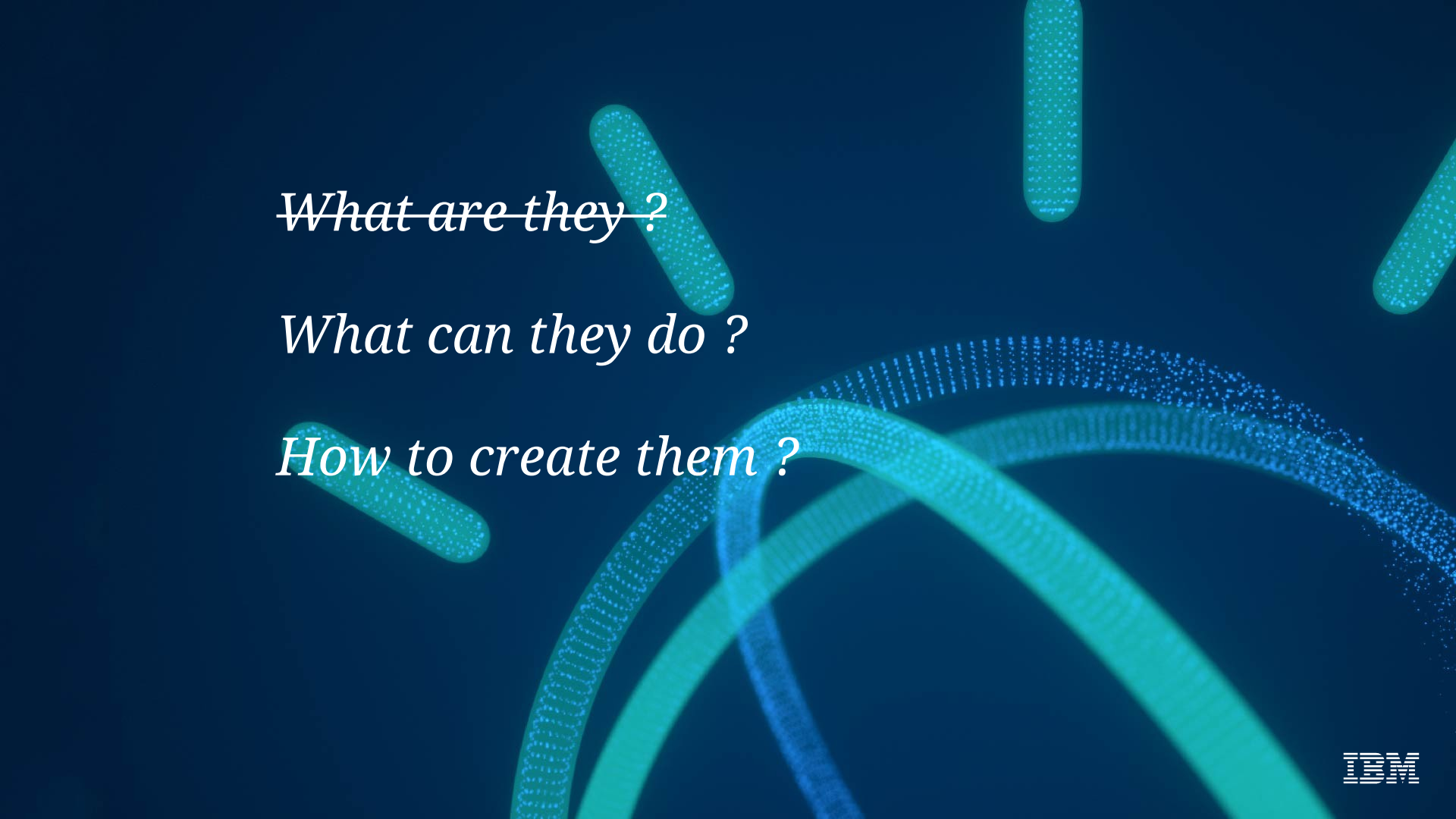
Designs on left have proven to be useful – each has its special powers

IBM

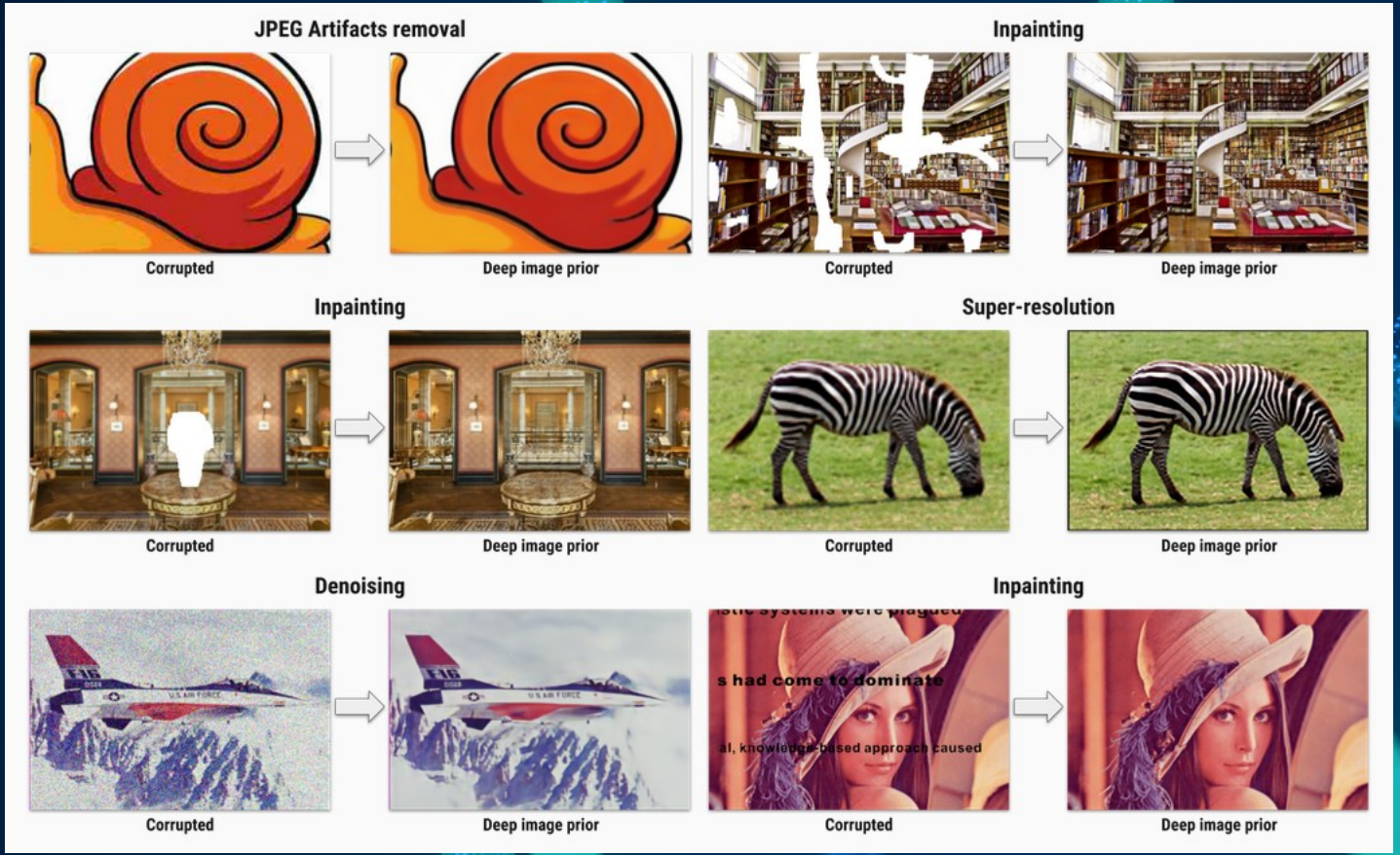
~~*What are they ?*~~

What can they do ?

How to create them ?

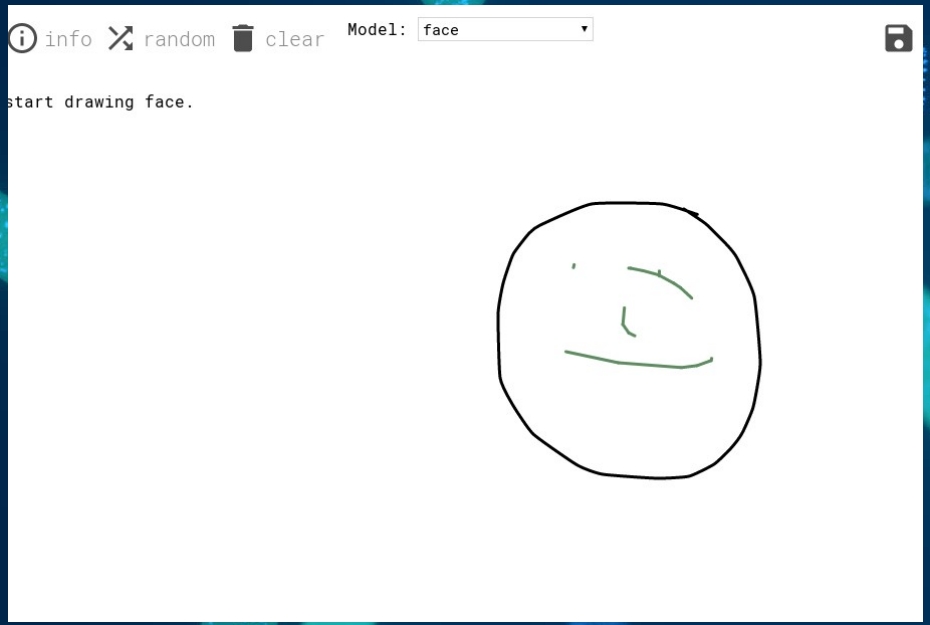


Fix a corrupted Image
https://dmitryulyanov.github.io/deep_image_prior



Networks Can Draw Objects – And help to draw

https://magenta.tensorflow.org/assets/sketch_rnn_demo/index.html



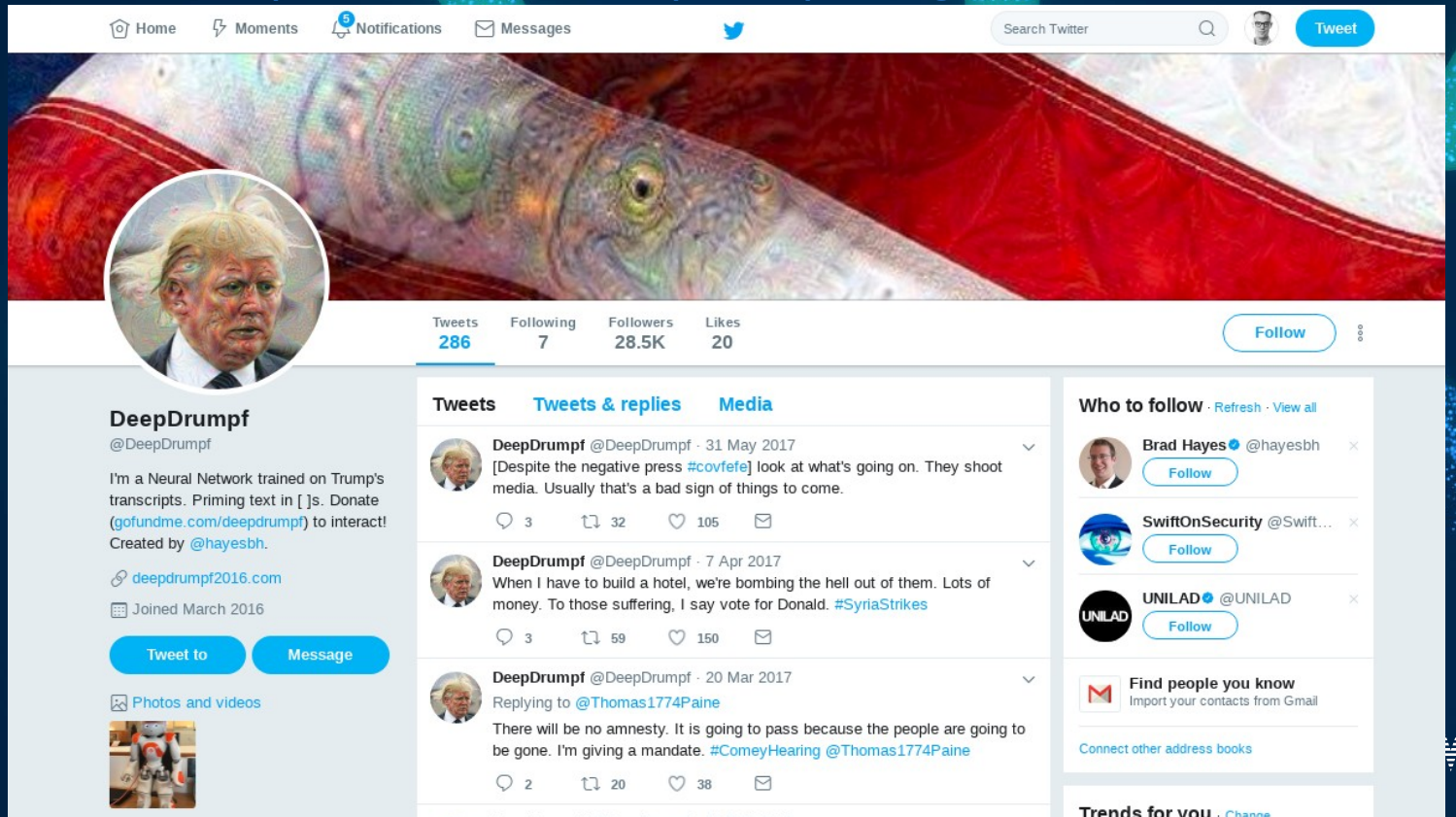
Networks can Play Music, Like Chopin

https://www.youtube.com/watch?v=j60J1cGINX4

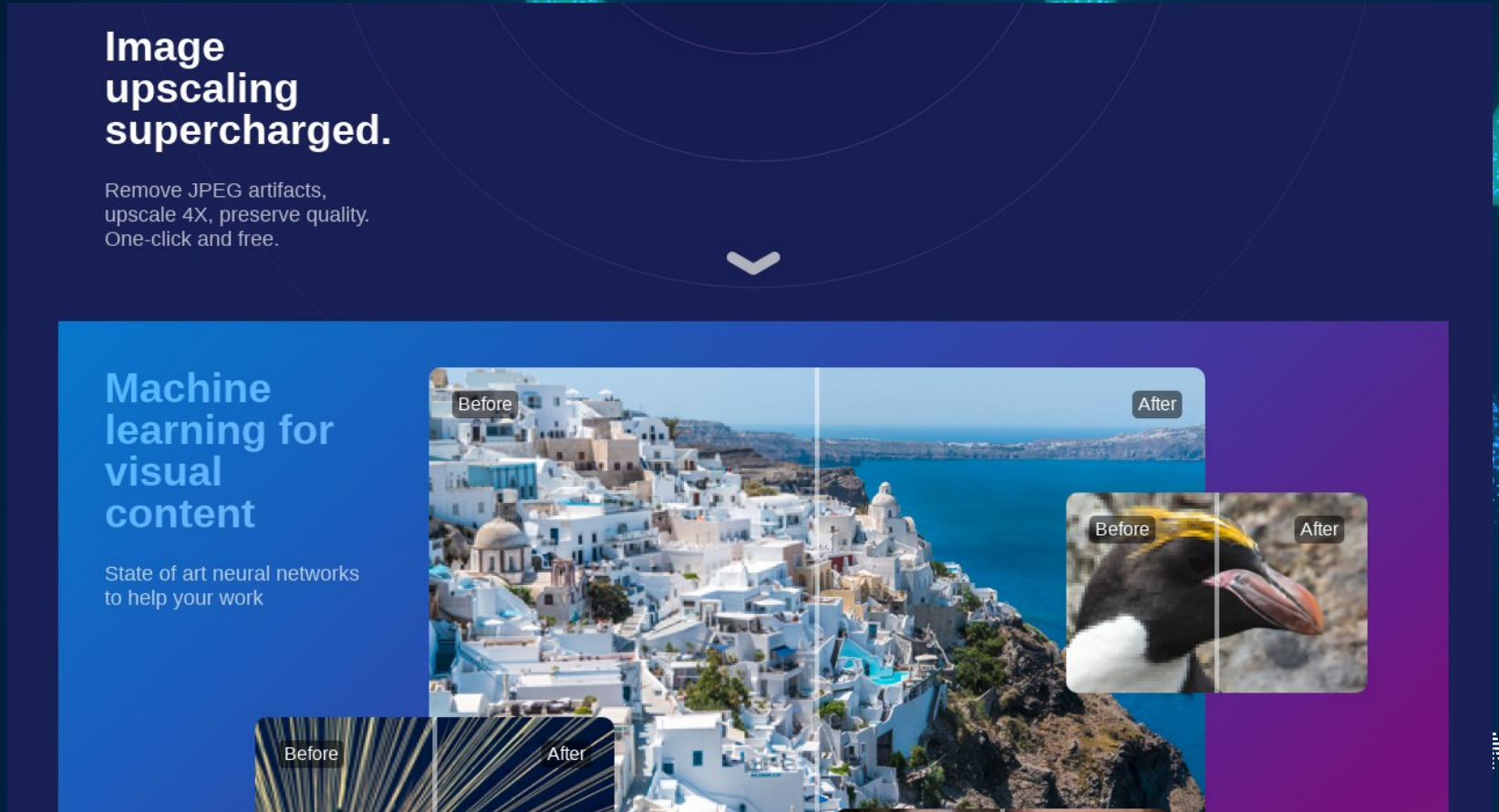
IBM

Networks can Tweet like Trump

https://twitter.com/deepdrumpf?lang=en



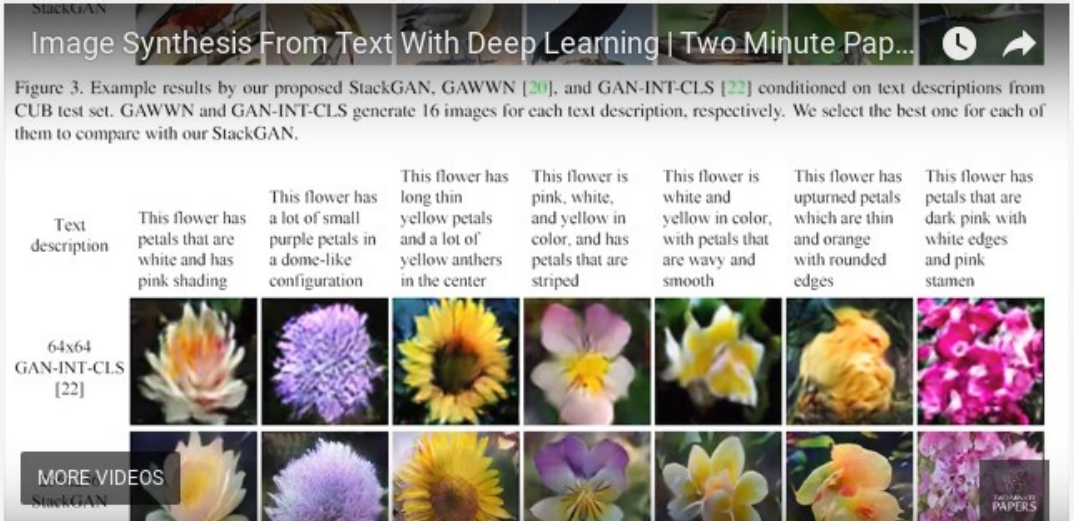
Do a CSI: “Enhance!”
https://letsenhance.io/



Neural Networks can to Text-To-Image

Neural network can create high-res images based on a text description

By **Luke Dormehl** — Posted on December 30, 2016 3:06 pm



Neural Networks can to Text-To-Speech

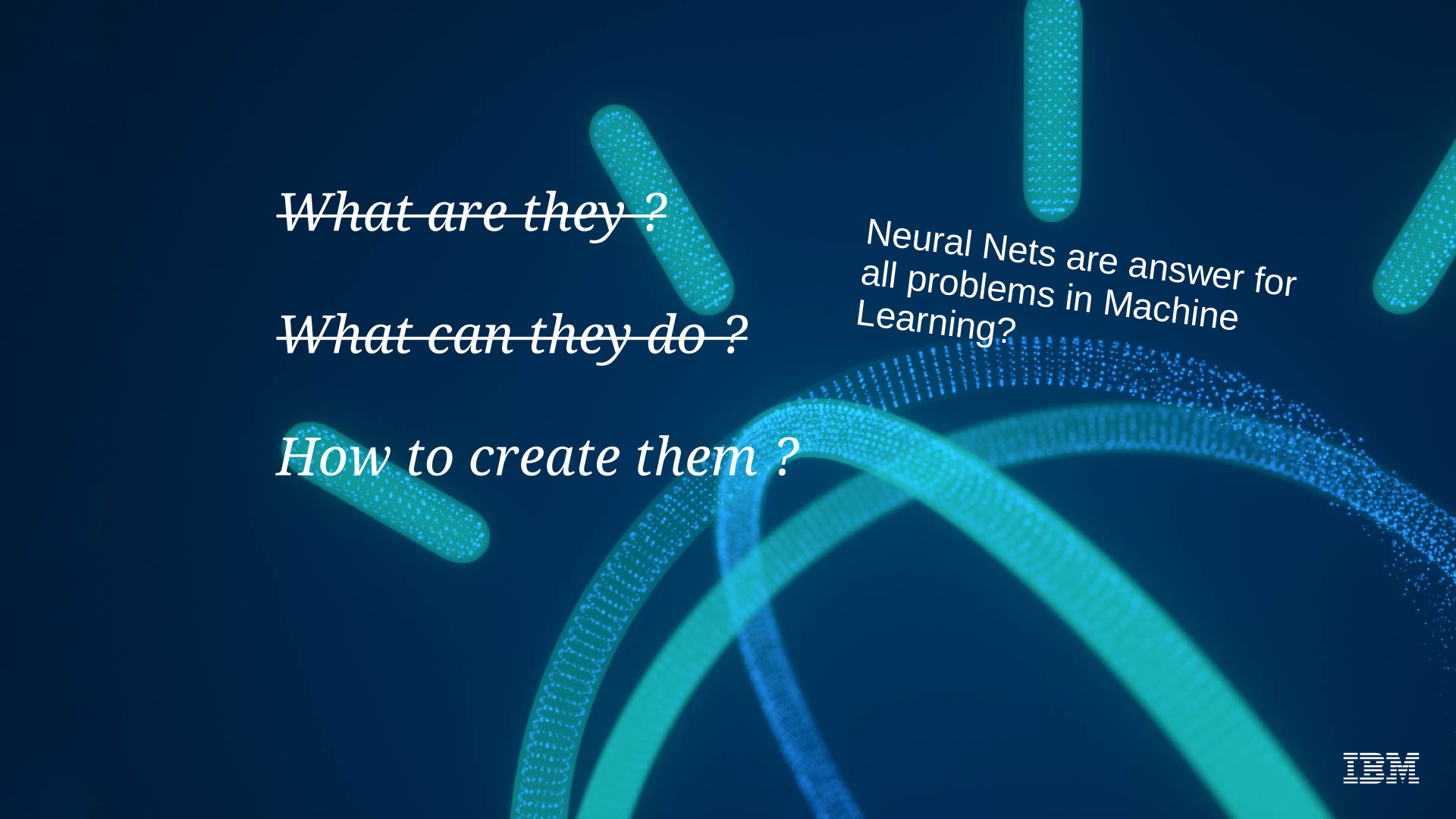
https://www.youtube.com/watch?v=EODKgrYMoS0

~~*What are they ?*~~

~~*What can they do ?*~~

How to create them ?

Neural Nets are answer for all problems in Machine Learning?



Measure Performance

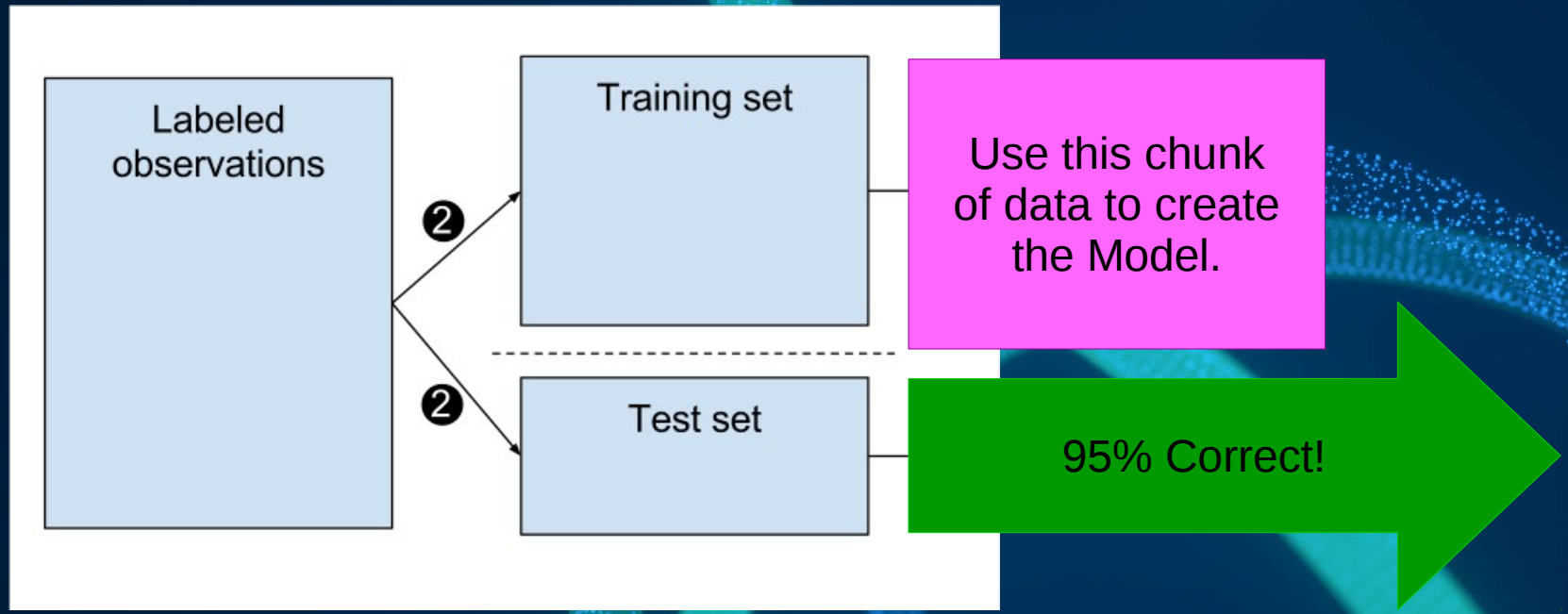
Machine Learning - Predicting:

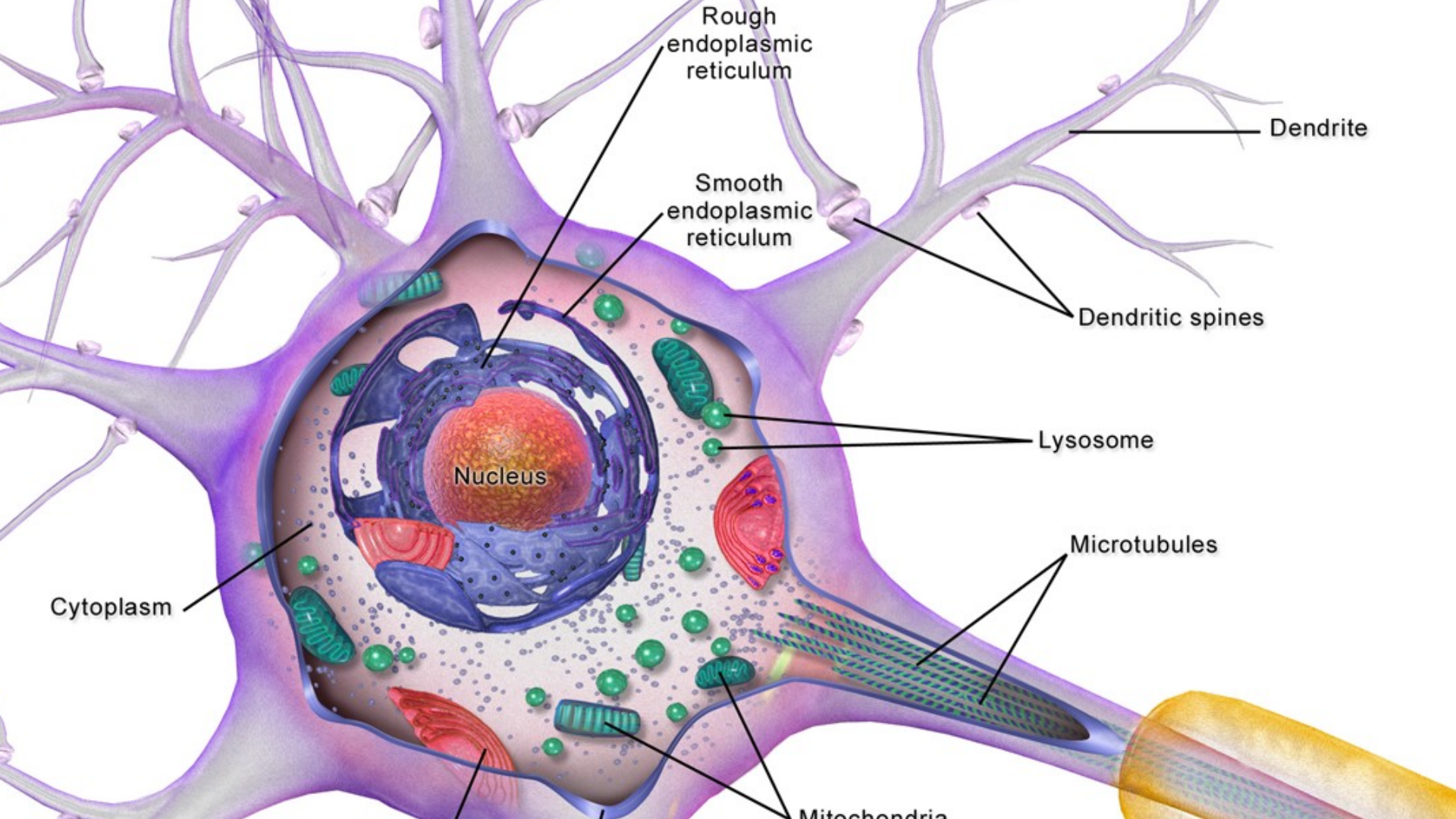
When we apply a Model to predict a value, we want to measure the performance.

How do we often do that?

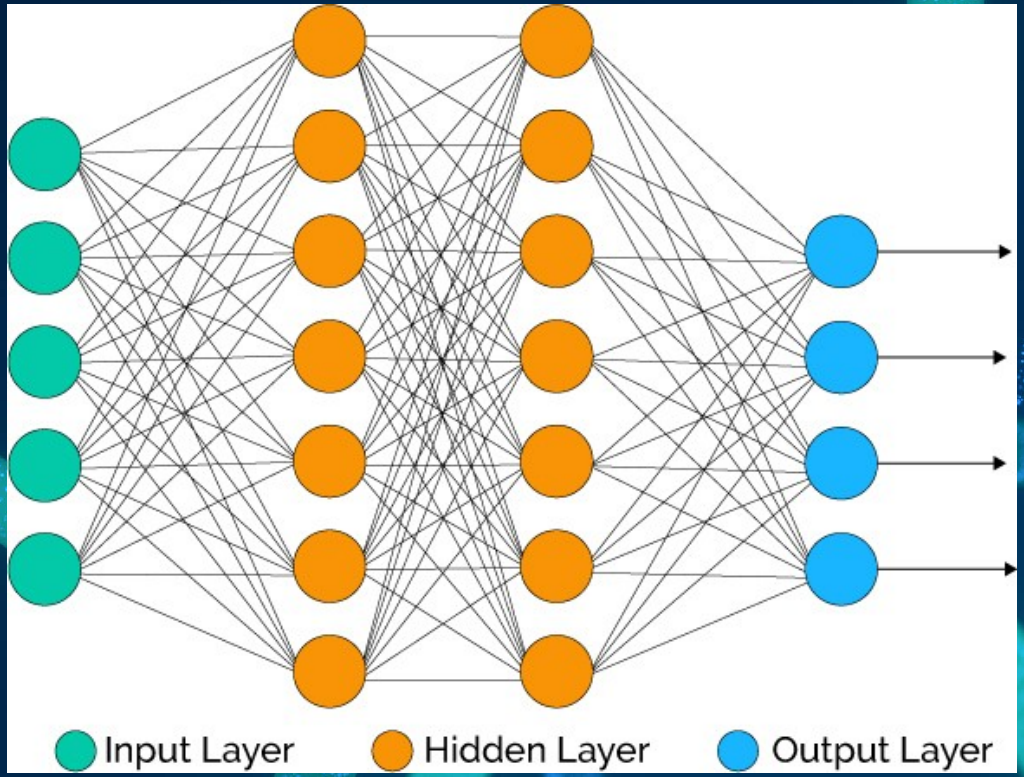
IBM

Train/Test Set




Rough endoplasmic reticulum
Dendrite
Smooth endoplasmic reticulum
Dendritic spines
Lysosome
Nucleus
Microtubules
Cytoplasm
Mitochondria

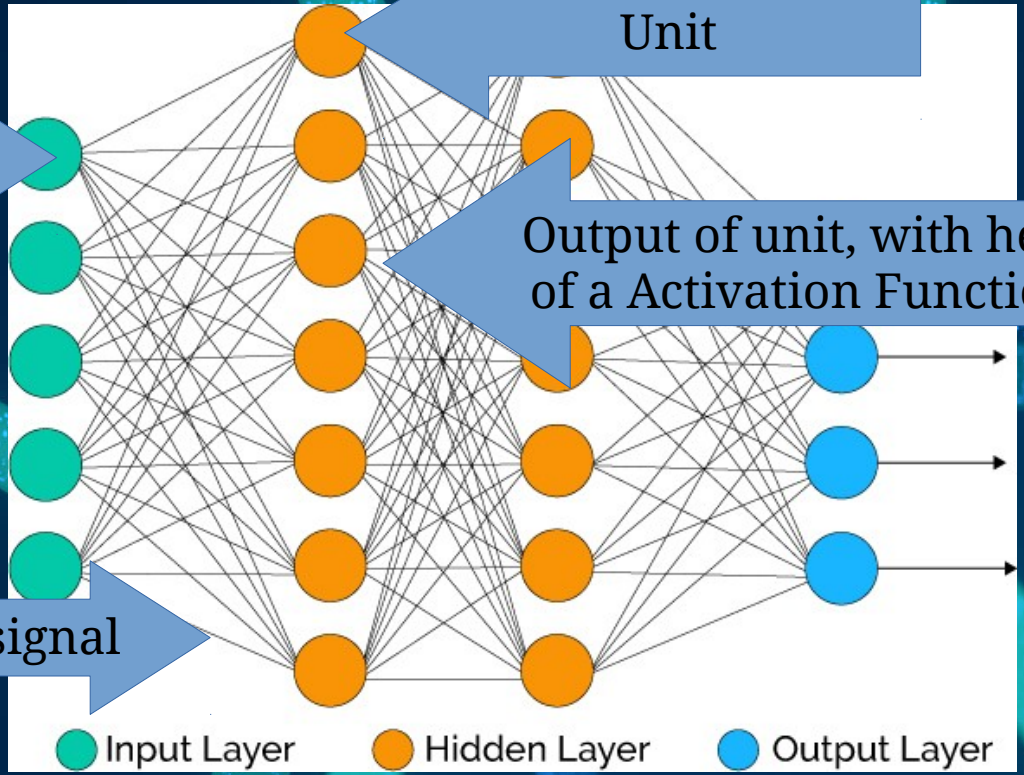
Artificial Neural Network (ANN)



The more Hidden Layers, The more Deep-learning.
1 Layer sometimes represents Classical Models

IBM

Artificial Neural Network (ANN)



Artificial Neural Network (ANN)

Feature, e.g. Age

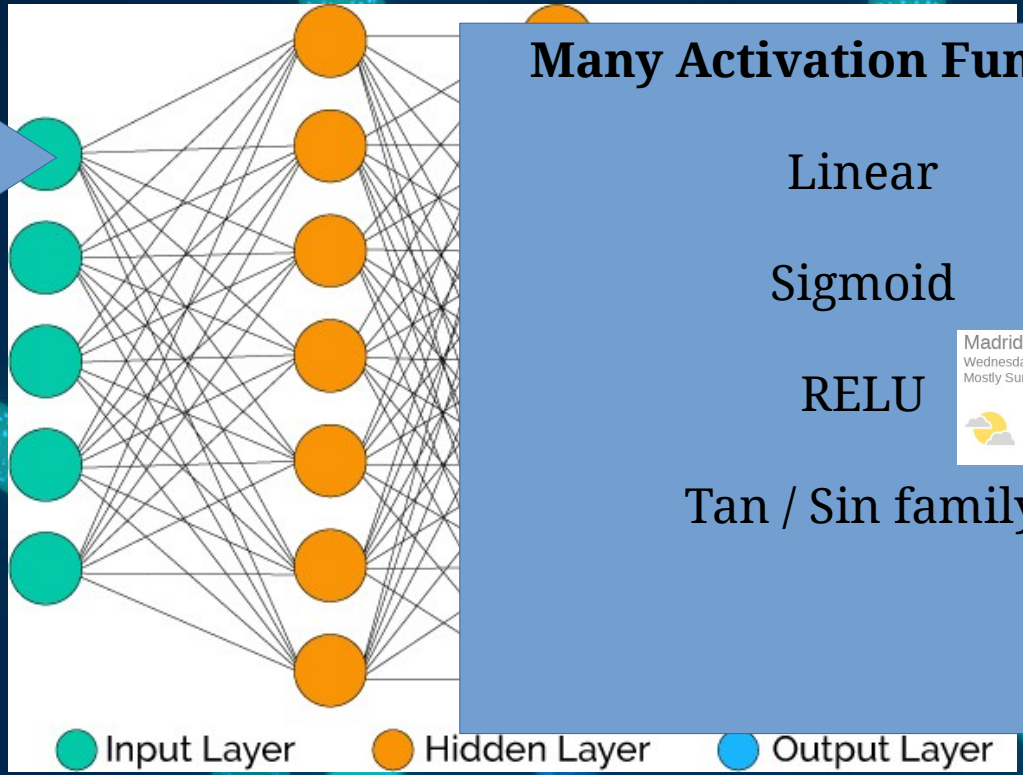
Pixel color value

Black White:
28x28 = 784 inputs

Color Picture:
28x28x3 = 2352 inputs

Voices, Text, Videos
(Can all be numbers)

Values, Means, Vars,
Differences, *, +



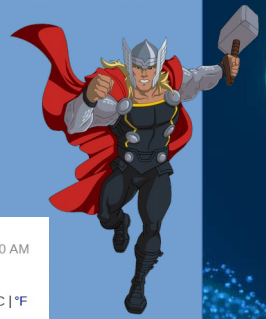
Many Activation Functions:

Linear

Sigmoid

RELU

Tan / Sin family



Artificial Neural Network (ANN)

Common Activation Functions

Madrid
Wednesday 10:00 AM
Mostly Sunny
6 °C | °F

Sigmoid

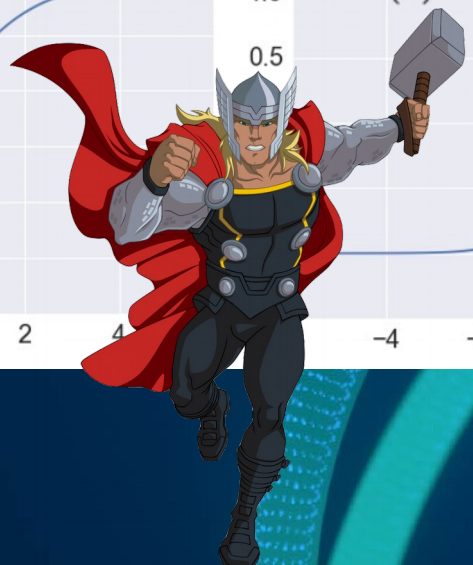
$$f(x) = \frac{1}{1 + e^{-x}}$$

TanH

$$\tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

ReLU

$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$$

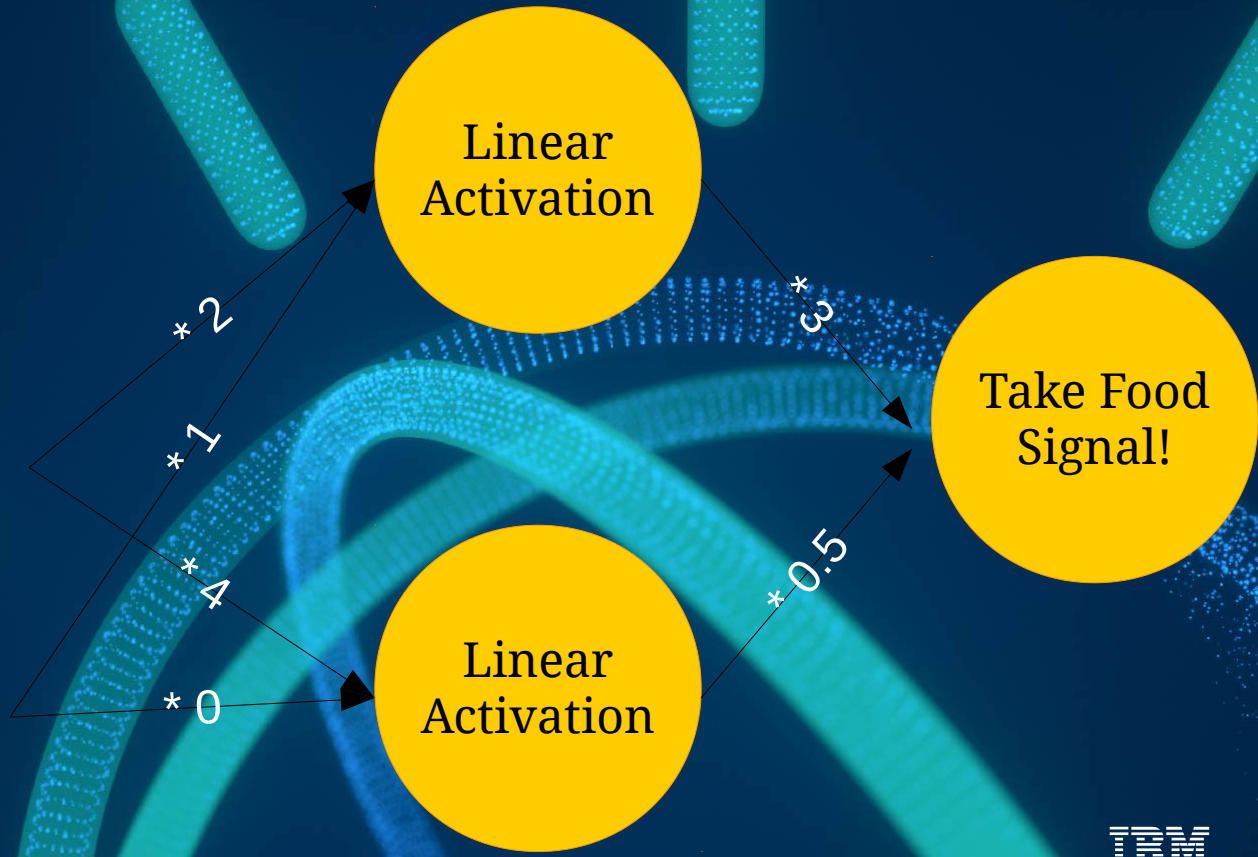


IBM

Example - Snail

Feature 1: Hunger Level
Input H: 0.5

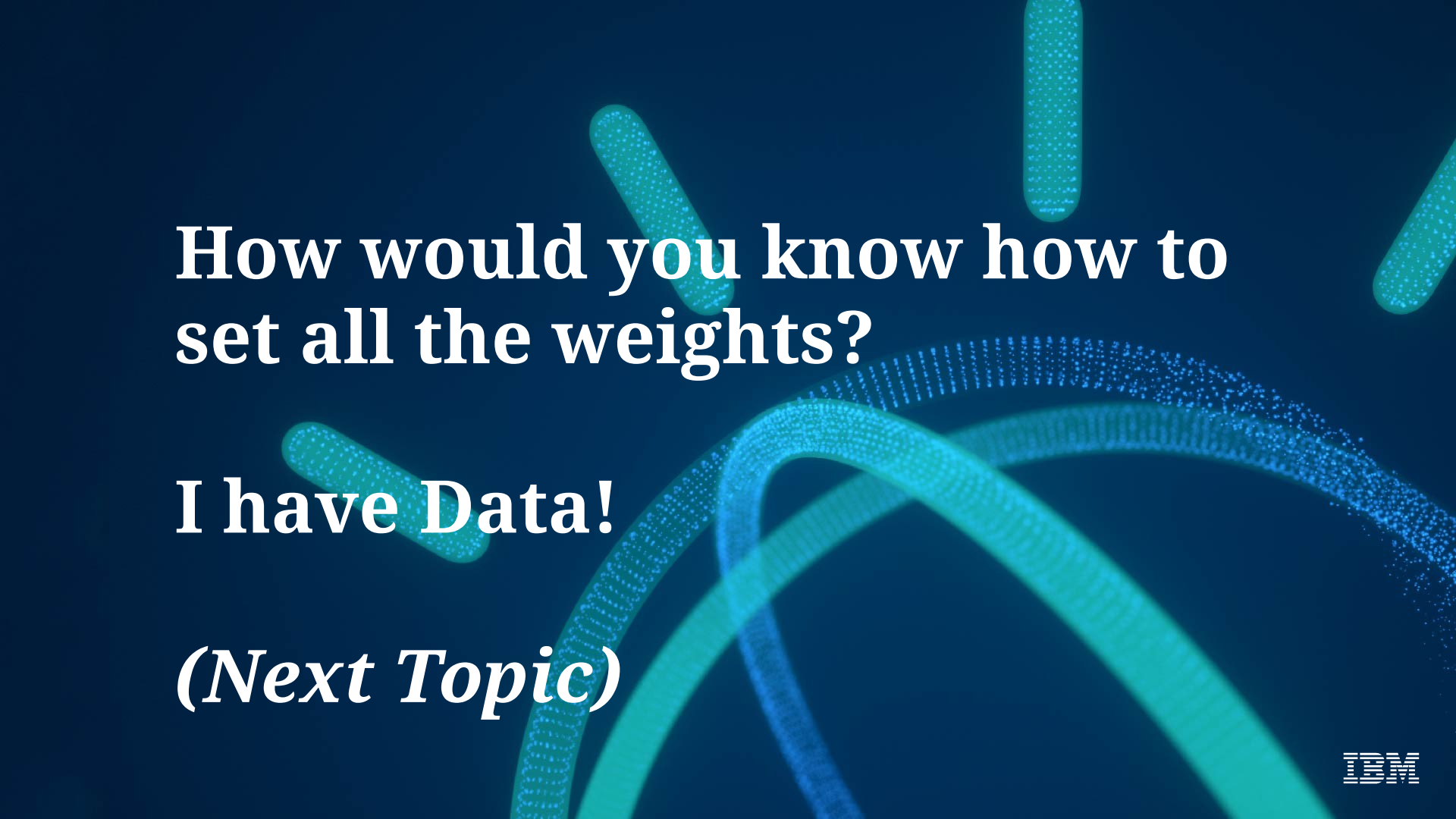
Feature 2: Sees food in Range Level
Input S: 1



IBM

What is the Most Used Activation Function?

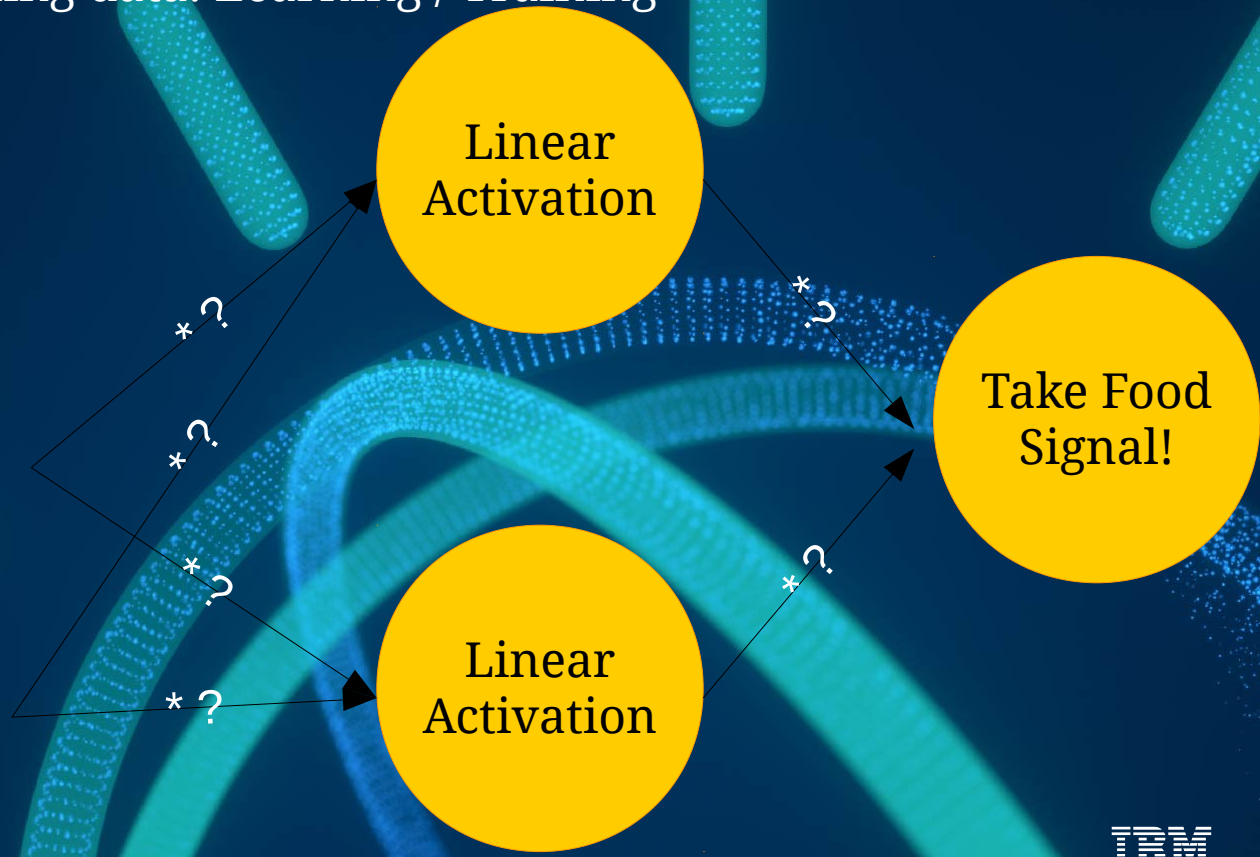
Which Activation mimics our Brain the Best?


How would you know how to set all the weights?
I have Data!
(Next Topic)

Weights need to be set, such that the network performs well, on our available Training data. Learning / Training

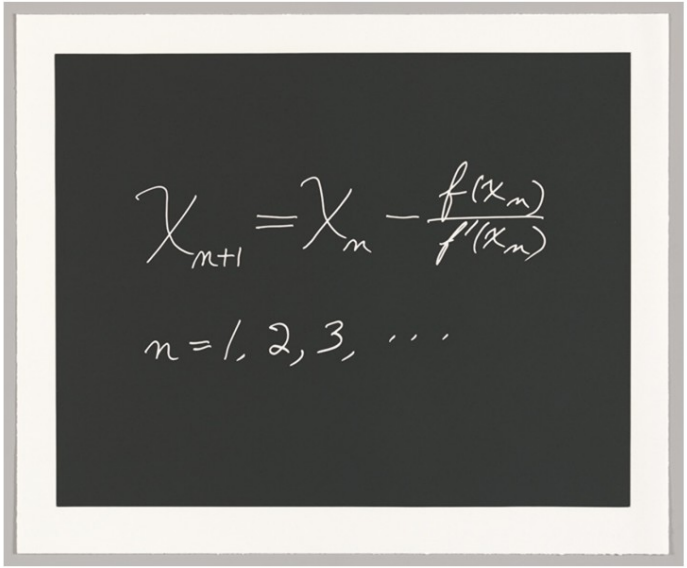
Feature 1: Hunger Level
Input H: 0.5

Feature 2: Sees food in Range Level
Input S: 1

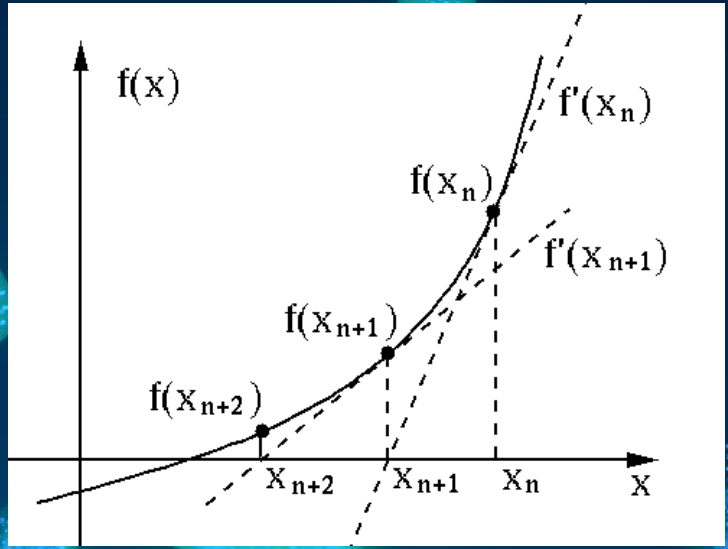


IBM

Stephen Smale: "Newton's Method"



Stephen Smale (American, born 1930). *Concinnitas*, 2014

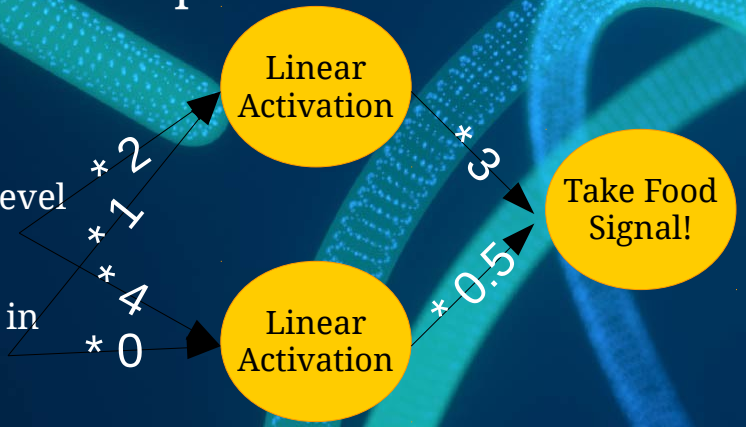


During the training Phase, of a Neural Network,

Data is fired through the network, and with help of the derivative, a **optimizer** will find the (almost) best **weights**

1. Take a data point
2. Calculate output
3. Use output error (in combination with derivative) to re-calibrate all weights
4. repeat , with next data point

Example - Snail



Feature 1: Hunger Level
Input H: 0.5

Feature 2: Sees food in Range Level
Input S: 1

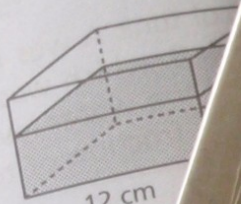
Network Output: 6

Labeled Data: 10

Error: 4
Tweak all weights to reduce error

IBM

ains water filled to $\frac{3}{4}$ of its height.
be the water level in the tank,
m^3 of water is poured out from the



e volume = length × breadth × height,

$$\text{Height} = \left(\frac{\text{volume}}{\text{length} \times \text{breadth}}\right)$$

n:

l 1:

of water in the tank $= \frac{3}{4} \times 8$

$= 6$ cm

e of water in the tank $= 12 \times 10$

$= 720$ cr

water poured out $= 60$ c

$= 72$

Take Note

ember that 1 l = 1000 cm^3 or 1000 ml. For instance in Example
ay give the initial volume of water in tank A as 1.728 l.
1000 = 1.728 l)

counting cubes. This will help you to
ting the same cube twice. One
nt vertically.

, sometimes
or calculation
xample 2(b)—

ks are awarded for
simplified answer (as
o make problem-solvin

ght given the volume,

r breadth, if the other
ave been included for

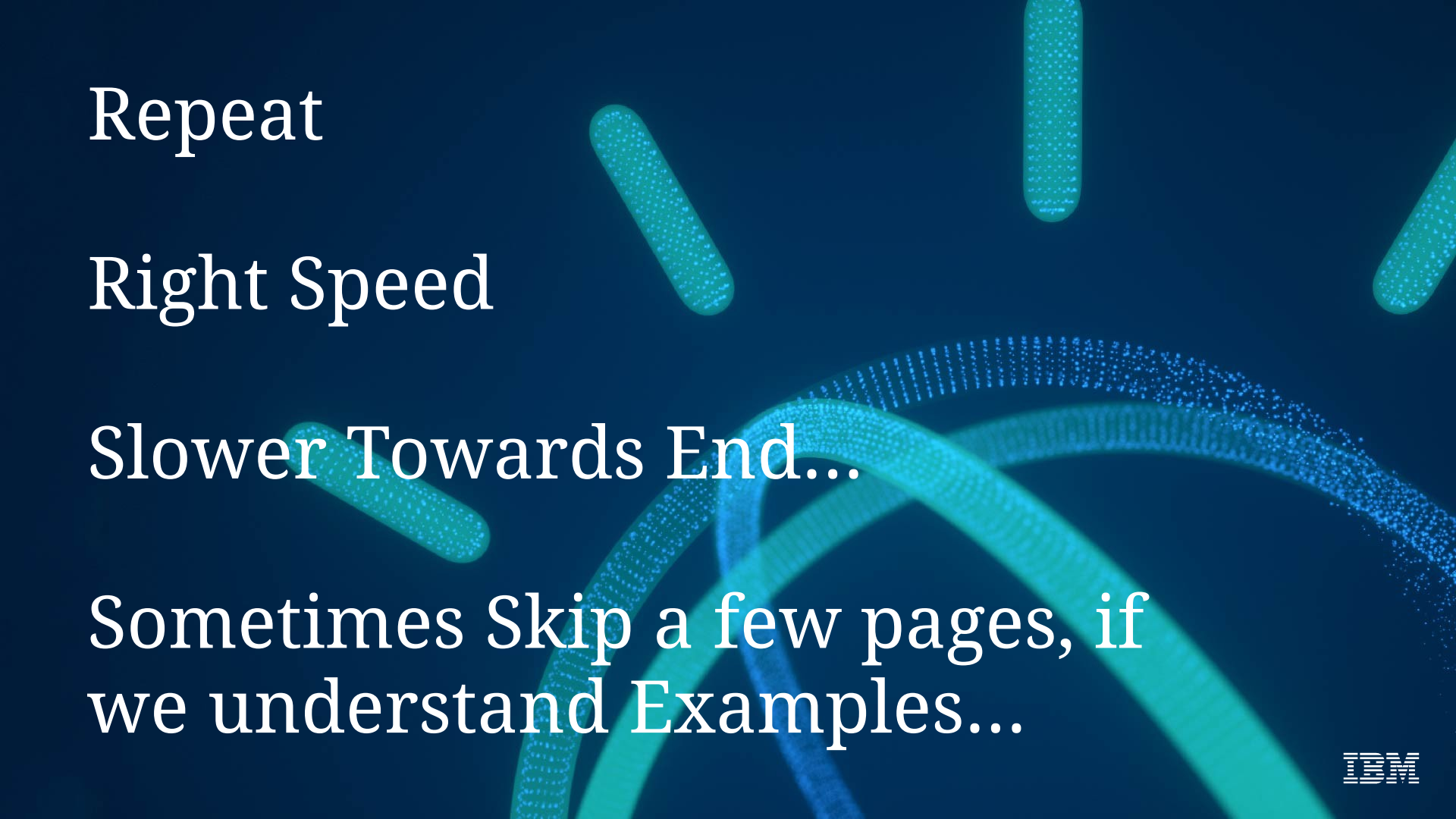
crease) in height may be co
nple 3. Using this method
orking.

Repeat

Right Speed

Slower Towards End…

Sometimes Skip a few pages, if we understand Examples…



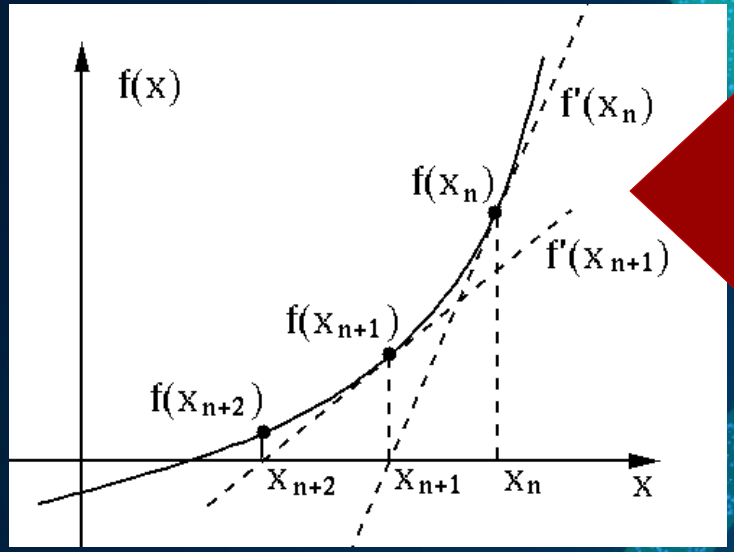
Repeat, **Epochs**

Right Speed, **Learning Rate**

Slower Towards End… **Decay**

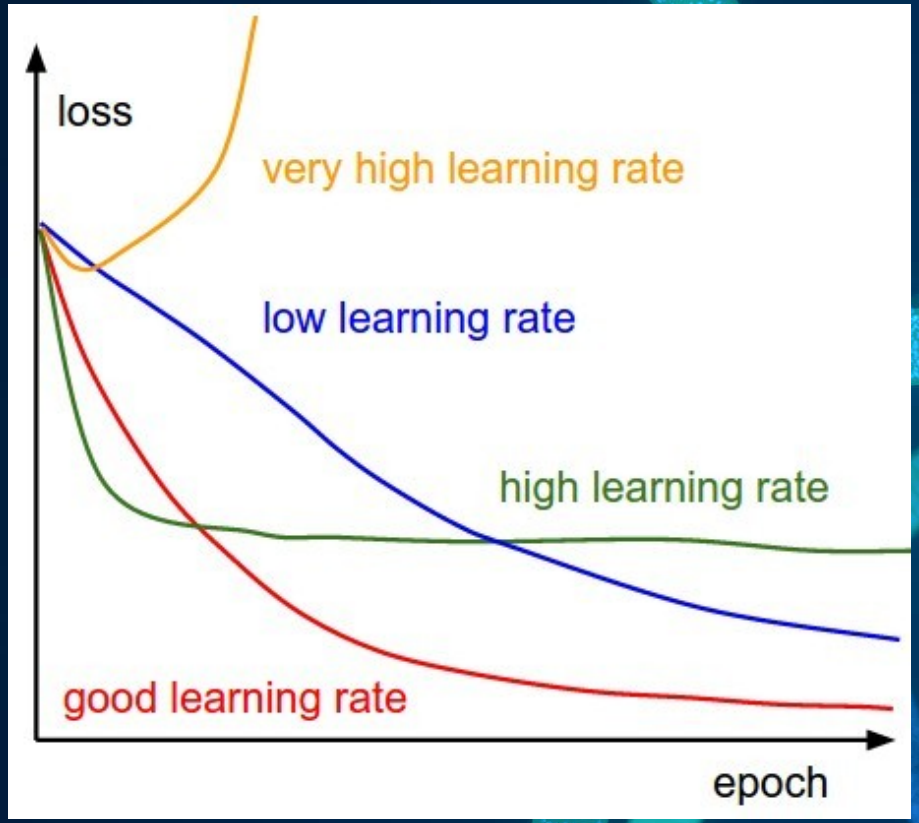
Sometimes Skip a few pages **Momentum**

Right Speed, **Learning Rate**



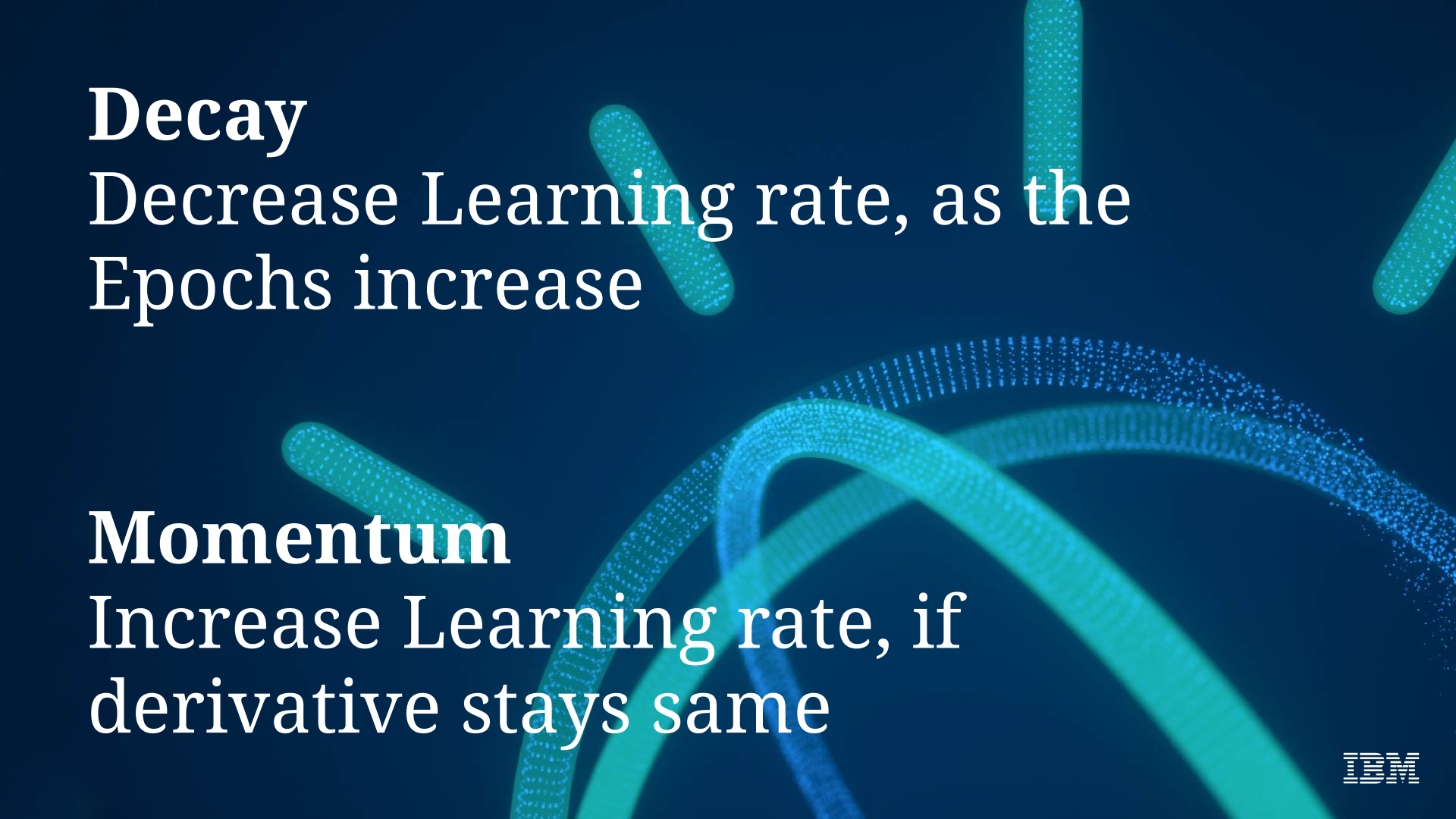
The size of the step, towards the right direction

Right Speed, **Learning Rate**



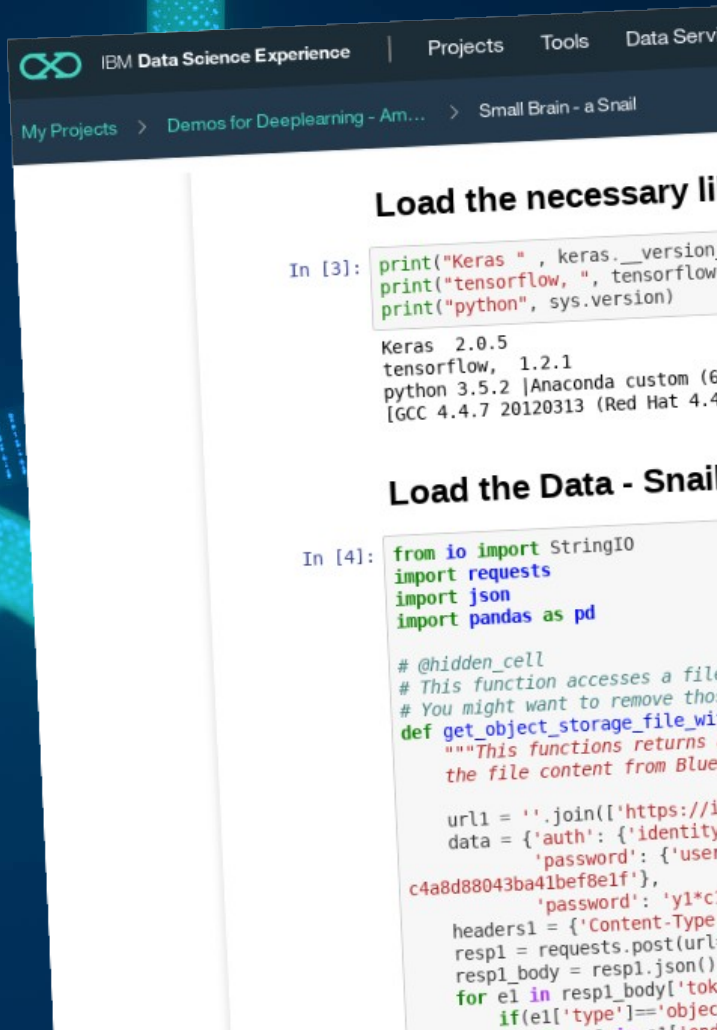
Decay
Decrease Learning rate, as the Epochs increase

Momentum
Increase Learning rate, if derivative stays same



SNAIL DEMO

1) Load labeled data
2) Design a Network
   - *Number of Layers*
   - *Number of Units within Layer*
   - *Activation Function of Unit*
3) Train it!
   - *Learning Rate*



Voorpagina
Net binnen
Algemeen
Achtergronden
Economie
Sport
Tech
Entertainment
Lifestyle
Overig
Wetenschap
Opmerkelijk
Dieren
NUcheckt
Auto
Video's
Regionaal

NU.nl > Overig > Wetenschap

Slakken nemen complexe beslissingen met twee hersencellen

Gepubliceerd: 03 juni 2016 19:17
Laatste update: 04 juni 2016 05:23

Wetenschappers hebben een simpel hersenmechanisme in kaart gebracht waarmee slakken besluiten of ze gaan eten of niet.

Als een slak langs een blaadje sla kruipt, bepaalt de activiteit van twee hersencellen of het dier stopt en het voedsel opeet.

Dat melden onderzoekers van Sussex in het wetenschappelijk tijdschrift *Nature Communications*.

Bij hun onderzoek brachten biologen met elektronische apparatuur de hersenactiviteit in kaart van tientallen slakken die door een ruimte kropen waarin

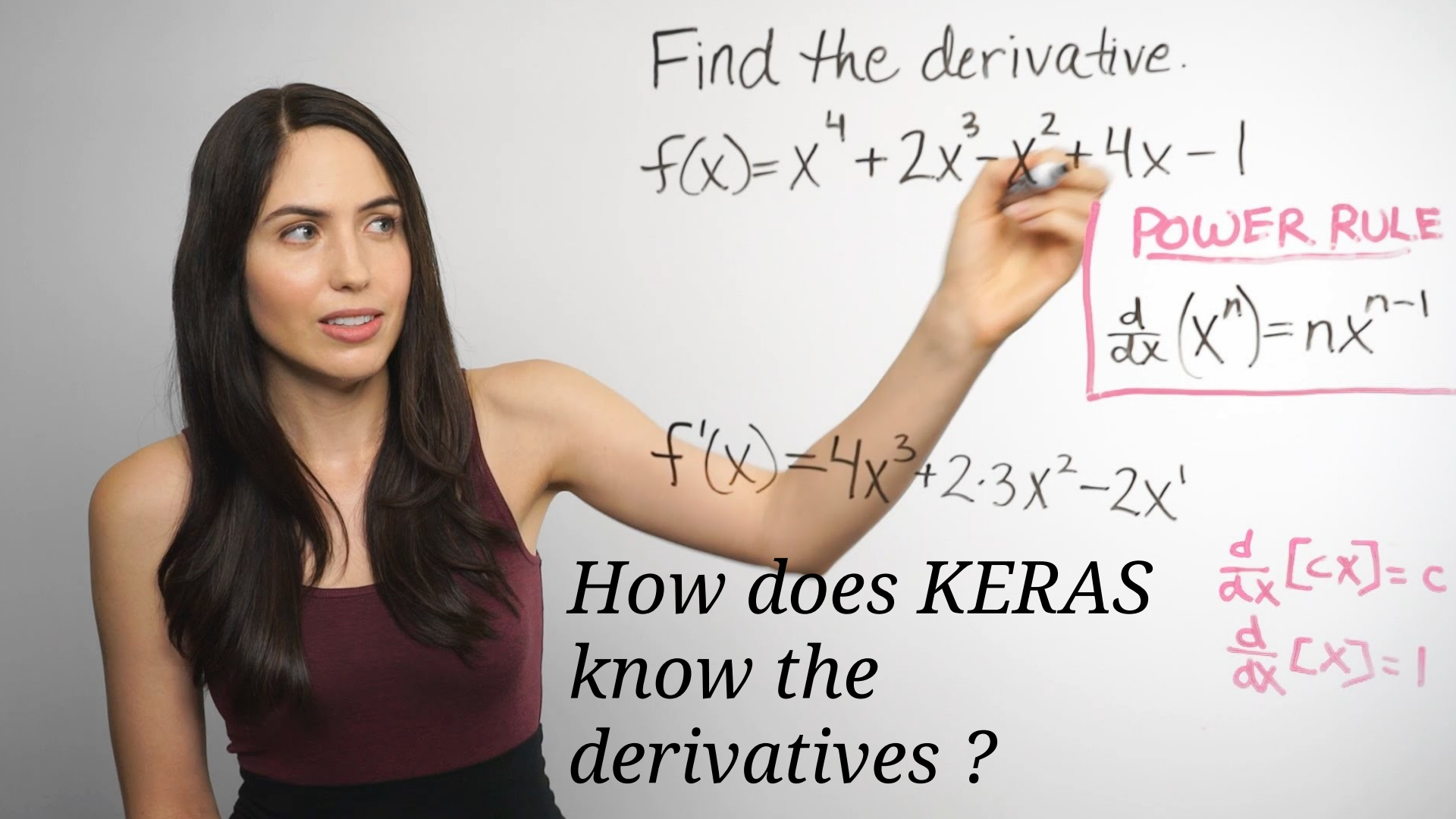
Net binnen

13:21 - Fossiel van nieuwe prehistorische dolfijne...
13:11 - Ook Locadia meldt zich geblesseerd af v...
13:01 - Harry Piekema krijgt rol in Sinterklaasjour...
12:57 - Rus na partij van ruim drie uur onderuit in...

Meest gelezen

1. Dode en gewonde bij schietpartij bij nachtclub...
2. Zeker 26 doden bij schietpartij in kerk Amerika...
3. Autoriteiten tasten nog in duister over motief s...
4. Stations Amsterdam en Schiphol gaan weer o...

Meer nieuws >


Find the derivative.
$f(x) = x^4 + 2x^3 - x^2 + 4x - 1$
POWER RULE
$\frac{d}{dx}(x^n) = nx^{n-1}$
$f'(x) = 4x^3 + 2 \cdot 3x^2 - 2x^1$
$\frac{d}{dx}[cx] = c$
$\frac{d}{dx}[x] = 1$
How does KERAS know the derivatives ?

Road Blocks:

- Need for Data (synthetic?)
- Over-fitting
- Design Choice?
- Computers Too Slooow
- “ERROR: Cannot use operator on 232x3 and 12x3”
- …

Why do we need Compute Power ?
Because N^3 !!!!

Matrix algebra [edit]

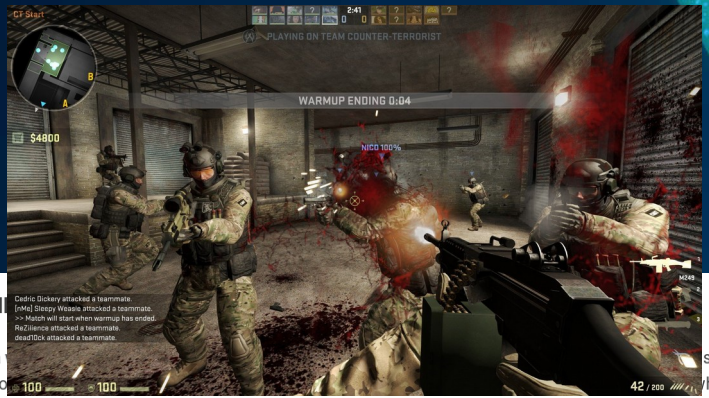
The following complexity figures assume that arithmetic with individual elements has complexity $O(1)$, as is the case with fixed-precision floating-point arithmetic or operations on a finite field.

Operation	Input	Output	Algorithm	Complexity
Matrix multiplication	Two $n\times n$ matrices	One $n\times n$ matrix	Schoolbook matrix multiplication	$O(n^3)$
			Strassen algorithm	$O(n^{2.807})$
			Coppersmith–Winograd algorithm	$O(n^{2.376})$
			Optimized CW-like algorithms[14][15][16]	$O(n^{2.373})$
Matrix multiplication	One $n\times m$ matrix & one $m\times p$ matrix	One $n\times p$ matrix	Schoolbook matrix multiplication	$O(nmp)$
Matrix inversion*	One $n\times n$ matrix	One $n\times n$ matrix	Gauss–Jordan elimination	$O(n^3)$
			Strassen algorithm	$O(n^{2.807})$
			Coppersmith–Winograd algorithm	$O(n^{2.376})$
			Optimized CW-like algorithms	$O(n^{2.373})$
Singular value decomposition	One $m\times n$ matrix	One $m\times m$ matrix, one $m\times n$ matrix, & one $n\times n$ matrix		$O(mn^2)$ ($m \le n$)
		One $m\times r$ matrix, one $r\times r$ matrix, & one $n\times r$ matrix		
			Laplace expansion	$O(n!)$
			Division-free algorithm[17]	$O(n^4)$

	Reference:	cost of calculation	n of nxn
0	Fav Icon of Browser	78.4	28
1	Icon Size in your Computer	1638.4	128
2	Icon Size in iPad / Tablet	6553.6	256
3	Small webcam	26214.4	512
4	Background of old Windows XP	104857.6	1024

IBM

Relation 3D game and GPU and Linear Algebra



Fou

Just a ... some trigono... what you have to give each function and what you will get out, and for the in-between stuff you can just use any calculator or math library that you may have.

Tip: If you do want to have a better grasp of the inner workings of these equations, then you should watch this video and read this PDF.
All linear transformations take this form:

$$B = F(A)$$

This states that if you have a linear transformation function $F()$, and your input is the vector A, then your output will be the vector B.

Each of these pieces - the two vectors and the function - can be represented as a matrix: the vector B as a 1x3 matrix, the vector A as another 1x3 matrix, and the linear transformation F as a 3x3 matrix (a *transformation matrix*).

A Big Thx to all the Gamerz out there!

Janusz "*snax*" Pogorzelski.

Freddy "*KRiMZ*" Johansson.

Kévin "*Ex6TenZ*" Droolans.

Ladislav "*GuardiaN*" Kovács.

Nicolai "*device*" Reedtz.

Patrik "*f0rest*" Lindberg.

Olof "*olofmeister*" Kajbjer.

Kenny "*kennyS*" Schrub.

IBM

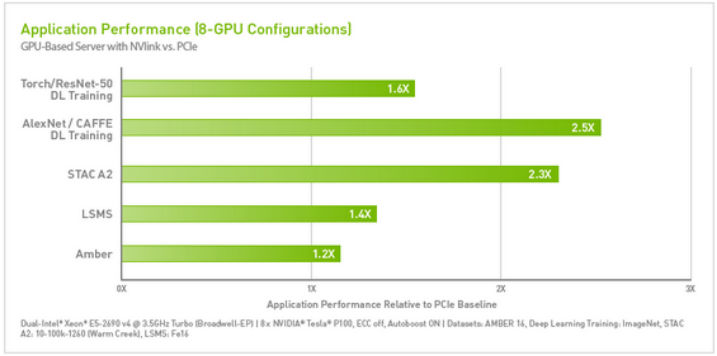
IBM & NVIDIA

NVIDIA NVLINK HIGH-SPEED INTERCONNECT

MAXIMUM THROUGHPUT FOR SUPERIOR APPLICATION PERFORMANCE

NVIDIA® NVLink™ is a high-bandwidth, energy-efficient interconnect that enables ultra-fast communication between the CPU and GPU, and between GPUs. The technology allows data sharing at rates 5 to 12 times faster than the traditional PCIe Gen3 interconnect, resulting in dramatic speed-ups in application performance and creating a new breed of high-density, flexible servers for accelerated computing. Download the whitepaper for more details on NVLink.

Designed to meet the challenges of exascale computing, NVLink is a fundamental ingredient of the U.S. Department of Energy's next-generation supercomputers. One such system—"Titan" at Oak Ridge National Laboratory—is currently the fastest supercomputer in the US*. Read more about these supercomputers by downloading the technology brief. Also, for Deep Learning applications, Tesla GPUs can accelerate you time to results from months to days. Read more about deep learning on Tesla platforms here.



NVIDIA

DRIVERS PRODUCTS DEEP LEARNING AND AI COMMUNITIES SUPPORT SHOP ABOUT NVIDIA

HOME DEEP LEARNING VIRTUAL REALITY DRIVING PRO GRAPHICS GAMING DATA CENTER

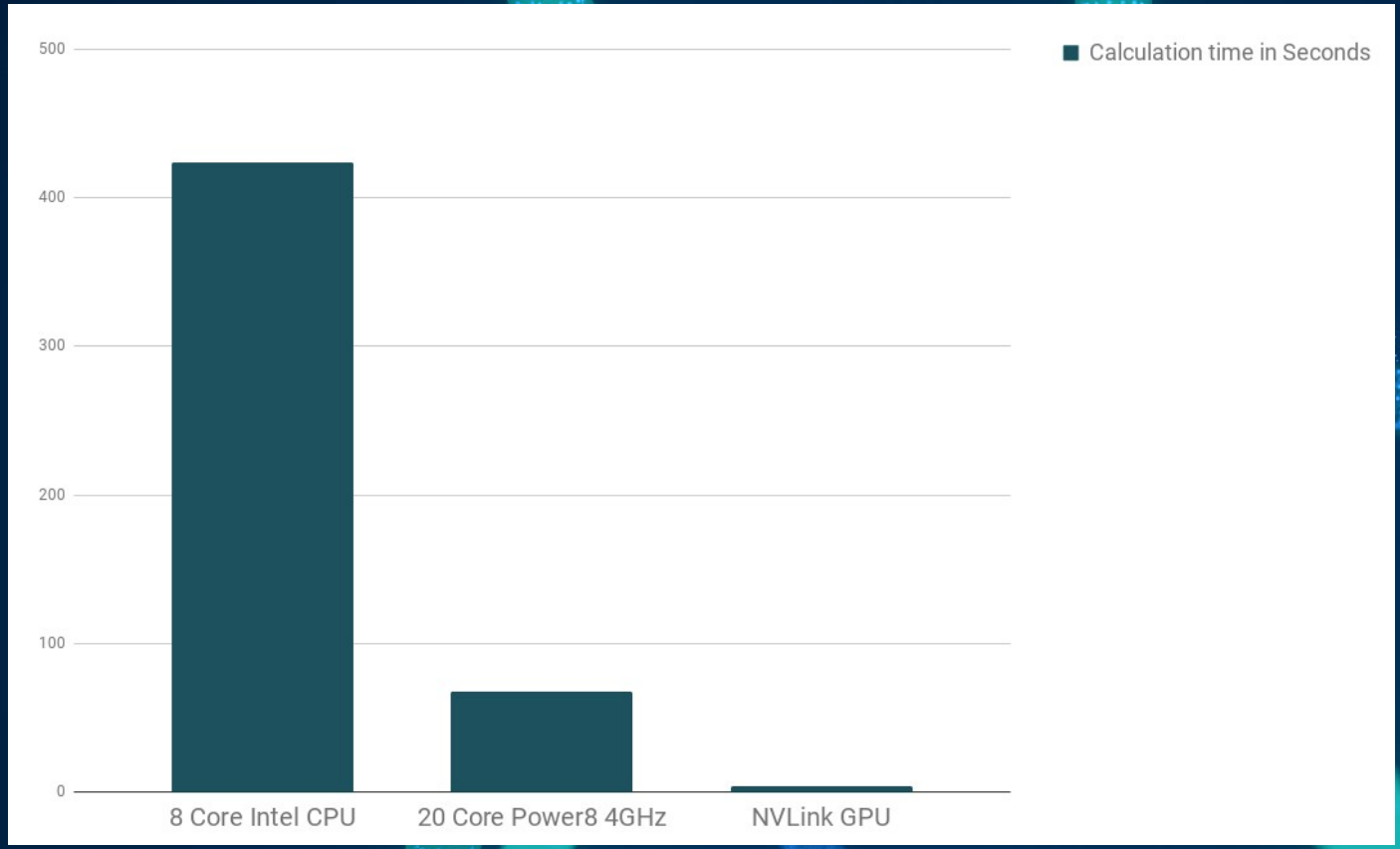
Posted on SEPTEMBER 8, 2016 by IAN BUCK

New IBM Servers with Tesla P100 GPUs and NVLink Mark a Milestone in High Performance Computing

Data center workloads are changing. Not long ago these systems were primarily used to handle storage and serve up web pages, but now they're increasingly tasked with AI workloads like understanding speech, text, images and video or analyzing big data for insights.

Billions of consumers want instant answers to a multitude of questions, while enterprise companies want to analyze mountains of data to better serve their customers' needs. Where do those answers

Fastest Data Bus to GPU in the Netherlands is in IBM Amsterdam, NVLink on Power



Road Blocks:

- Need for Data (synthetic?)
- Over-fitting
- Design Choice?
- ~~Computers Too Slooow~~
- “ERROR: Cannot use operator on 232x3 and 12x3”
- …

Over-fitting....

Too Focused on details, when learning:

- Don't look on page number, when learning in a book, it is not related

When A network is too complex for a problem, it tries to find relations which are not part of the phenomenon we try it to teach itself...

IBM

Brain Myth: Drinking alcohol kills brain cells

Brain Fact: Moderate alcohol use doesn't kill brain cells, and while rampant alcohol use can damage the brain, it's not due to cell death.

Does alcohol kill brain cells? You've probably heard this myth, but it's not really true. Moderate alcohol intake doesn't kill brain cells, or even damage them. That's because the amount of alcohol needed to kill brain cells would also kill the person drinking it!

That doesn't mean that alcohol can't damage the brain, though. Alcoholics can experience brain damage related to drinking, but it's not because alcohol kills brain cells. There are a few things that can happen when people drink a lot of alcohol over a long period of time. **While it can't kill brain cells, it can damage the dendrites, which are the branch-like ends of the brain cells. Dendrites are key for passing messages from one neuron to another**, so a degradation of the dendrites can cause cognitive problems. Recent research shows that dendrite damage can be reversed with certain kinds of therapy and training.

https://www.brainhq.com/brain-resources/brain-facts-myths/brain-mythology/brain-myth-alcohol-kills-brain-cells

Keras Documentation

Search docs

- Home
- Getting started
  - Guide to the Sequential model
  - Guide to the Functional API
  - FAQ
- Models
  - About Keras models
  - Sequential
  - Model (functional API)
- Layers
  - About Keras layers
- **Core Layers**
  - Dense
  - Activation
  - Dropout
  - Flatten
  - Reshape
  - Permute

Same shape as input.

Dropout [source]

```
keras.layers.Dropout(rate, noise_shape=None, seed=None)
```

Applies Dropout to the input.

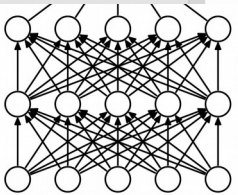
Dropout consists in randomly setting a fraction `rate` of input units to 0 at each update during training time, which helps prevent overfitting.

Arguments

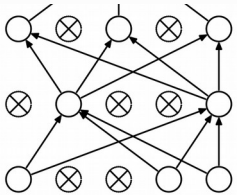
- **rate**: float between 0 and 1. Fraction of the input units to drop.
- **noise_shape**: 1D integer tensor representing the shape of the binary drop
  multiplied with the input. For instance, if your inputs have shape
  `(batch_size, timesteps, features)` and you want the dropout m
  timesteps, you can use `noise_shape=(batch_size, 1, features)` .
- **seed**: A Python integer to use as random seed.

References

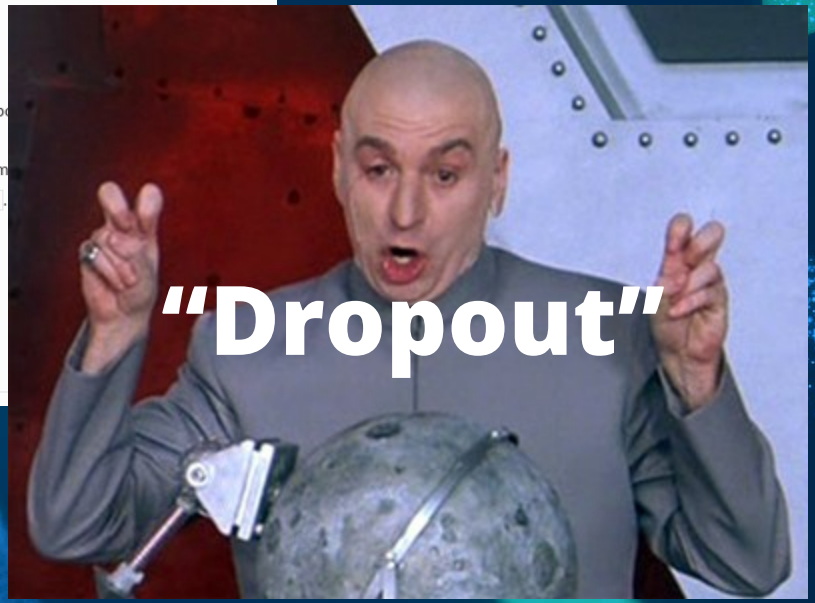
- Dropout: A Simple Way to Prevent Neural Networks from Overfitting



(a) Standard Neural Net

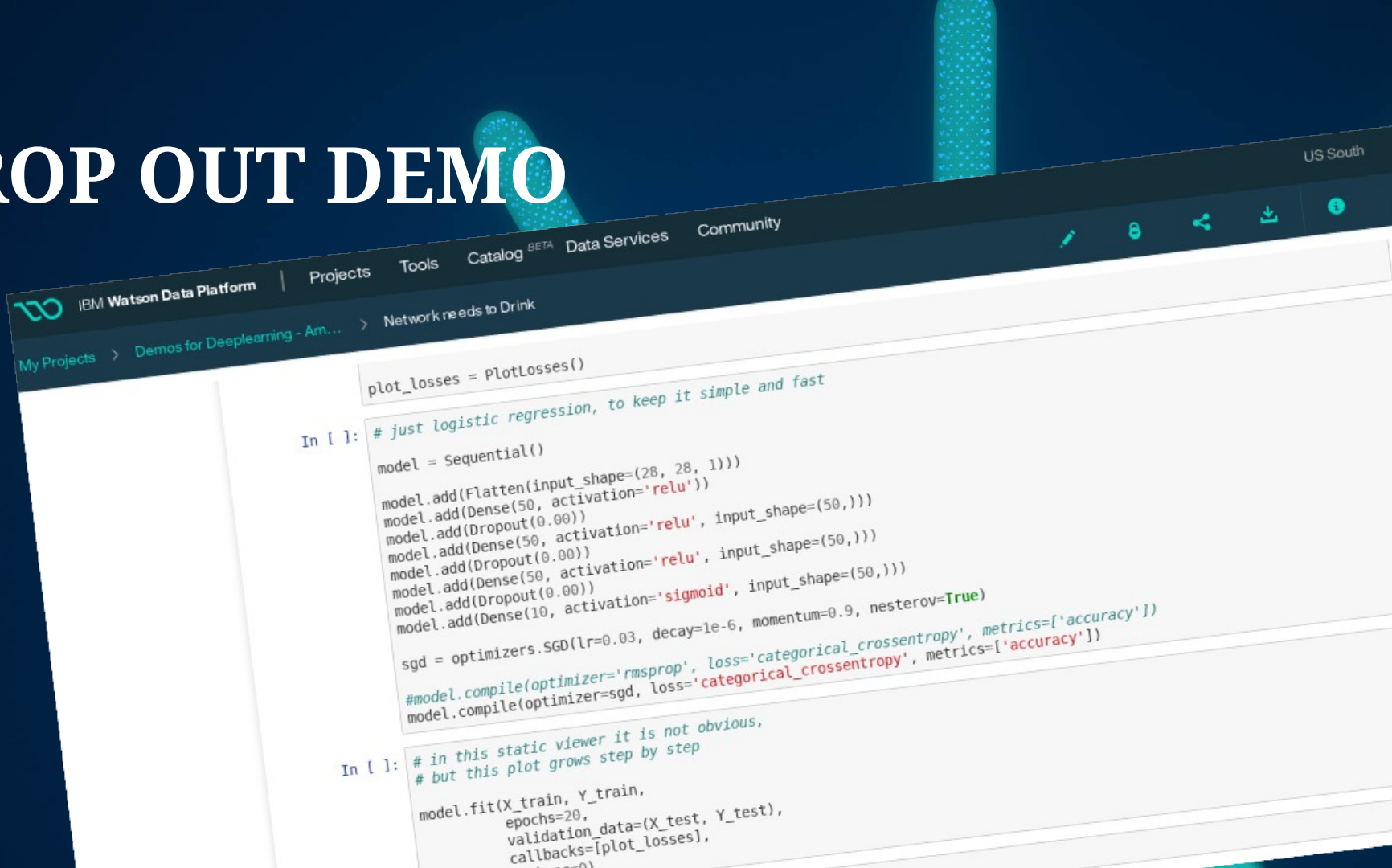


(b) After applying dropout.



IBM

DROP OUT DEMO



Do we accept errors made by NN?

Trends?

How do we Humans learn from Networks?

What is the best design for a problem?

slides & code:
bit.ly/intro_deeplearning_dsx

IBM