

A Small Intro in Deeplearning

What are they ?

What can they do ?

How to create them ?

Willem Hendriks
willem.hendriks@nl.ibm.com





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":

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 expertise, as there's a lot more to it than just a matter of focus and experience. (This is true

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

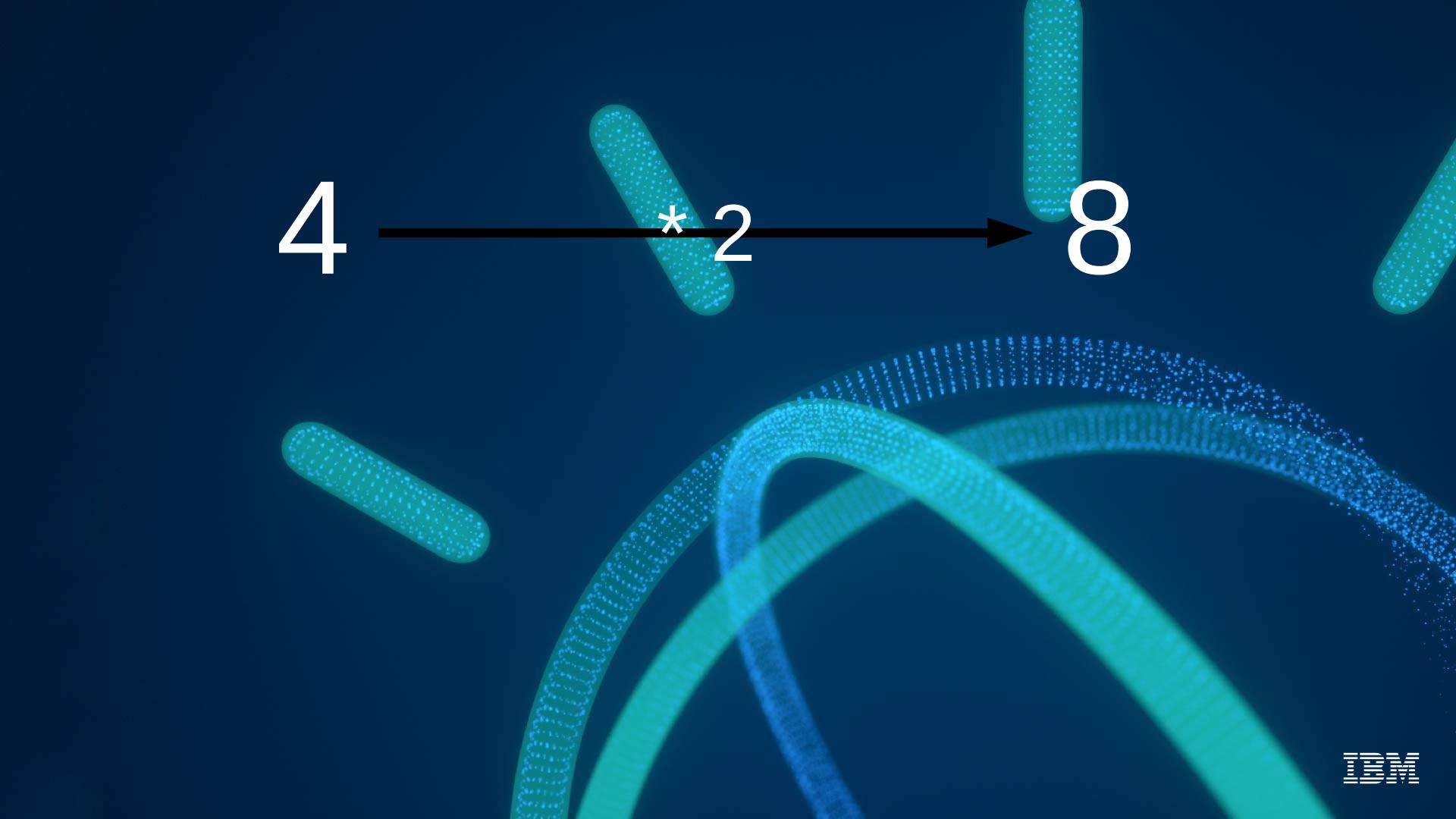
This course also teaches you how Deep Learning actually works, rather than presenting only a cursory or surface-level description. So after completing it, you will be able to apply deep learning to your own applications. If you are looking for a job in AI, after this course you will also be able to answer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

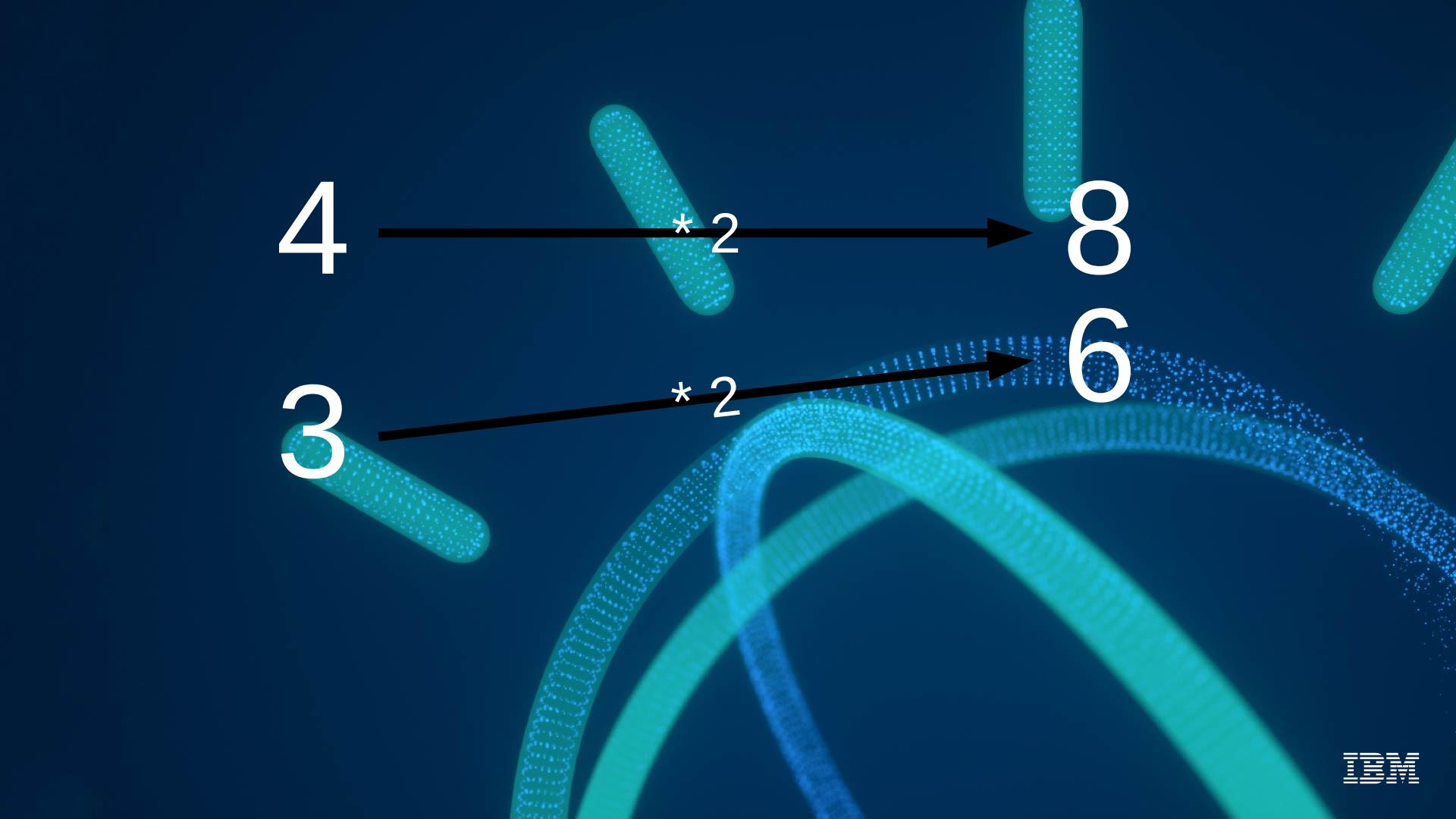
$$4 * 2 = 8$$



4

* 2

8



4

* 2

8

3

* 2

6

4

* 2

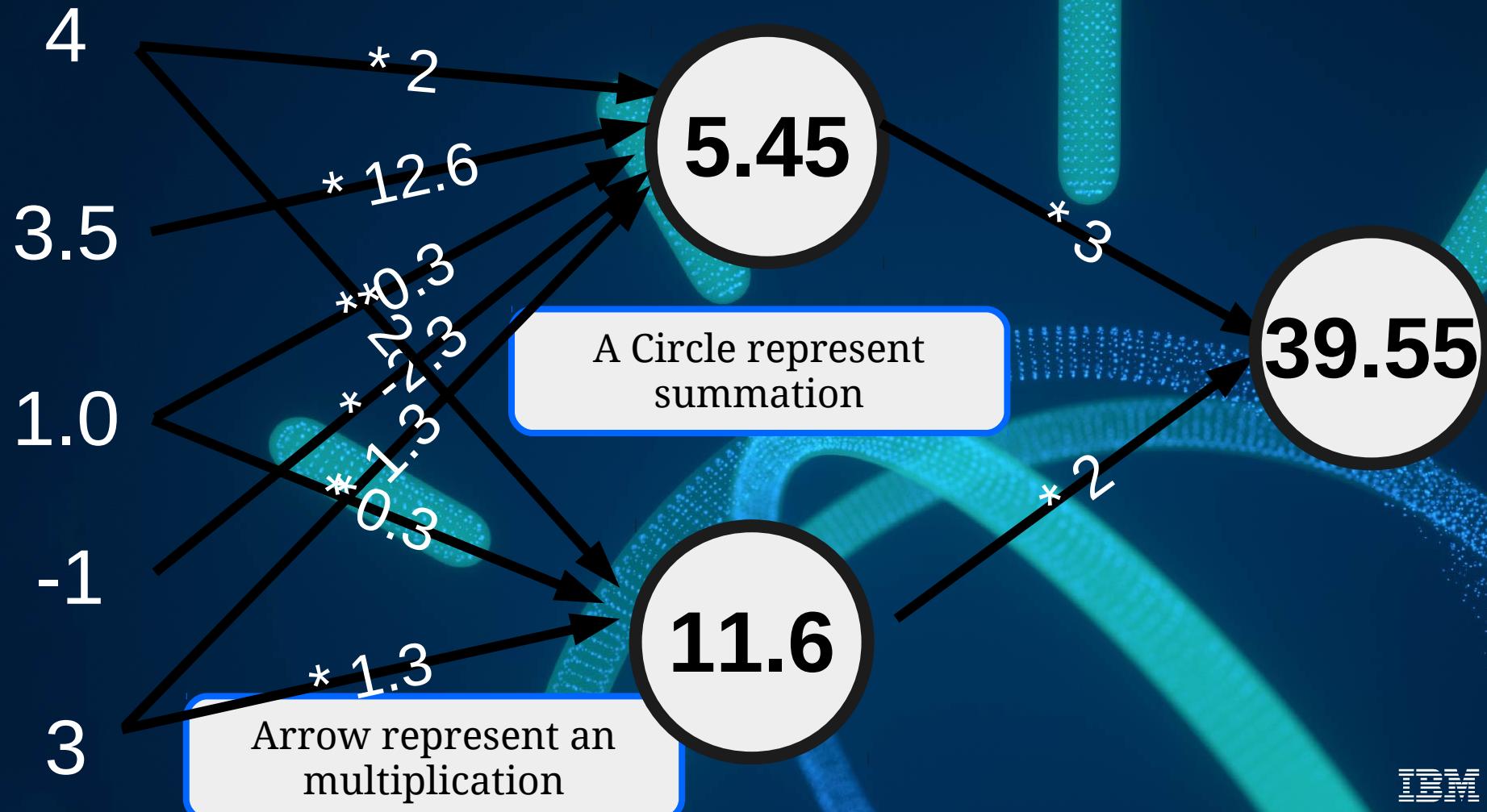
3

* 2

Arrow represent an
multiplication

14

A Circle represent
summation



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

6 °C | °F

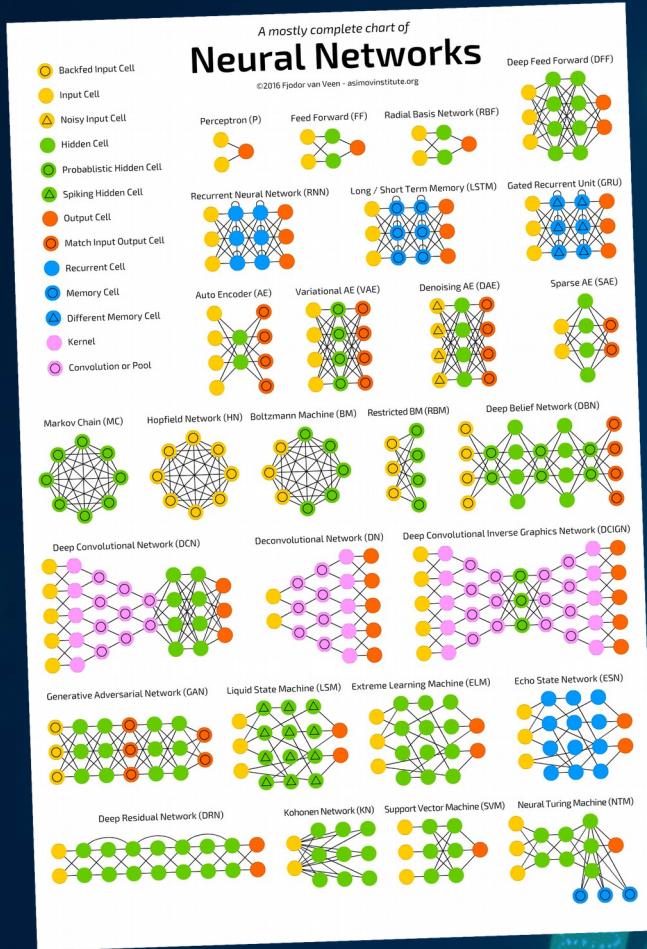
0

34.4

Thor! Smack
extreme
numbers
to between
0.0 and 1.0

0.9



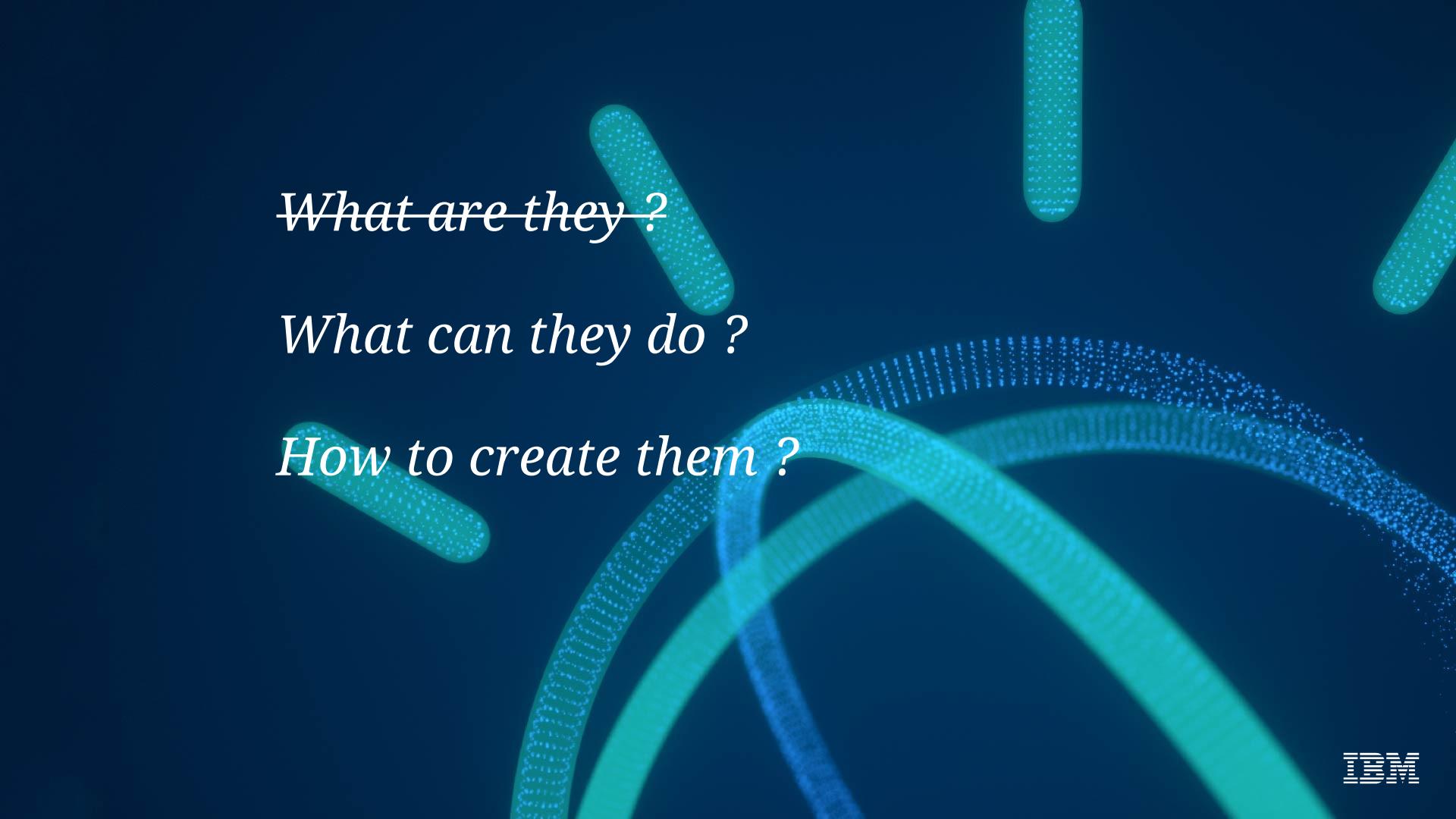


Many Different Design.

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



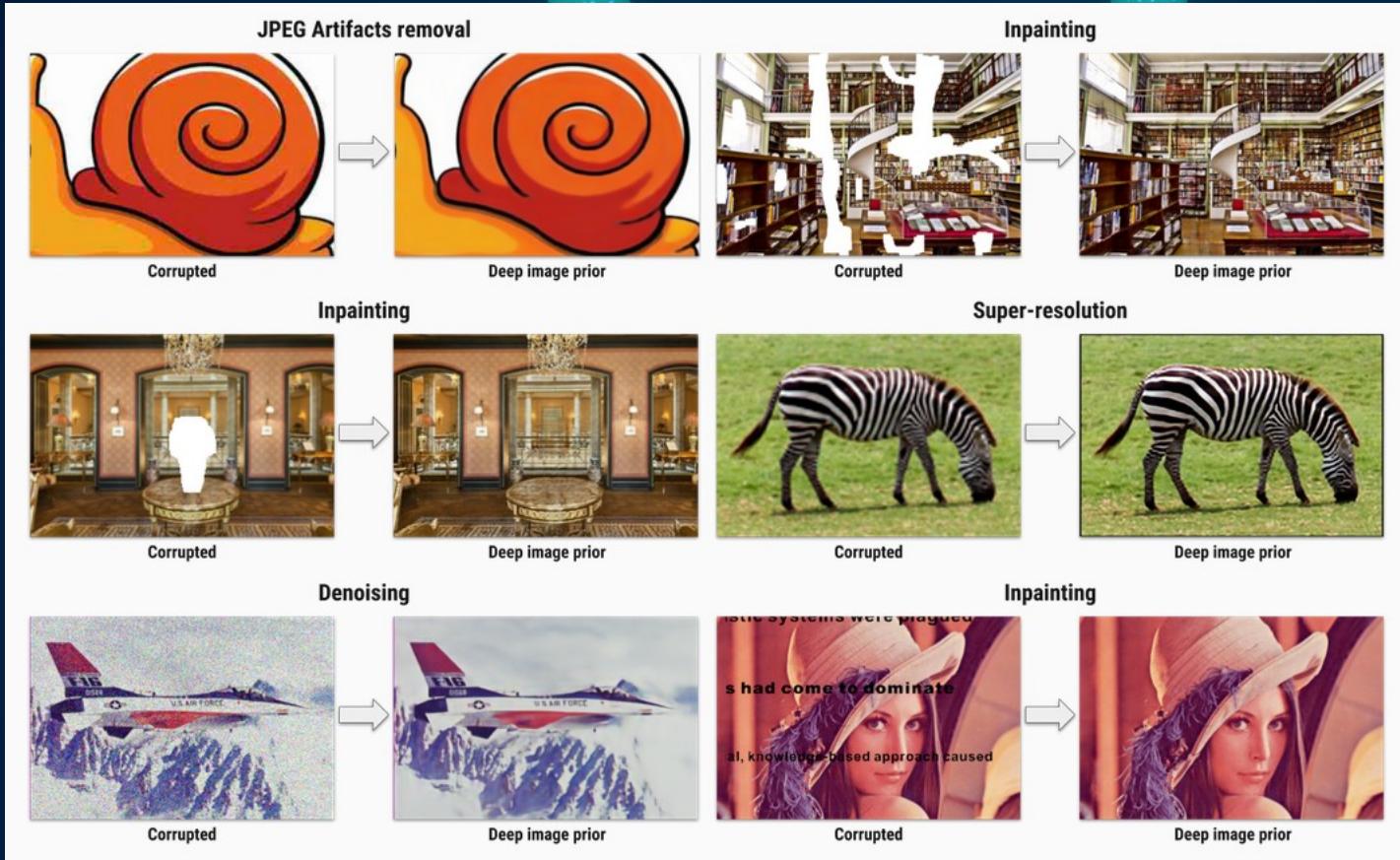
What are they ?

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>



Chopin Music Generation with RNN (Recurrent Neural Networks) and Deep Learning

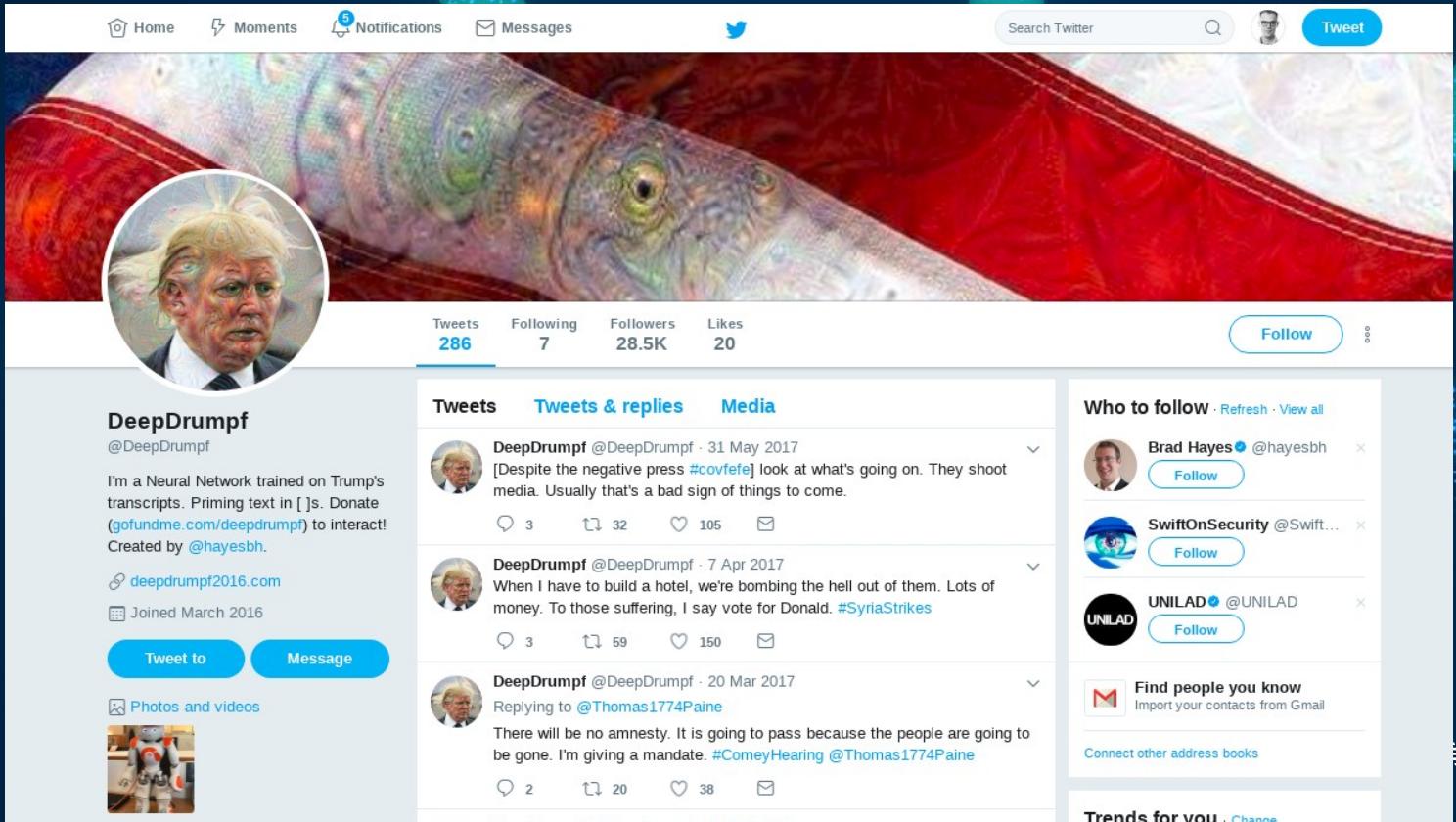
17,707 views

134 9 SHARE ...

IBM

Networks can Tweet like Trump

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



The screenshot shows a Twitter profile for a user named "DeepDrumpf". The profile picture is a circular, colorful, abstract representation of Donald Trump's face, likely generated by a neural network. The background of the profile page features a similar abstract, multi-colored pattern.

Profile Summary:

- Tweets: 286
- Following: 7
- Followers: 28.5K
- Likes: 20

Profile Description:

I'm a Neural Network trained on Trump's transcripts. Priming text in []. Donate (gofundme.com/deepdrumpf) to interact! Created by @hayesbh.

Links:

- [@deepdrumpf2016.com](http://deepdrumpf2016.com)
- Joined March 2016

Actions:

- [Tweet to](#)
- [Message](#)

Media:

 Photos and videos

Tweets & replies:

- DeepDrumpf** @DeepDrumpf · 31 May 2017
[Despite the negative press #covfefe] look at what's going on. They shoot media. Usually that's a bad sign of things to come.
3 32 105
- DeepDrumpf** @DeepDrumpf · 7 Apr 2017
When I have to build a hotel, we're bombing the hell out of them. Lots of money. To those suffering, I say vote for Donald. #SyriaStrikes
3 59 150
- DeepDrumpf** @DeepDrumpf · 20 Mar 2017
Replying to @Thomas1774Paine
There will be no amnesty. It is going to pass because the people are going to be gone. I'm giving a mandate. #ComeyHearing @Thomas1774Paine
2 20 38

Who to follow:

- Brad Hayes** @hayesbh
[Follow](#)
- SwiftOnSecurity** @Swift...
[Follow](#)
- UNILAD** @UNILAD
[Follow](#)

Find people you know:
Import your contacts from Gmail

Connect other address books:

Trends for you:

Do a CSI: “Enhance!”

<https://letsenhance.io/>

**Image
upscaling
supercharged.**

Remove JPEG artifacts,
upscale 4X, preserve quality.
One-click and free.

**Machine
learning for
visual
content**

State of art neural networks
to help your work



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

f 6  + Subscribe 

StackGAN

Image Synthesis From Text With Deep Learning | Two Minute Pap...

Figure 3. Example results by our proposed StackGAN, GAWWN [20], and GAN-INT-CLS [22] conditioned on text descriptions from CUB test set. GAWWN and GAN-INT-CLS generate 16 images for each text description, respectively. We select the best one for each of them to compare with our StackGAN.

Text description	This flower has petals that are white and has pink shading	This flower has a lot of small purple petals in a dome-like configuration	This flower has long thin yellow petals and a lot of yellow anthers in the center	This flower is pink, white, and yellow in color, and has petals that are striped	This flower is white and yellow in color, with petals that are wavy and smooth	This flower has upturned petals which are thin and orange with rounded edges	This flower has petals that are dark pink with white edges and pink stamen
64x64 GAN-INT-CLS [22]	 	 	 	 	 	 	

MORE VIDEOS
StackGAN

ProF

Comprometidos con

#SomosPR



Neural Networks can to Text-To-Speech

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



WaveNet machine learning algorithm speaking

4,815 views

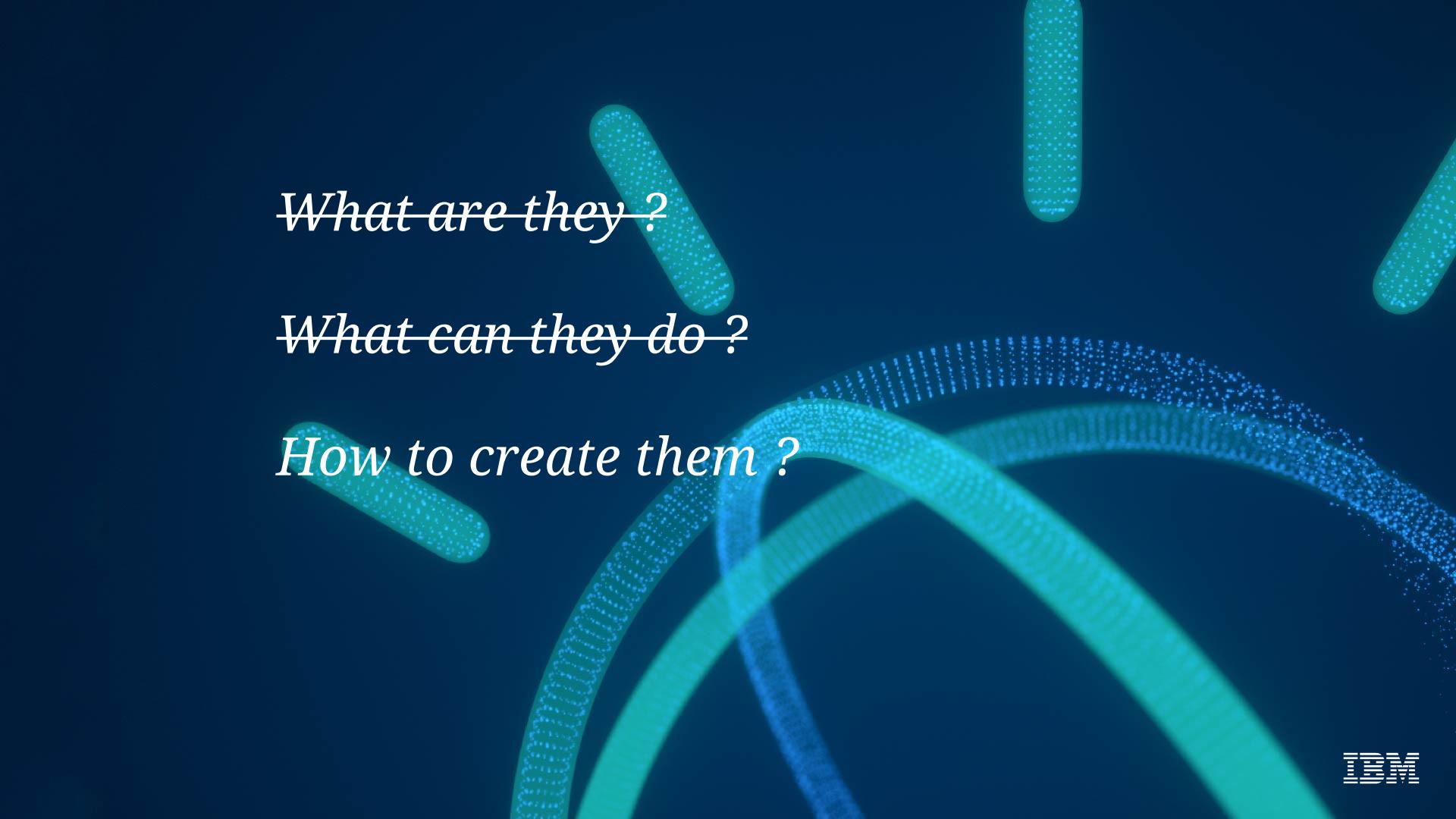
30 0 SHARE ...



sveinbjornt
Feb 11, 2017

SUBSCRIBE

IBM



What are they ?

What can they do ?

How to create them ?

Measure Performance

Machine Learning - Predicting:

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

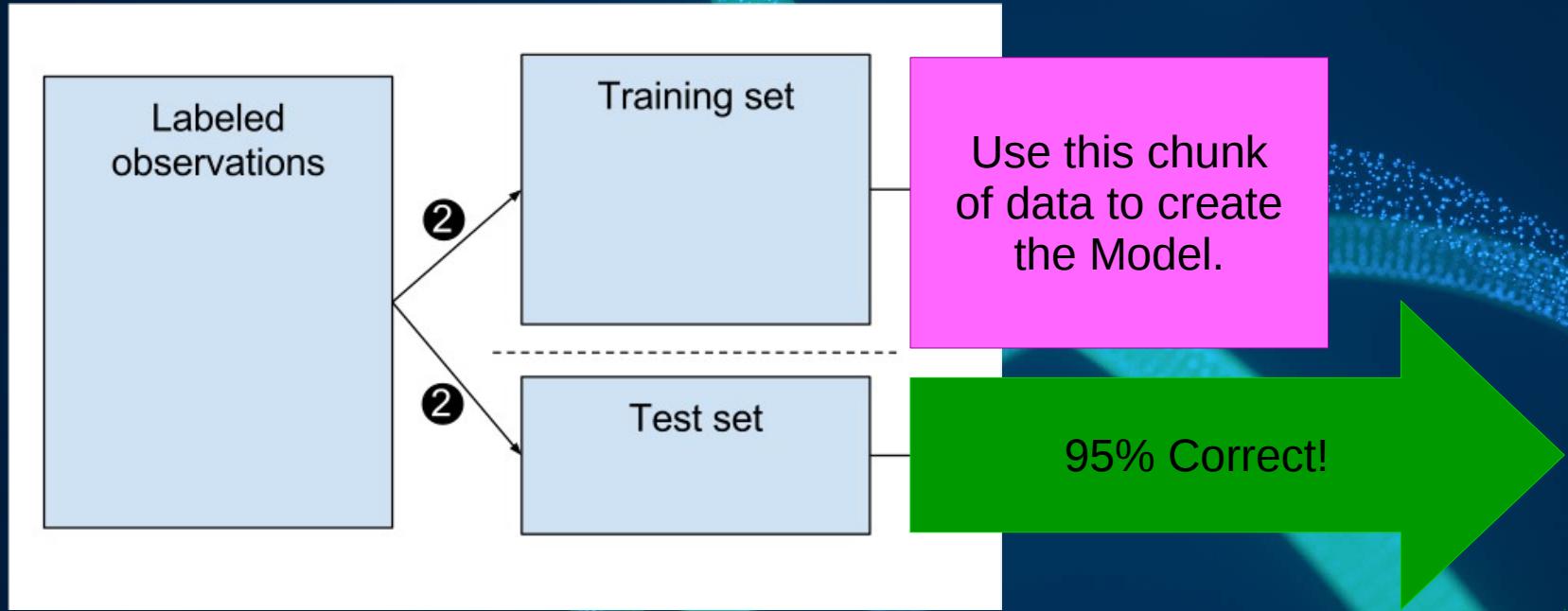
How do we often do that?

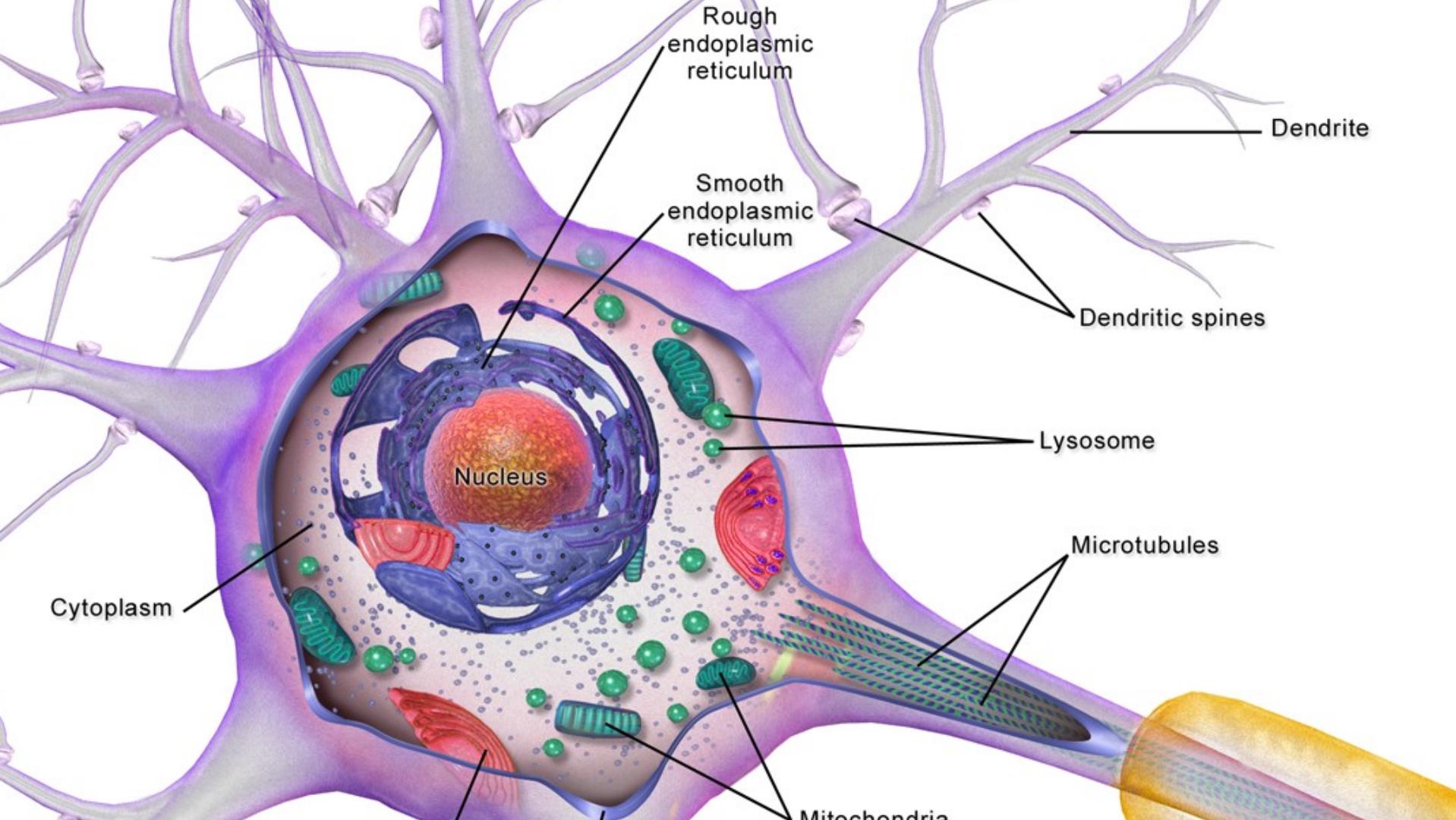
Data

(imaged of Cats & Dogs)

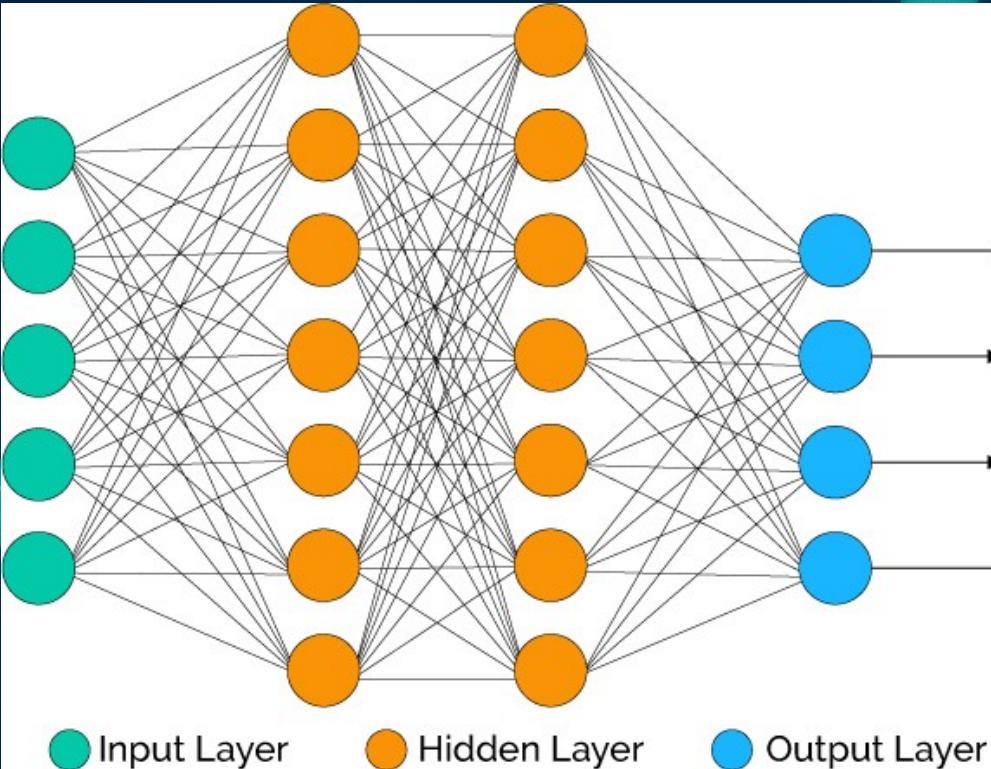
Model Performance?

Train/Test Set



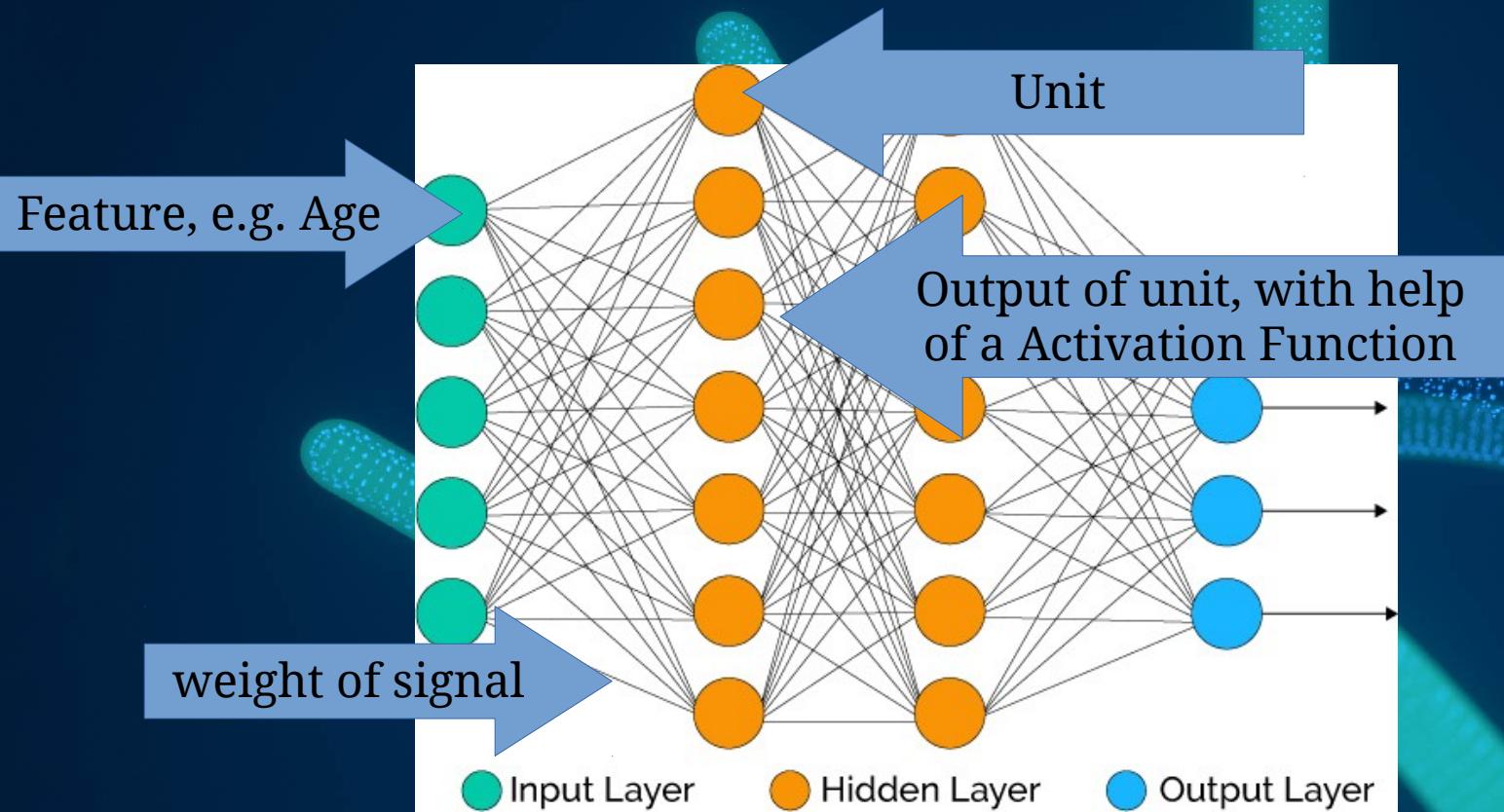


Artificial Neural Network (ANN)



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

Artificial Neural Network (ANN)



Artificial Neural Network (ANN)

Feature, e.g. Age

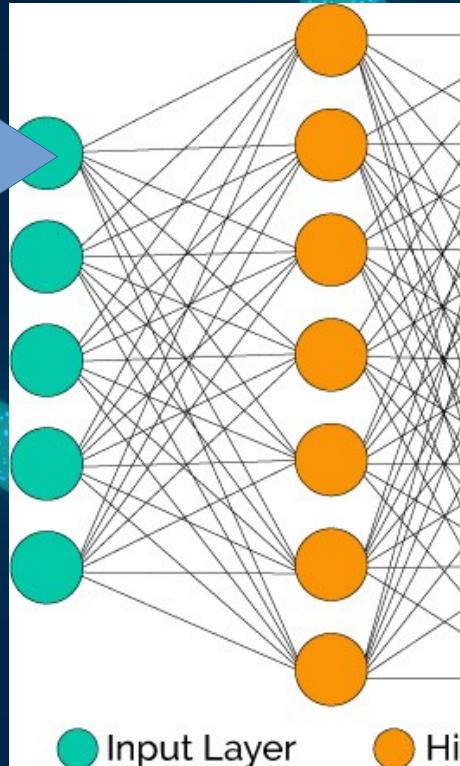
Pixel color value

Black White:
 $28 \times 28 = 784$ inputs

Color Picture:
 $28 \times 28 \times 3 = 2352$ 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

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



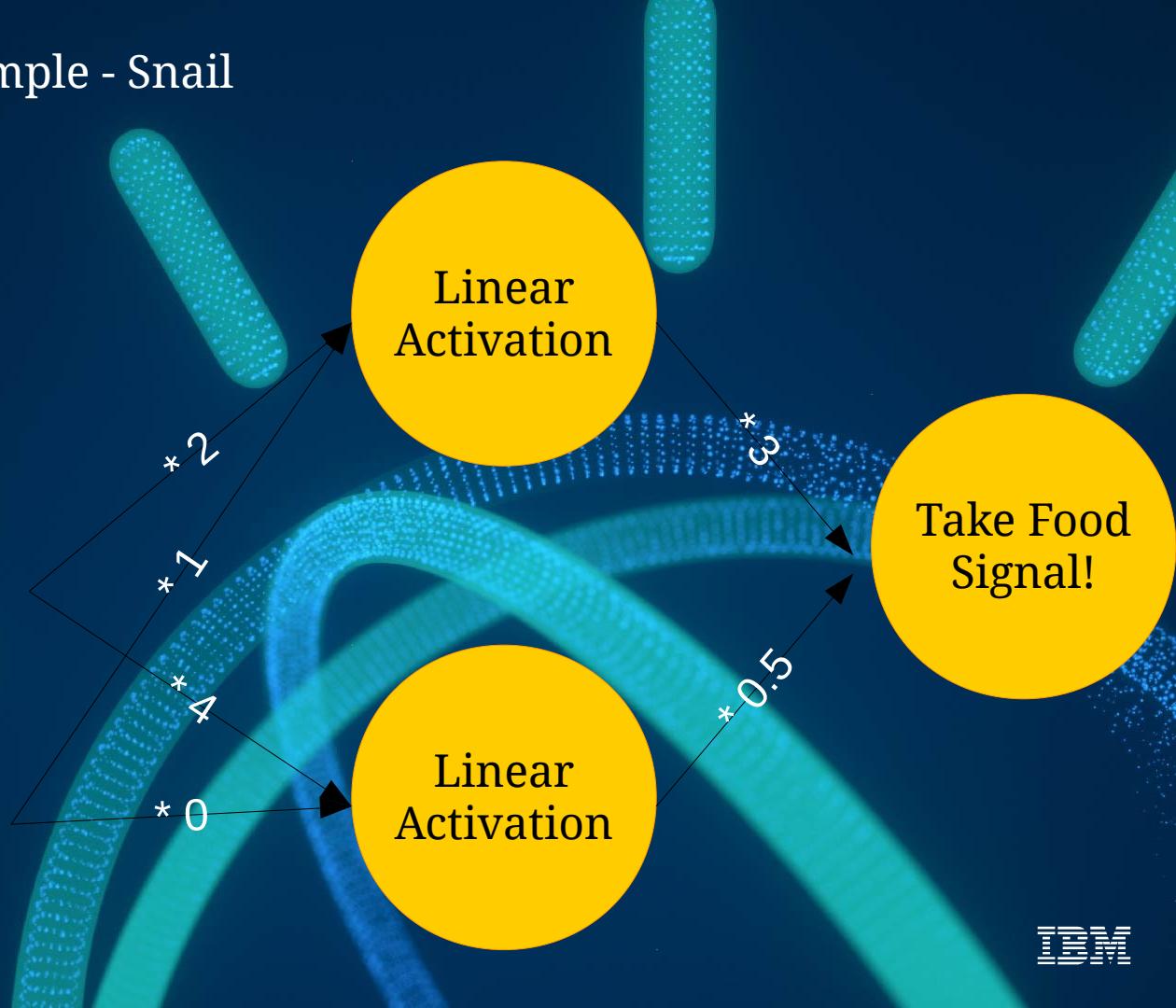
IBM

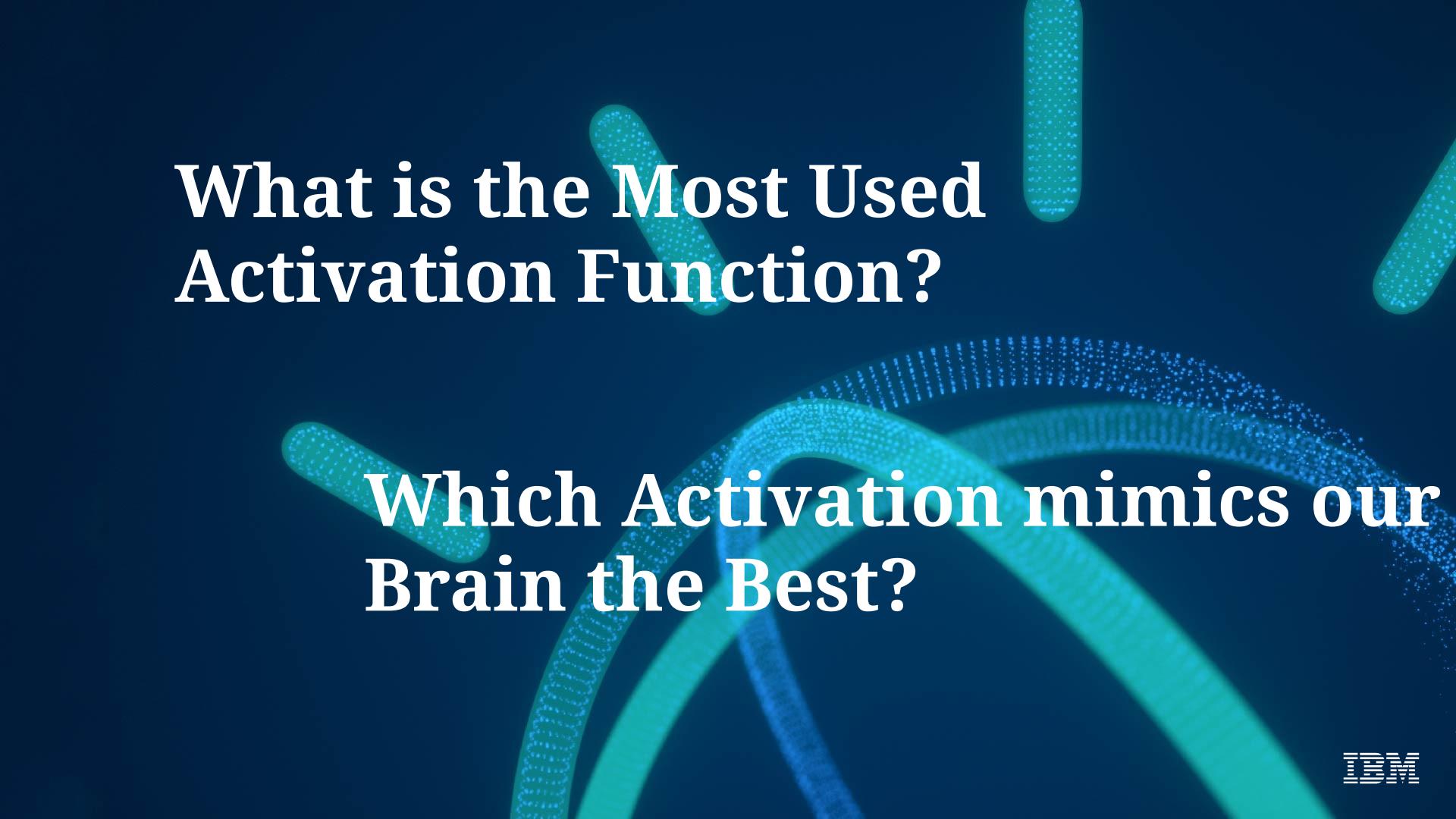
Example - Snail



Feature 1: Hunger Level
Input H: 0.5

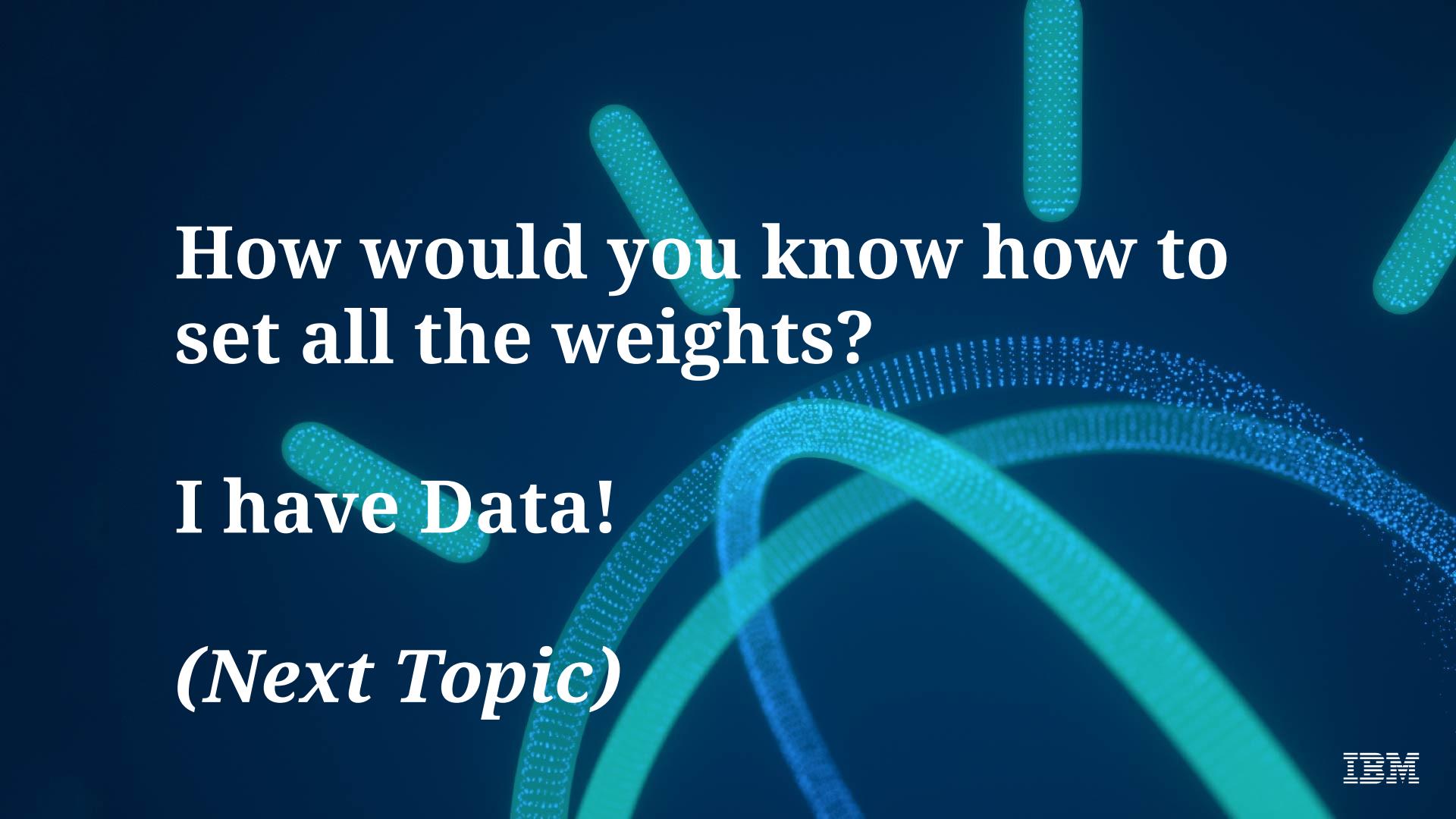
Feature 2: Sees food in
Range Level
Input S: 1





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!

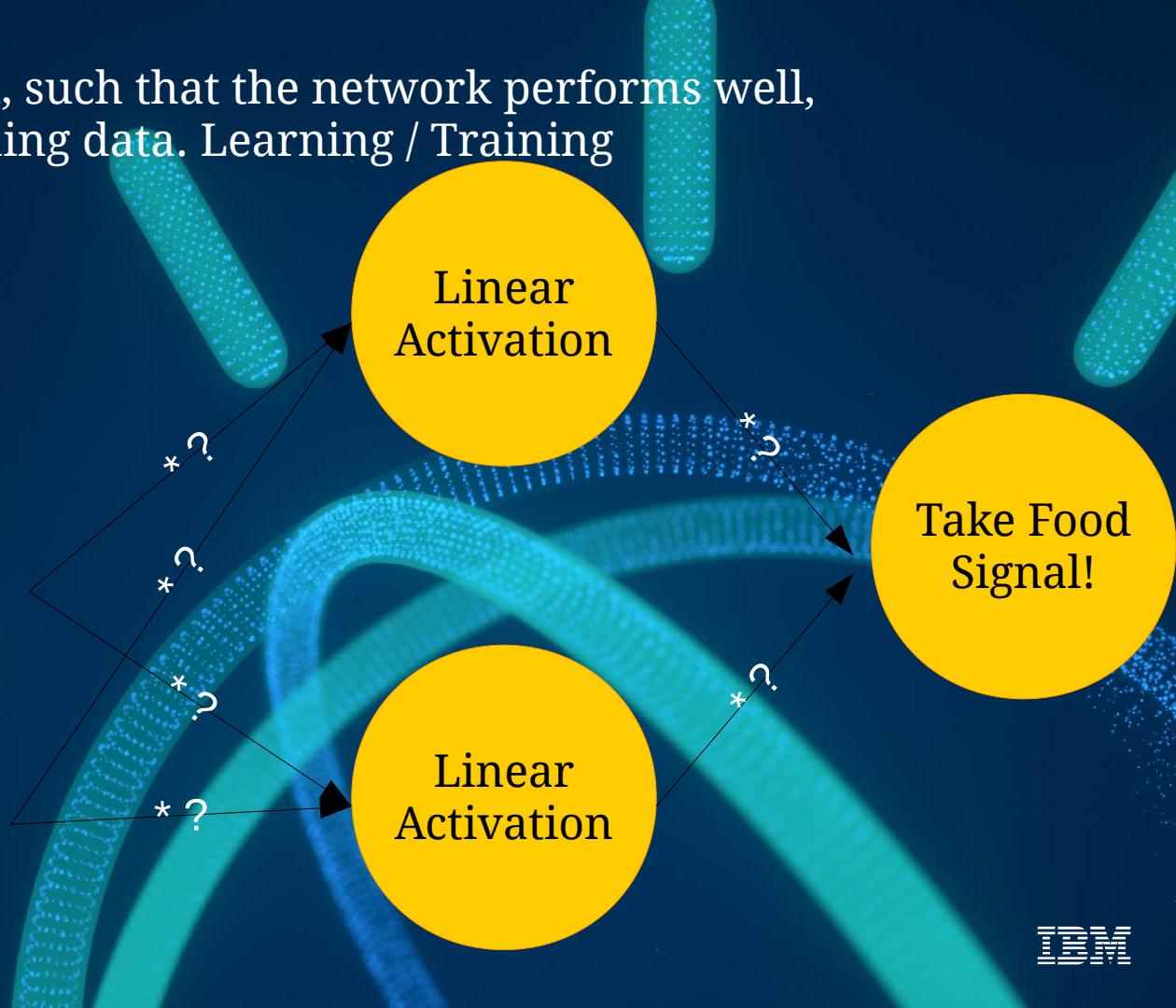
(Next Topic)

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

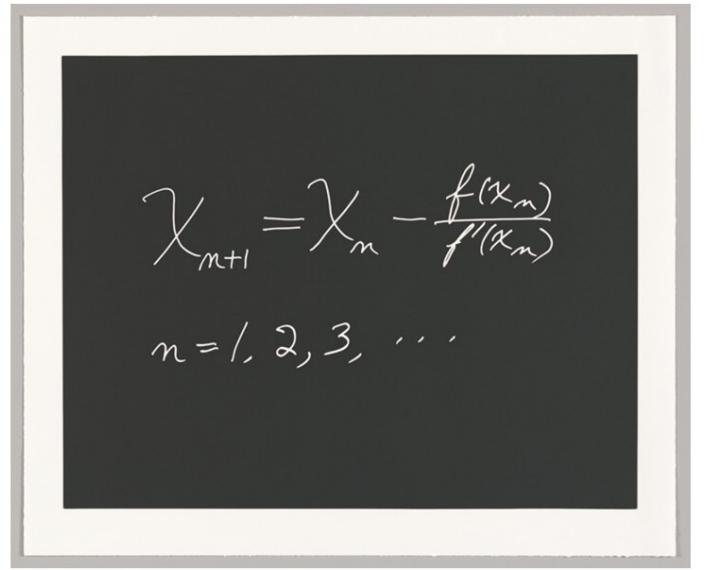


Feature 1: Hunger Level
Input H: 0.5

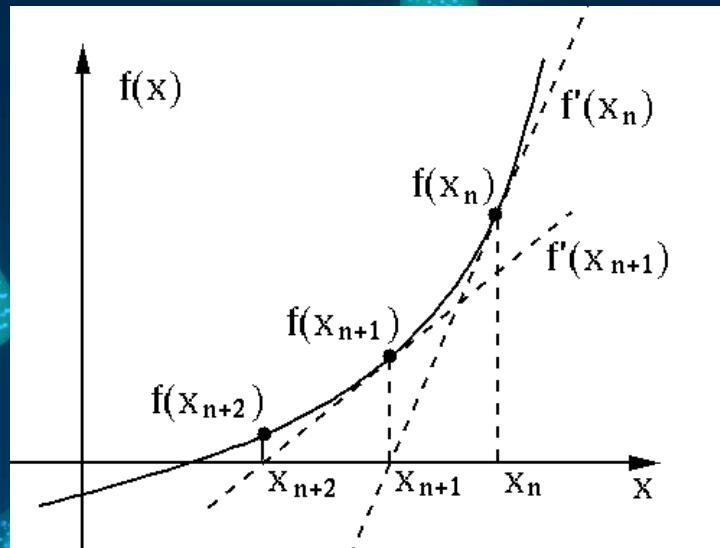
Feature 2: Sees food in
Range Level
Input S: 1



Stephen Smale: "Newton's Method"



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



IBM

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

Linear Activation

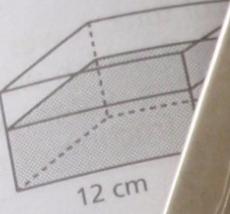
Take Food Signal!

Network Output: 6

Labeled Data: 10

Error: 4
Tweak all weights to reduce error

ains water filled to $\frac{3}{4}$ of its height.
be the water level in the tank,
 cm^3 of water is poured out from the
volume = length \times breadth \times height,
Height = $\frac{\text{volume}}{\text{length} \times \text{breadth}}$



1:
of water in the tank = $\frac{3}{4} \times 8$
= 6 cm
= 12×10
= 720 cm³
e of water in the tank = 60 cm^3
water poured out = 60 cm³
- ?

HOW TO STUDY

BY JOSEPH V. LANDY, S.J.

THE ART OF LEARNING
AND MAKING IT STICK

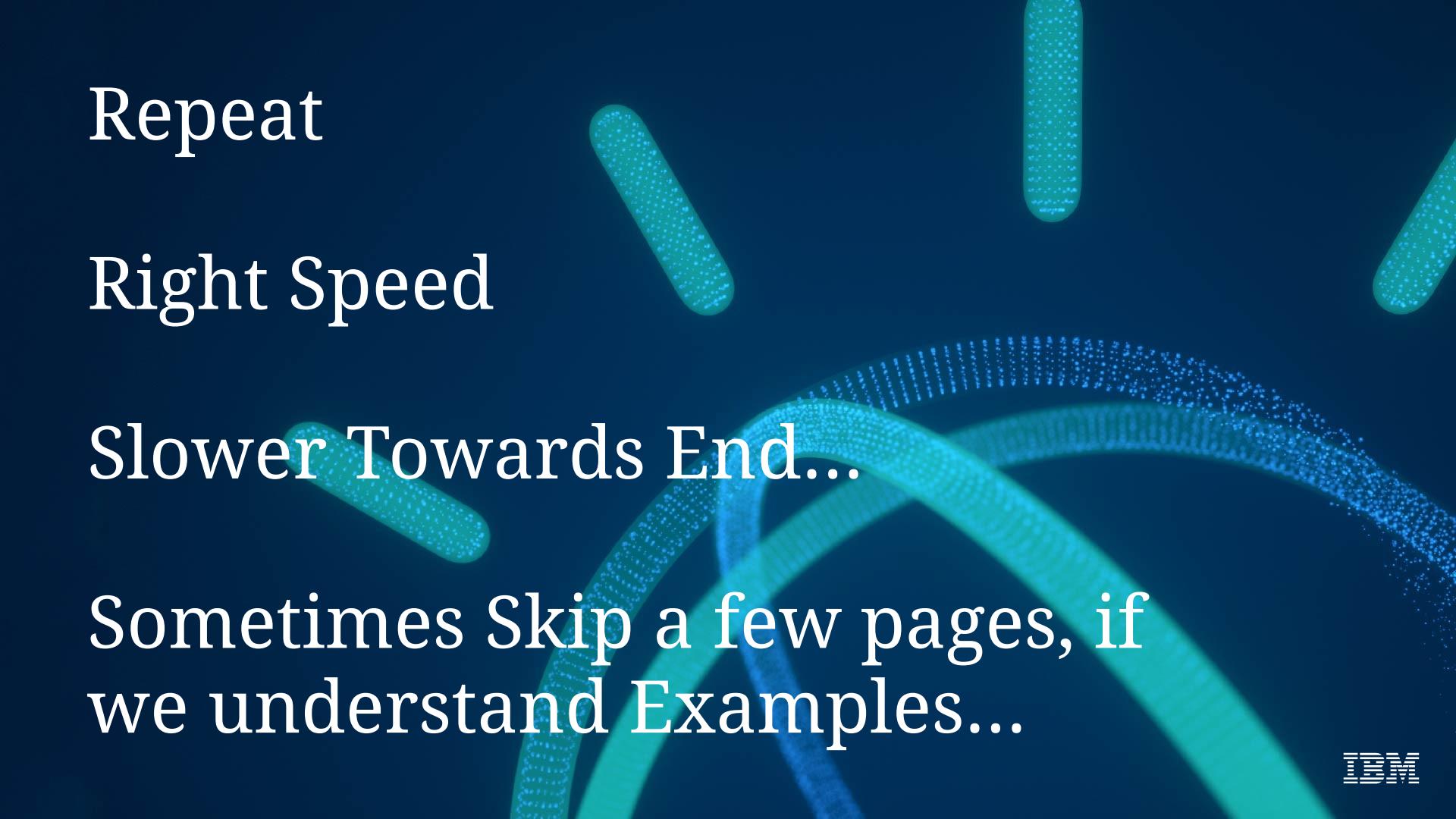


Take Note
Remember that $1 \text{ l} = 1000 \text{ cm}^3$ or 1000 ml. For instance in Example 1, if the volume of water in tank A is 1.728 l, then $1.728 \times 1000 = 1728 \text{ cm}^3$.

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

2, sometimes
or calculation
Example 2(b)—
s are awarded for
simplified answer (as
o make problem-solvin

ight given the volume,
or breadth, if the other
ave been included for
crease) in height may be cal
Example 3. Using this method
orking.



Repeat

Right Speed

Slower Towards End...

Sometimes Skip a few pages, if
we understand Examples...

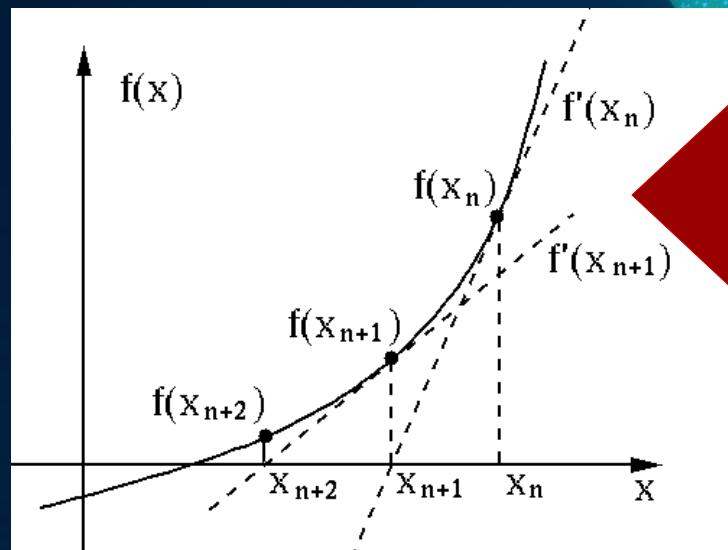
Repeat, Epochs

Right Speed, Learning Rate

Slower Towards End... Decay

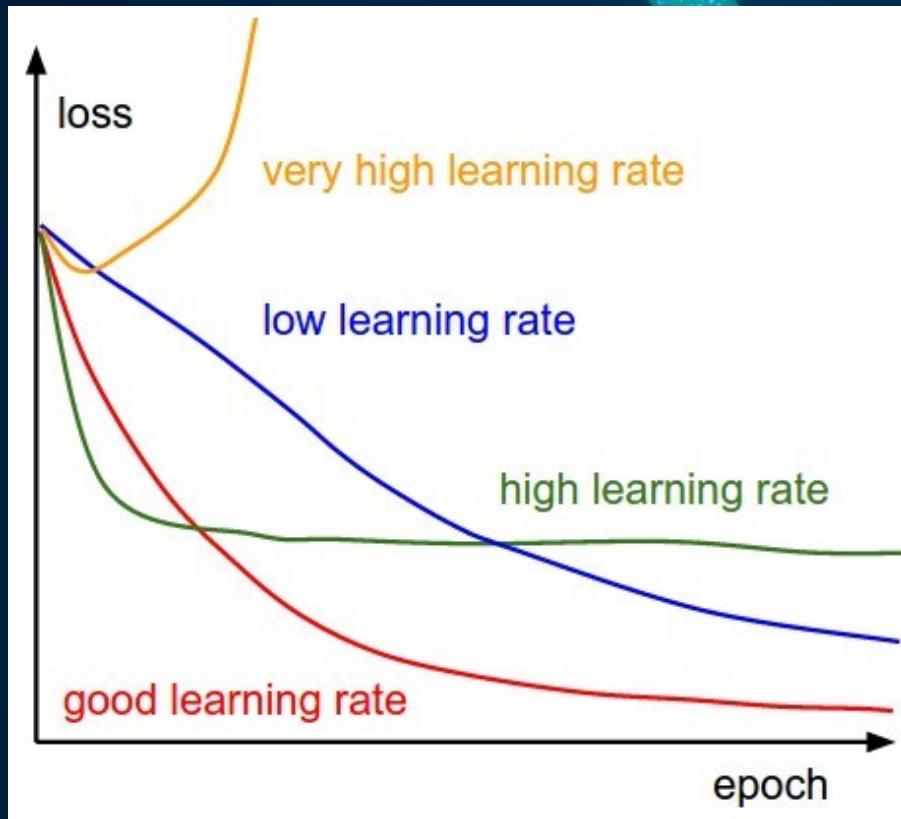
Sometimes Skip a few pages Momentum

Right Speed, Learning Rate



The size of the step, towards
the right direction

Right Speed, Learning Rate



Decay

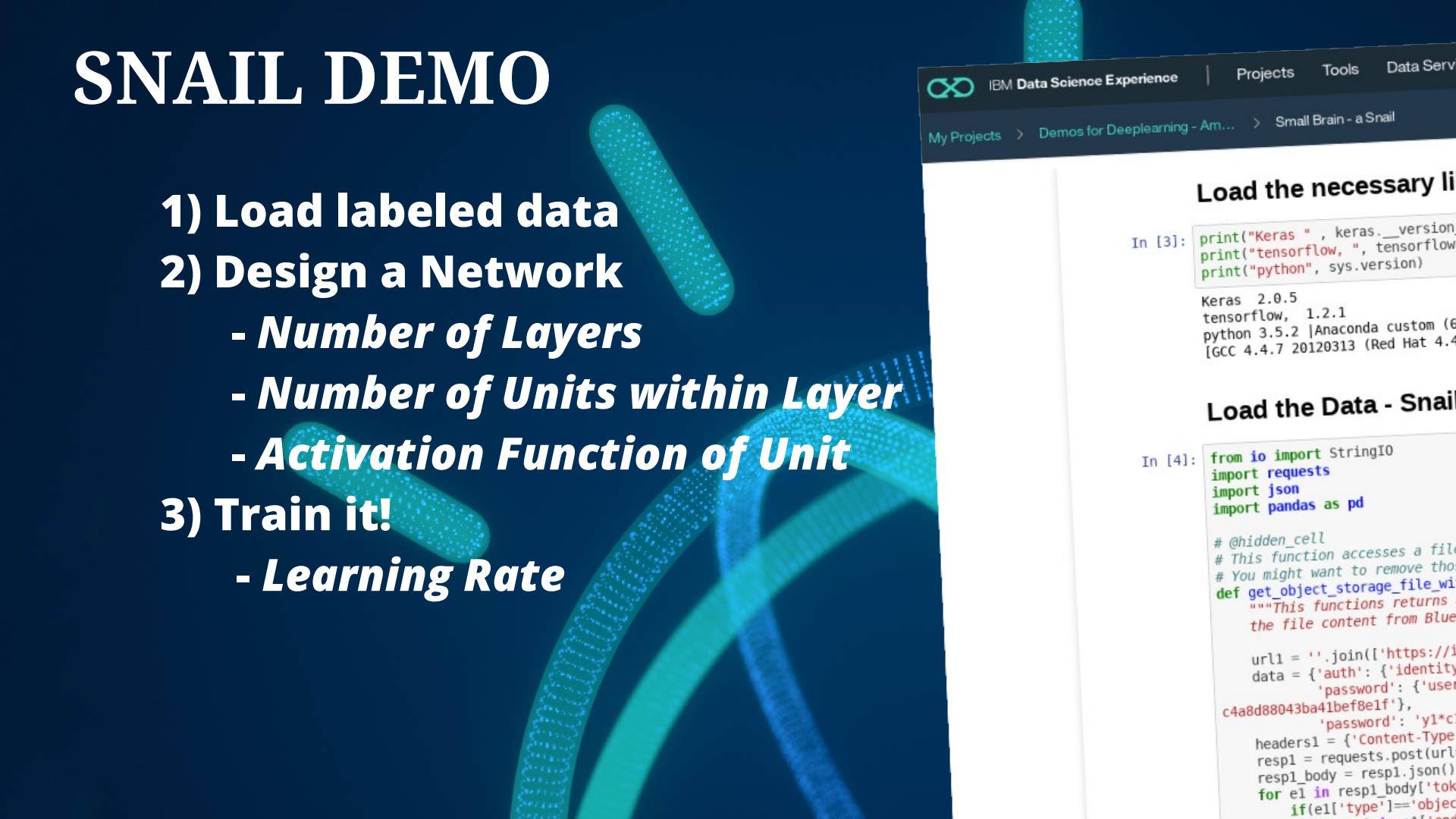
Decrease Learning rate, as the Epochs increase

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*



IBM Data Science Experience | Projects Tools Data Serv

My Projects > Demos for Deeplearning - Am... > Small Brain - a Snail

Load the necessary libraries

```
In [3]: print("Keras ", keras.__version__)
print("tensorflow, ", tensorflow.__version__)
print("python", sys.version)
```

Keras 2.0.5
tensorflow, 1.2.1
python 3.5.2 |Anaconda custom (64-bit)
[GCC 4.4.7 20120313 (Red Hat 4.4.7-1)]

Load the Data - Snail

```
In [4]: from io import StringIO
import requests
import json
import pandas as pd

# @hidden_cell
# This function accesses a file
# You might want to remove this
def get_object_storage_file_with_token():
    """This function returns the file content from BlueMix
    """

    url1 = ''.join(['https://', os.environ['STORAGE_URL'], '/'])
    data = {'auth': {'identity': os.environ['STORAGE_ID'],
                    'password': os.environ['STORAGE_PASSWORD']},
            'c4a8d88043ba41bef8e1f': {
                'password': 'y1*c*'
            }
    }
    headers1 = {'Content-Type': 'application/json'}
    resp1 = requests.post(url1, data=json.dumps(data))
    resp1_body = resp1.json()
    for el in resp1_body['tokens']:
        if(el['type']=='object')
```

Maandag 06 november 2017 | Het laatste nieuws het eerst op NU.nl

11 °C 0 NS 554,29 TV gids 0 Live

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

NU.nl > Overig > Wetenschap

Foto: Thinkstock

Slakken nemen complexe beslissingen met twee hersencellen

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

f t G+

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

Als een slak langs een blaadje sla kruip, 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 de temperatuur en luchtvochtigheid de levenskostie van de dieren.

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...

Google Pixel 2

Ask more of your phone.

Buy now Google Store

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

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

$$f'(x) = 4x^3 + 2 \cdot 3x^2 - 2x^1$$

*How does KERAS
know the
derivatives ?*

$$\frac{d}{dx}[cx] = c$$

$$\frac{d}{dx}[x] = 1$$

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		$O(mn^2)$ ($m \leq n$)
		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



Relation 3D game and GPU and Linear Algebra



Four

Cedric Dickey attacked a teammate.
[Me] Sleepy Weasle attacked a teammate.
=> Match will start when warmups has ended.
Re2lence attacked a teammate.
Bardick attacked a teammate.

Just a trigonometry tip: 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 1×3 matrix, the vector A as another 1×3 matrix, and the linear transformation F as a 3×3 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.



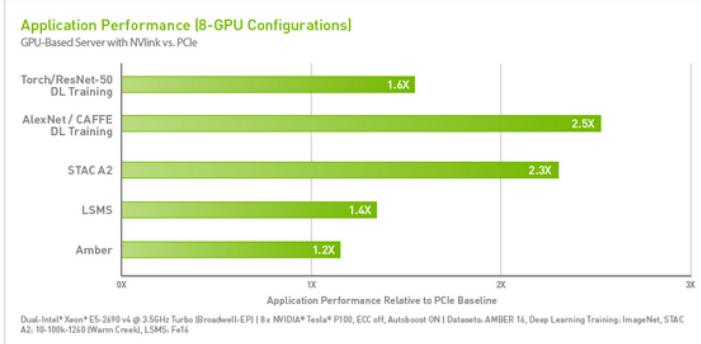
IBM & NVIDIA

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](#).



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

Posted on SEPTEMBER 8, 2016 by IAN BUCK

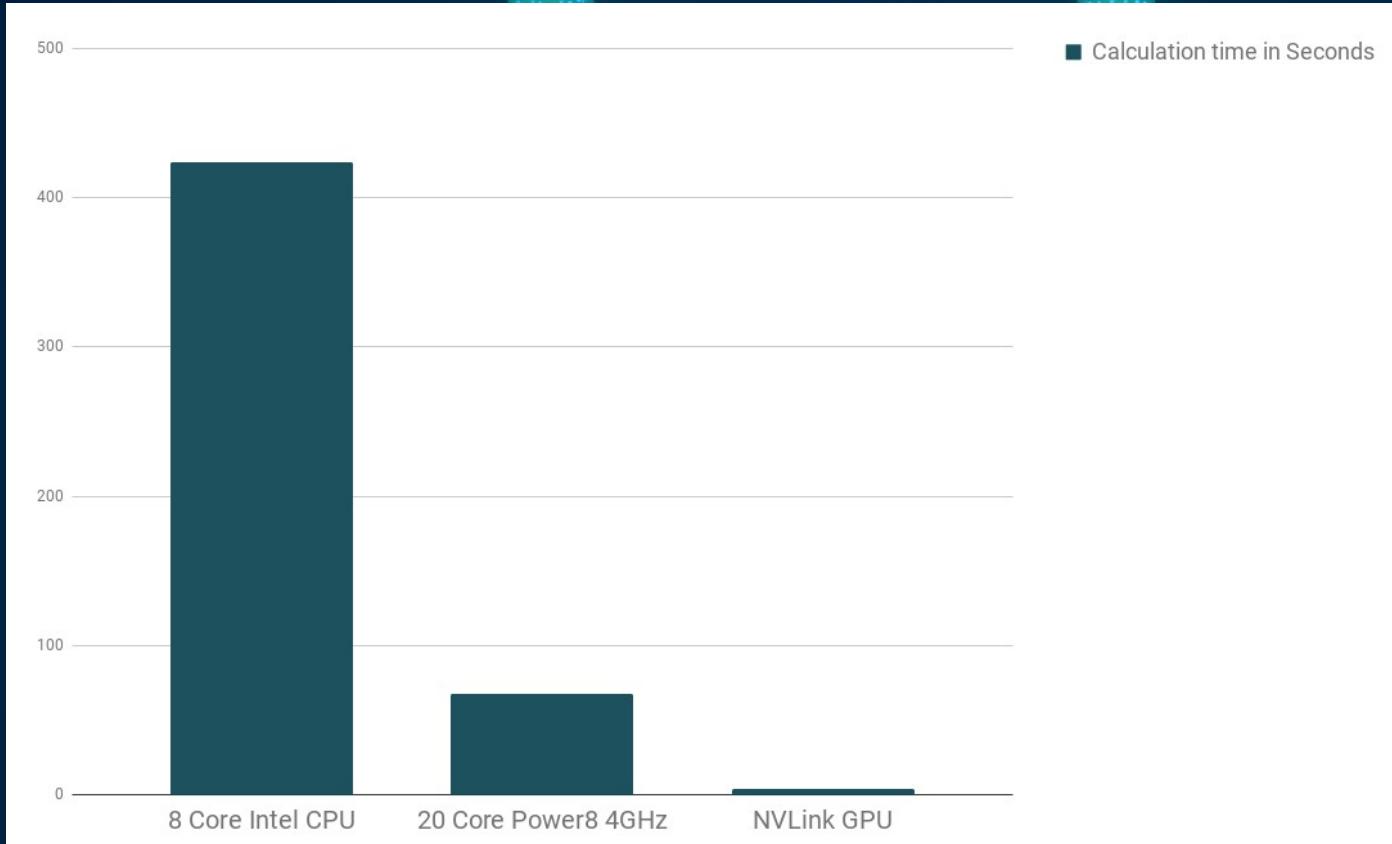
1.6k f g+ in tw

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 come from? In data centers.

... years ago, and partnered with us to ... reduced its

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



IBM

Road Blocks:

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

options continue!

sic // www.emindy.org

1

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...

Brain Myth: Drinking alcohol kills brain cells

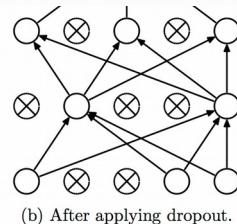
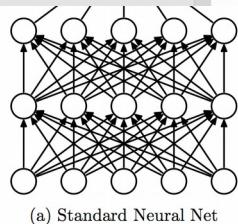
Brain Fact: Moderate alcohol use doesn't kill brain cells, and while rampant alcohol use can damage the brain, it's not due to cell death.

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-myth-alcohol-kills-brain-cells>

The screenshot shows a web browser displaying the brainHQ website. The URL in the address bar is https://www.brainhq.com/brain-resources/cool-brain-facts-myths/brain-myth-alcohol-kills-brain-cells. The page title is "Brain Myth: Drinking alcohol kills brain cells". The main content area contains the text from the slide, followed by a detailed explanation of how alcohol damage is not due to cell death but rather to the degradation of dendrites. The page also includes navigation links like "Home", "Why BrainHQ?", "World Class Science", "Brain Resources", and "Help". A sidebar on the right is titled "Making the Most of Your Brain" and discusses Wernicke-Korsakoff syndrome.

[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.

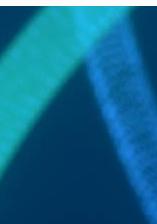
Dropout consists in randomly setting a fraction `rate` of input units to 0 at each update during training time, which helps prevent overfitting.

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 dropout mask that will be multiplied with the input. For instance, if your inputs have shape `(batch_size, timesteps, features)` and you want the dropout mask to have the same shape, you can use `noise_shape=(batch_size, timesteps, features)`.
- `seed`: A Python integer to use as random seed.

References

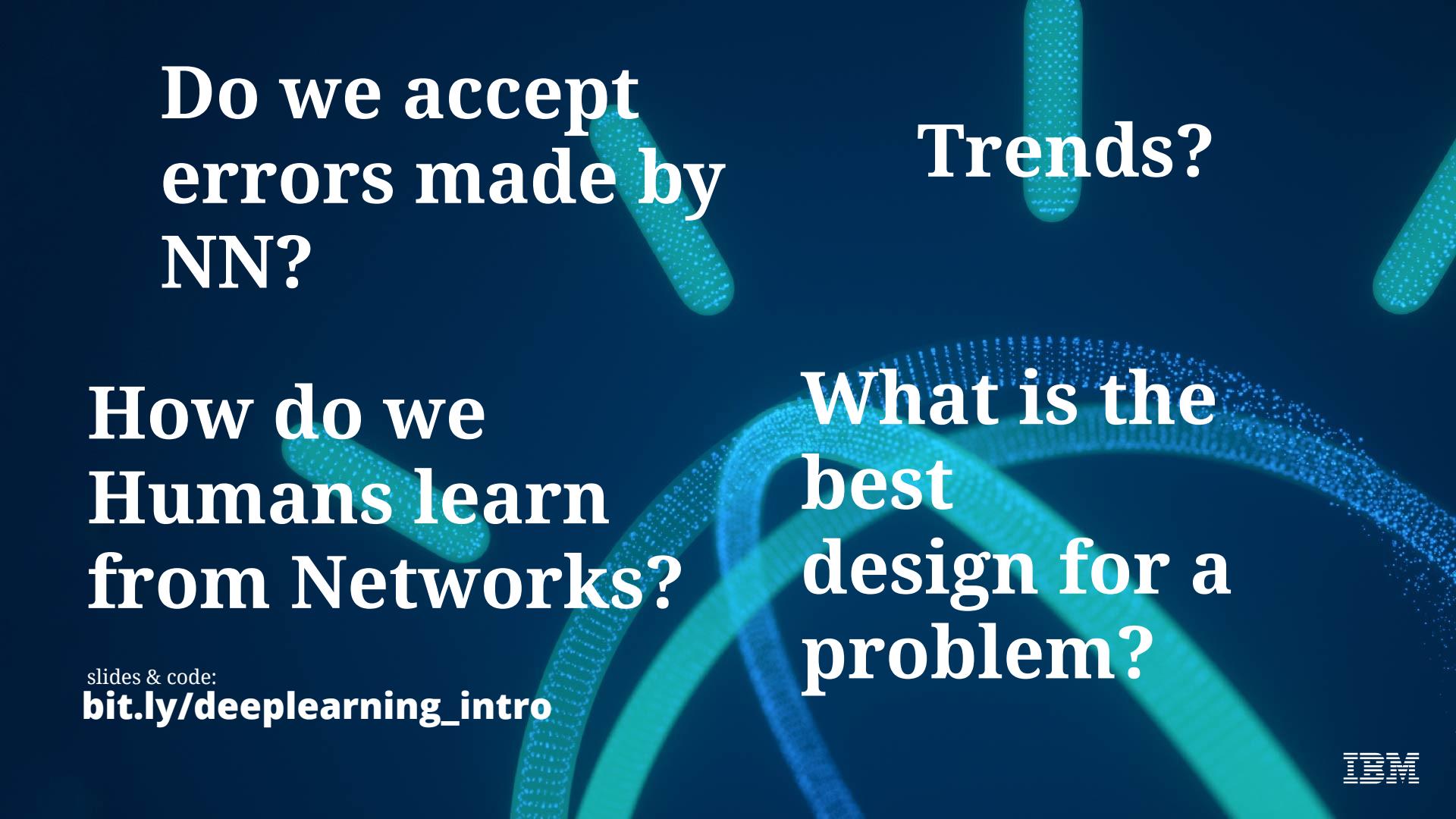
- [Dropout: A Simple Way to Prevent Neural Networks from Overfitting](#)



DROP OUT DEMO

The screenshot shows the IBM Watson Data Platform interface with a Jupyter notebook open. The notebook cell contains Python code for building a neural network with multiple layers of dropout.

```
plot_losses = PlotLosses()  
In [ ]: # just logistic regression, to keep it simple and fast  
model = Sequential()  
model.add(Flatten(input_shape=(28, 28, 1)))  
model.add(Dense(50, activation='relu'))  
model.add(Dropout(0.00))  
model.add(Dense(50, activation='relu', input_shape=(50,)))  
model.add(Dropout(0.00))  
model.add(Dense(50, activation='relu', input_shape=(50,)))  
model.add(Dropout(0.00))  
model.add(Dense(10, activation='sigmoid', input_shape=(50,)))  
sgd = optimizers.SGD(lr=0.03, decay=1e-6, momentum=0.9, nesterov=True)  
#model.compile(optimizer='rmsprop', loss='categorical_crossentropy', metrics=['accuracy'])  
model.compile(optimizer=sgd, loss='categorical_crossentropy', metrics=['accuracy'])  
In [ ]: # in this static viewer it is not obvious,  
# but this plot grows step by step  
model.fit(X_train, Y_train,  
          epochs=20,  
          validation_data=(X_test, Y_test),  
          callbacks=[plot_losses],  
          verbose=0)
```



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/deeplearning_intro

