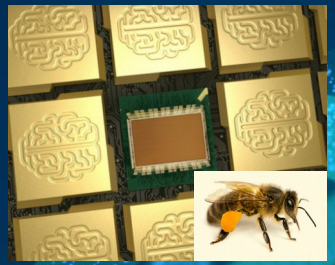
Data Science & Deeplearning for mere mortals

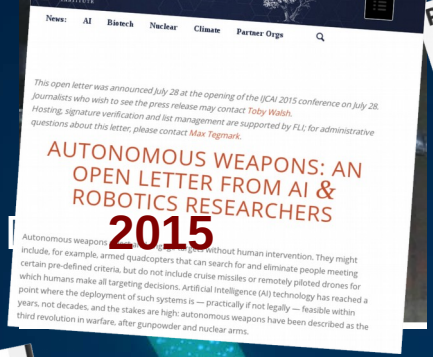
15:30 ~ (16:15)

Willem Hendriks
willem.hendriks@nl.ibm.com

IBM



IBM


News: AI Biotech Nuclear Climate Partner Orgs
AUTONOMOUS WEAPONS: AN OPEN LETTER FROM AI & ROBOTICS RESEARCHERS
2015

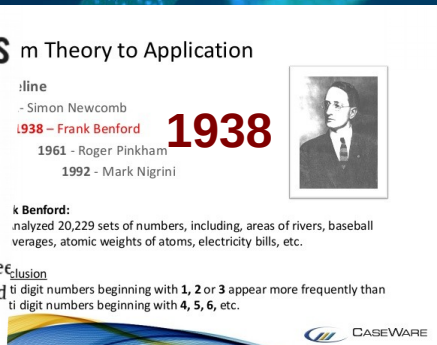

2016
Magento® Is Better
200 Top Internet Retailers Chose Magento Over Other eCommerce Platforms.
Prisma Photo App: Change The Picture Like Picasso Painting!


TECH
IBM's Watson AI Saves Woman's Life
Rare Form of Leukaemia
It took the artificial intelligence just 10 minutes to spot...
08/08/2016 11:24
2016


IBM's brain-mimicking computers are getting bigg
brains
2014


THE MULTIVERSE —
Movie written by AI algorithm turns out to be hilarious and intense
For Sunspring's exclusive debut on Ars, we talked to the filmmakers about collaborating with an AI.
ANNALEE NEWITZ (US) - 9/6/2016, 13:00
2016


LISTEN: Creepy AI Telemarketer Sounds Human, Denies Being A Robot
TOPICS: Activist Post Intelligence
DECEMBER 12, 2013
Activist Post
2013
Time Magazine is investigating a healthcare telemarketing firm who has bee
using an amazingly realistic robot caller which seems to operate on advanced
bit creepy artificial intelligence.


m Theory to Application
eline
- Simon Newcomb
1938 – Frank Benford
1961 - Roger Pinkham
1992 - Mark Nigrini
1938
k Benford:
nalyzed 20,229 sets of numbers, including, areas of rivers, baseball
verages, atomic weights of atoms, electricity bills, etc.
clusion
ti digit numbers beginning with 1, 2 or 3 appear more frequently than
ti digit numbers beginning with 4, 5, 6, etc.
CASEWARE


EMERGENCY STOP
2016


Microsoft
Tay.ai

Data Science

*We apply techniques from **Machine Learning**,*
*on **Data**,*
to gain knowledge / solve problems / create Terminator / recognize a cat / ...

Data Science

Data** has **Features

Example:

***Person** can have age (number), kids (number), married (True / False), as features*

***Image** can have 256x256x3 color values, as features*

Person apply for loan can have Age, Current Balance, and Has Credit-Card (True / False) as features

Data Science

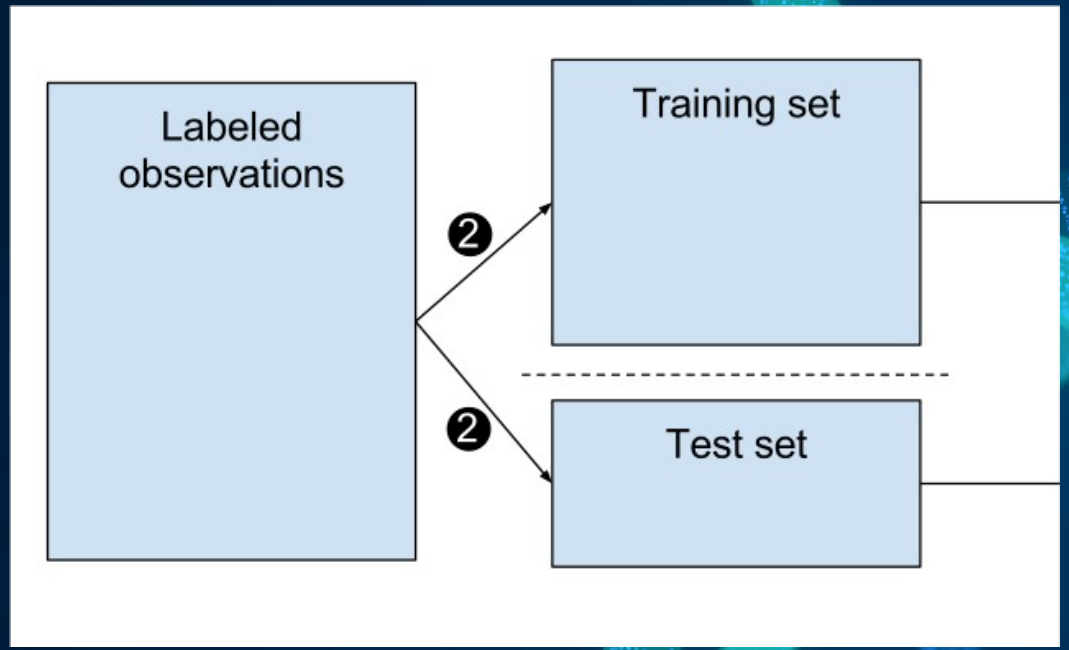
Measure Performance

Special Case of 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 Science

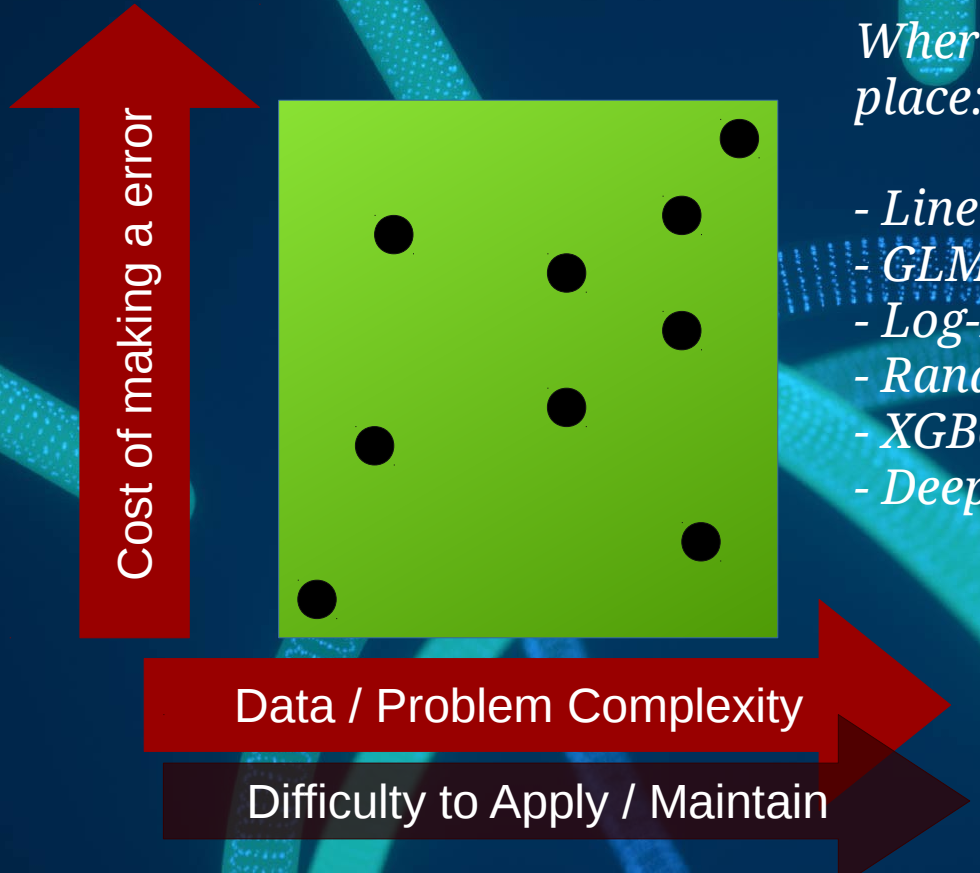


Data Science

Deep-learning!

Supervised Learning

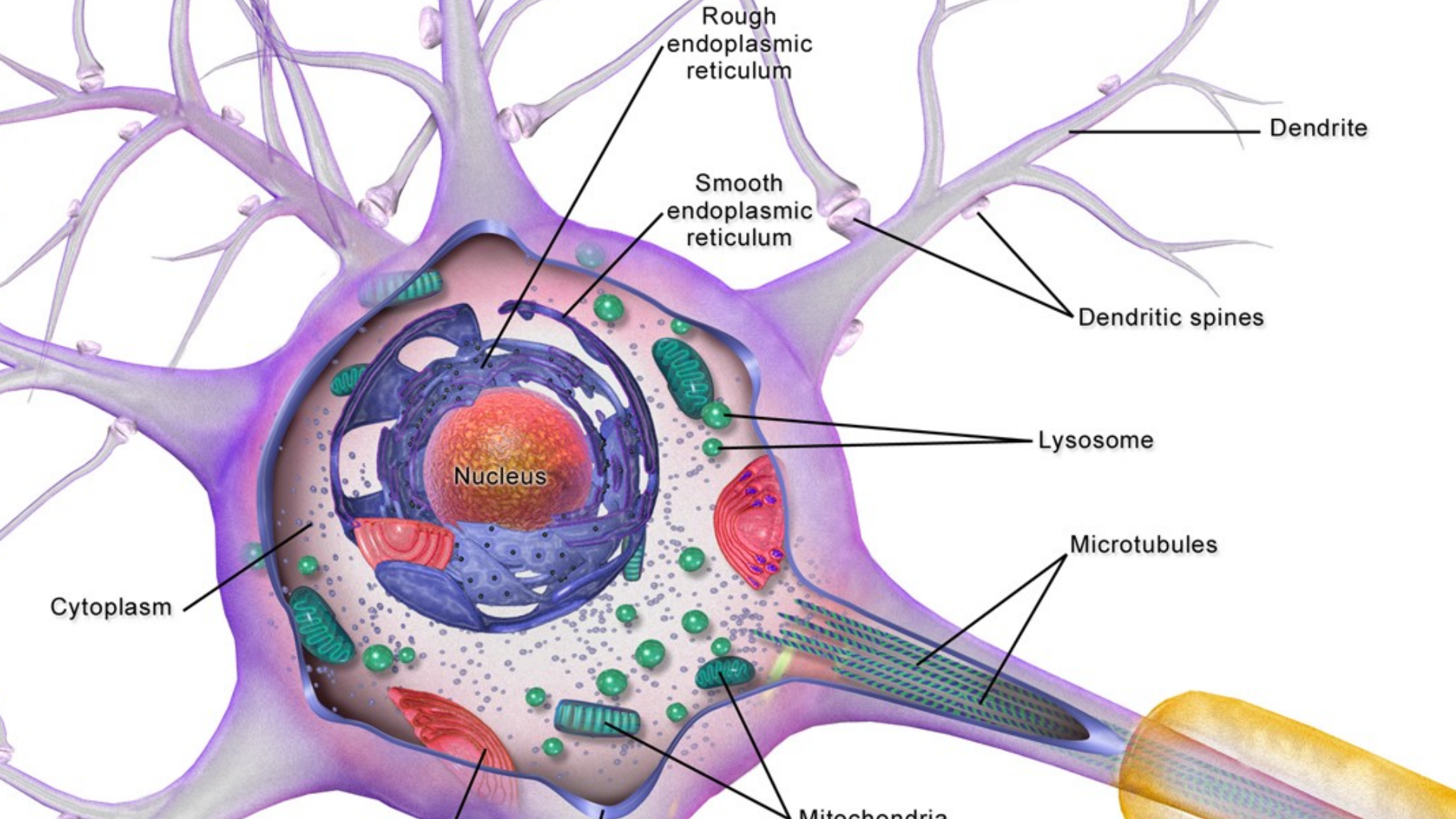
We need Labeled Data



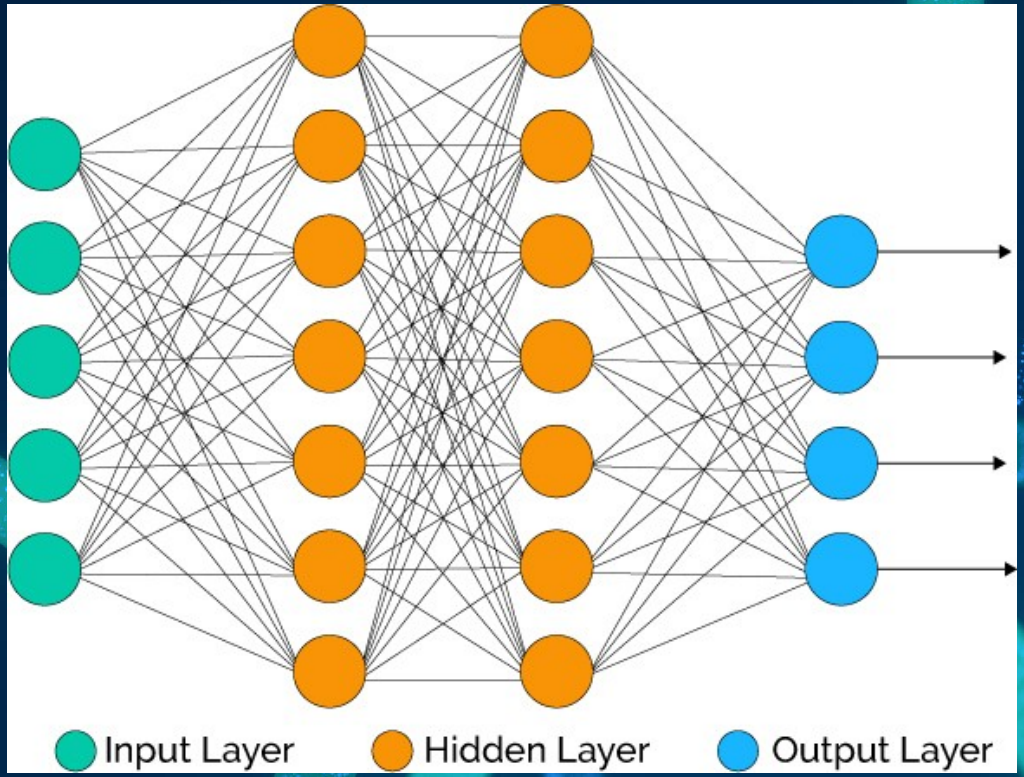
Where would you place:

- *Linear Model*
- *GLM*
- *Log-Logistic Regr.*
- *Random Forest*
- *XGBoost*
- *Deep Learning*

IBM


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

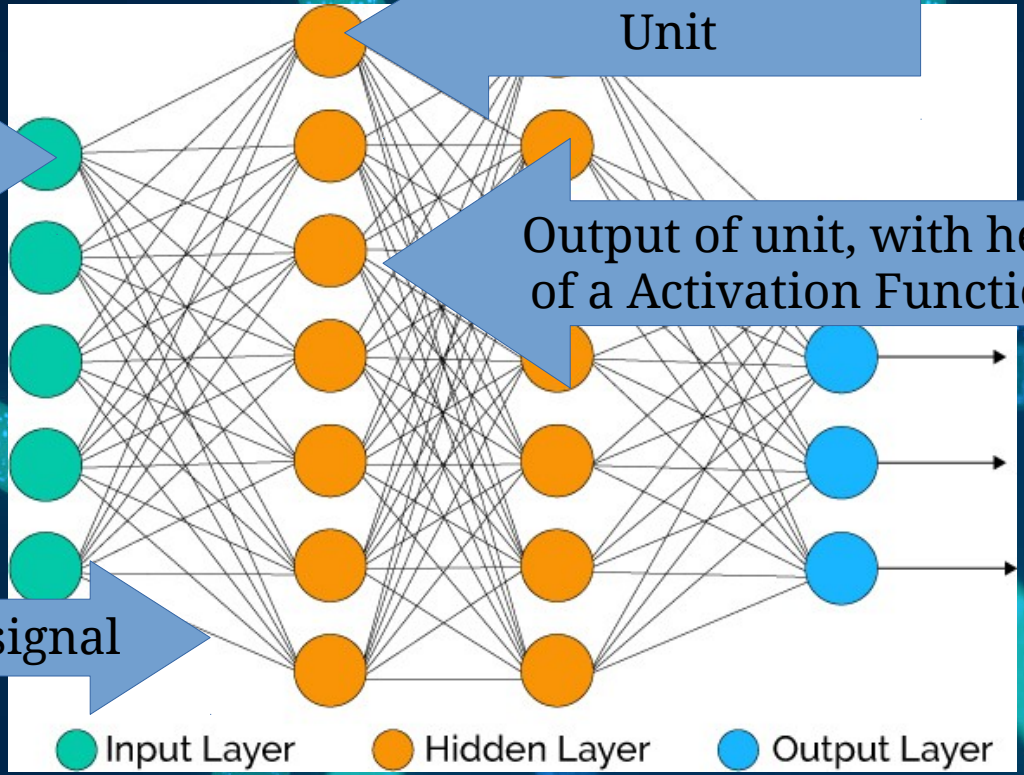
Artificial Neural Network (ANN)



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

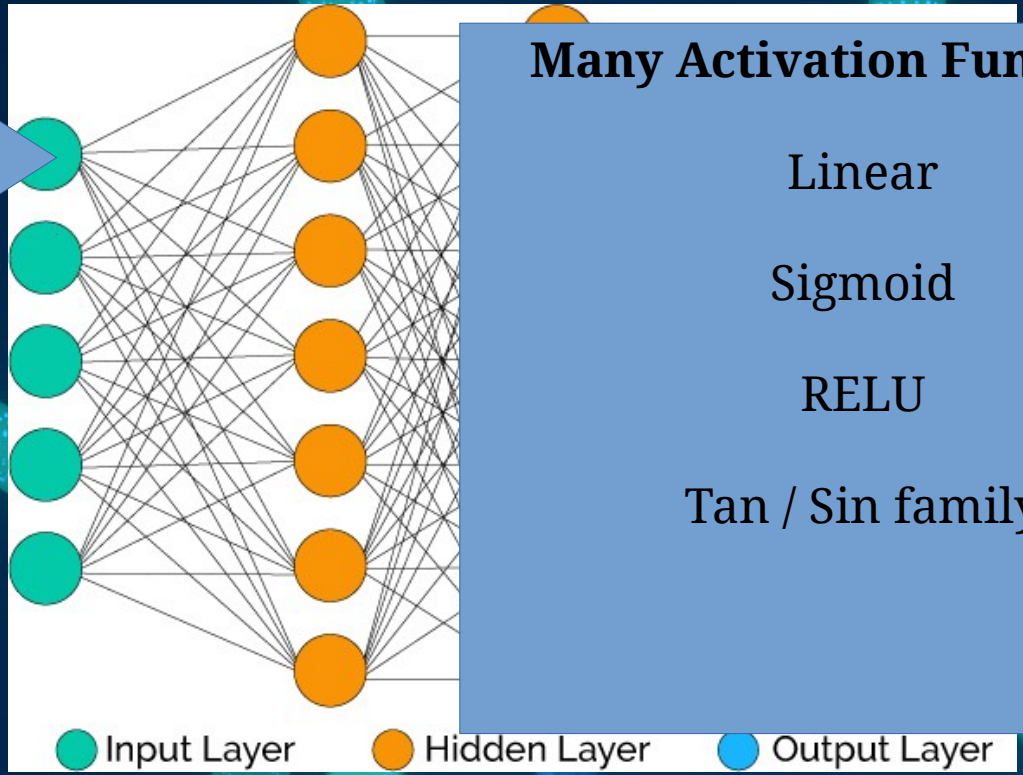
IBM

Artificial Neural Network (ANN)



Artificial Neural Network (ANN)

Feature, e.g. Age



Many Activation Functions:

Linear

Sigmoid

RELU

Tan / Sin family

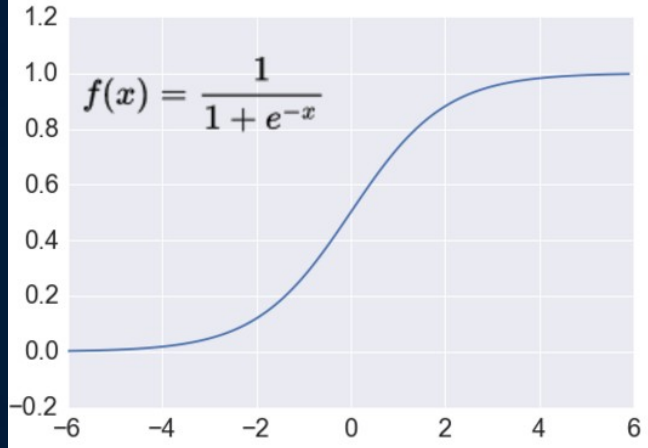
IBM

Artificial Neural Network (ANN)

Common Activation Functions

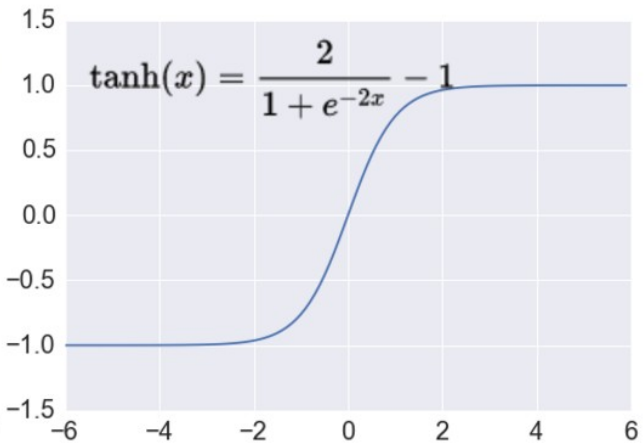
Sigmoid

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



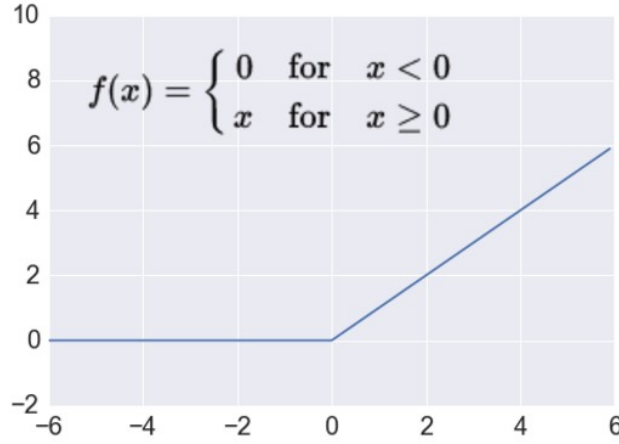
TanH

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



ReLU

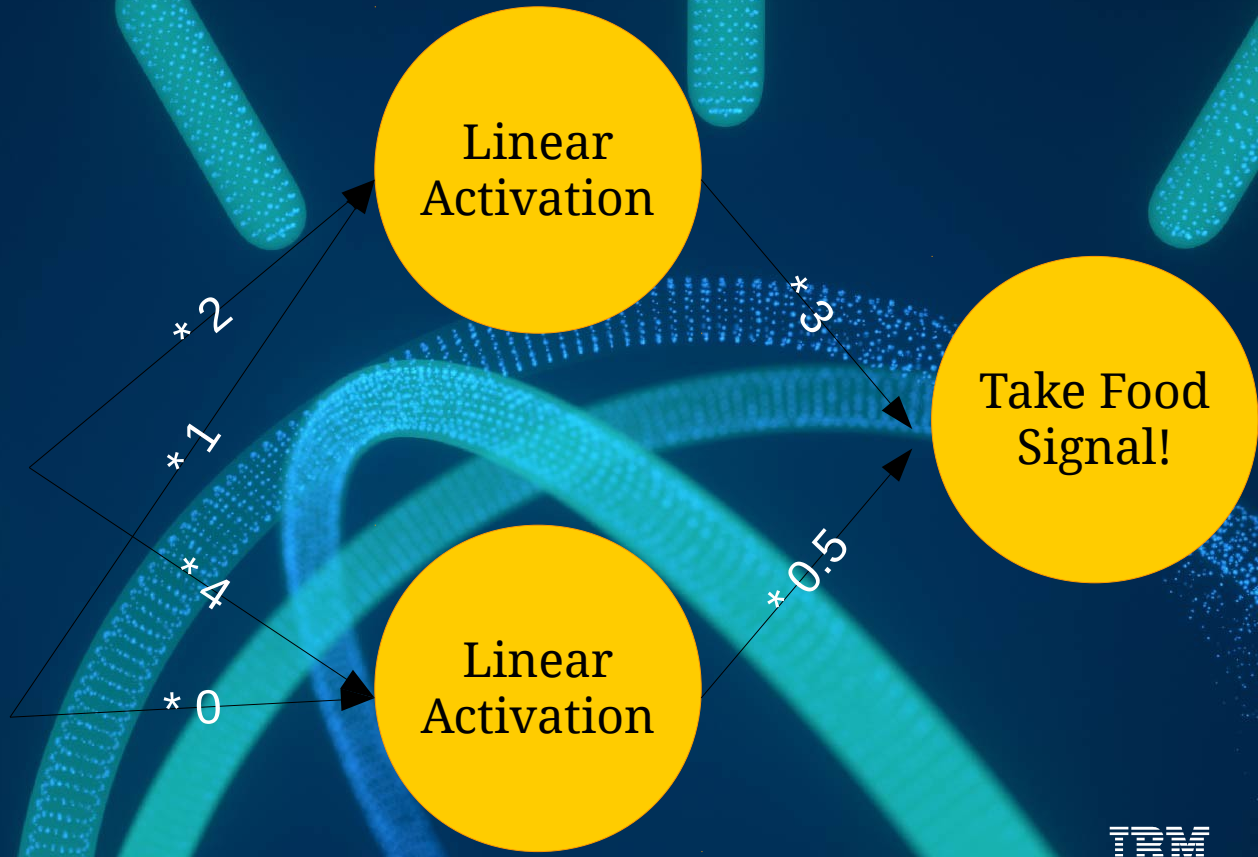
$$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$$



Example - Snail

Feature 1: Hunger Level
Input H: 0.5

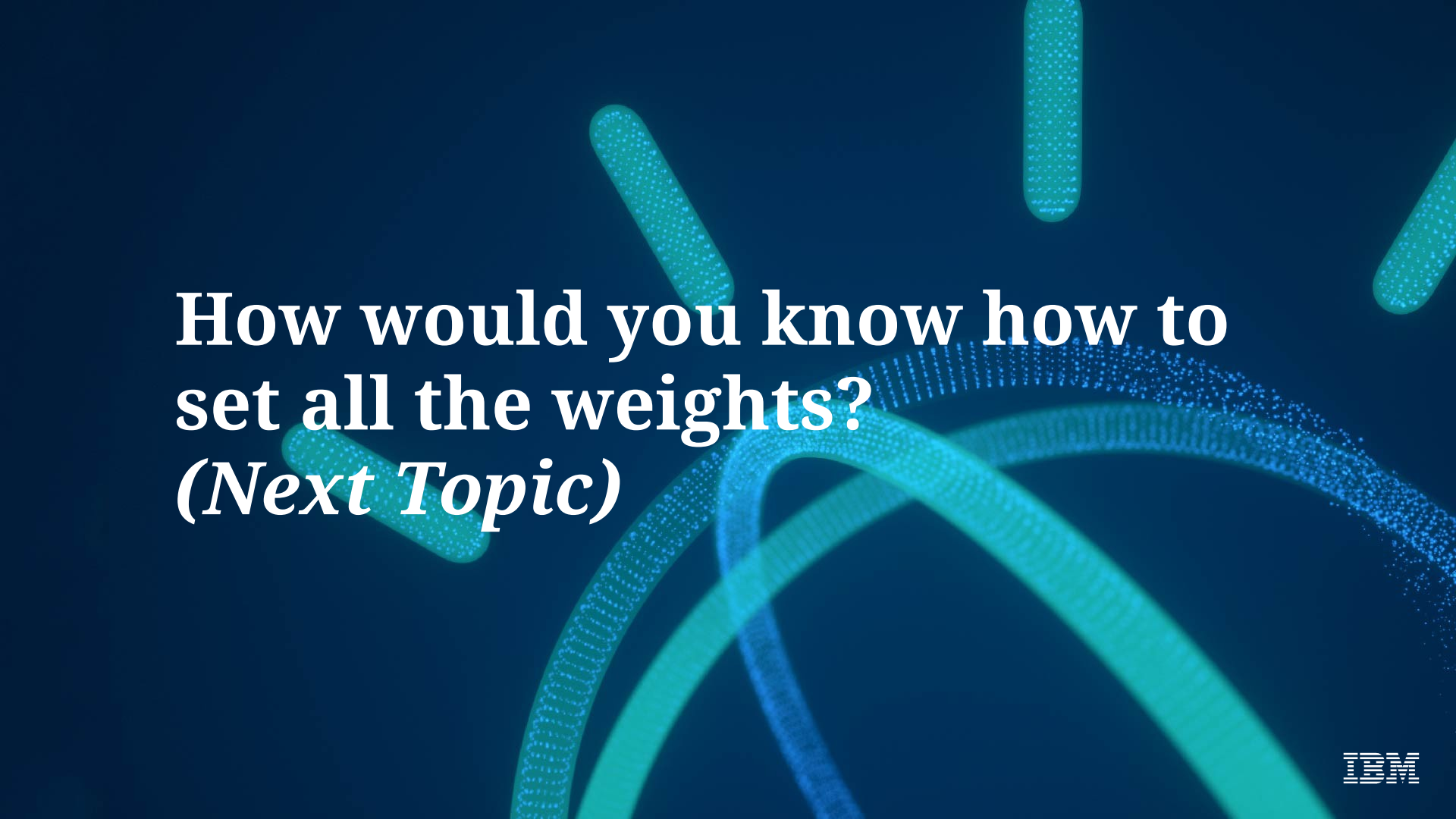
Feature 2: Sees food in Range Level
Input S: 1



IBM

What is the Most Used Activation Function?

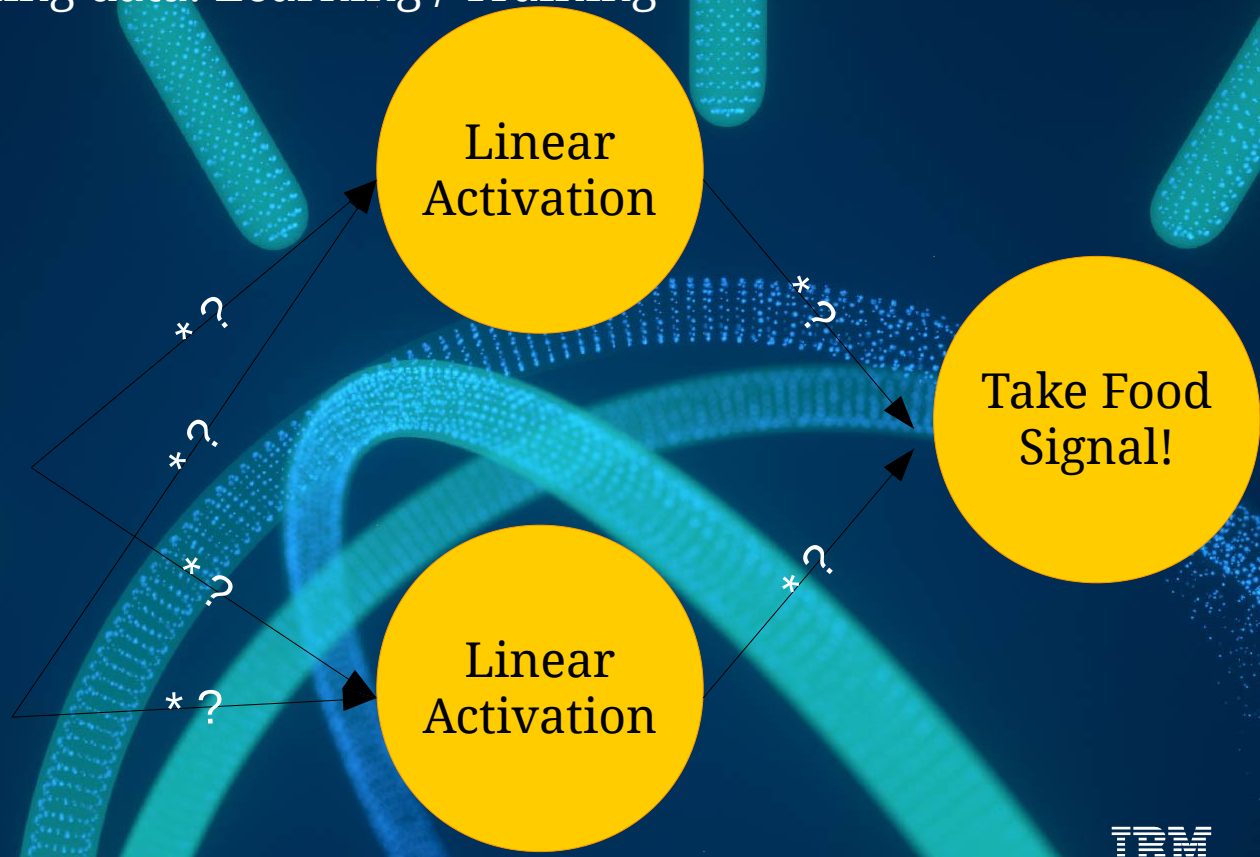
Which Activation mimics our Brain the Best?


How would you know how to set all the weights?
(Next Topic)
IBM

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

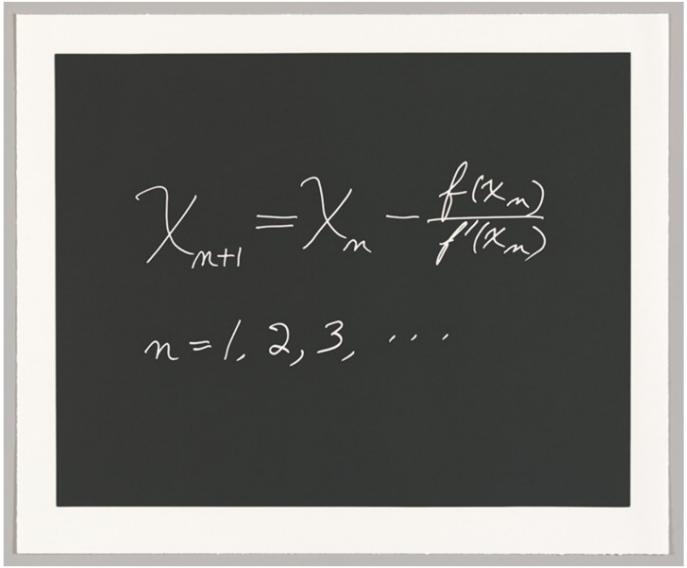
Feature 1: Hunger Level
Input H: 0.5

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

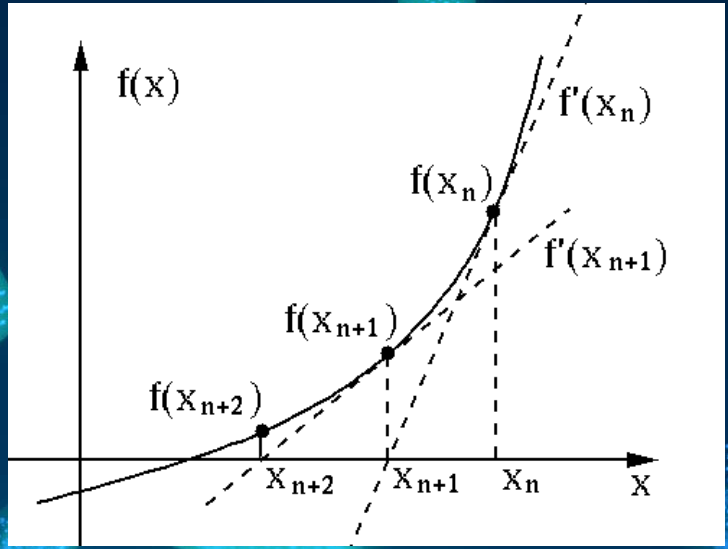


IBM

Stephen Smale: "Newton's Method"



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

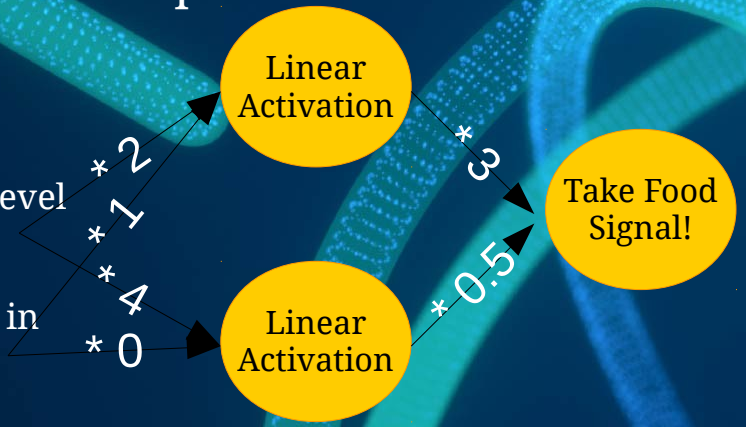


During the training Phase, of a Neural Network,

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

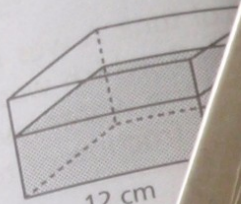
Network Output: 6

Labeled Data: 10

Error: 2
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

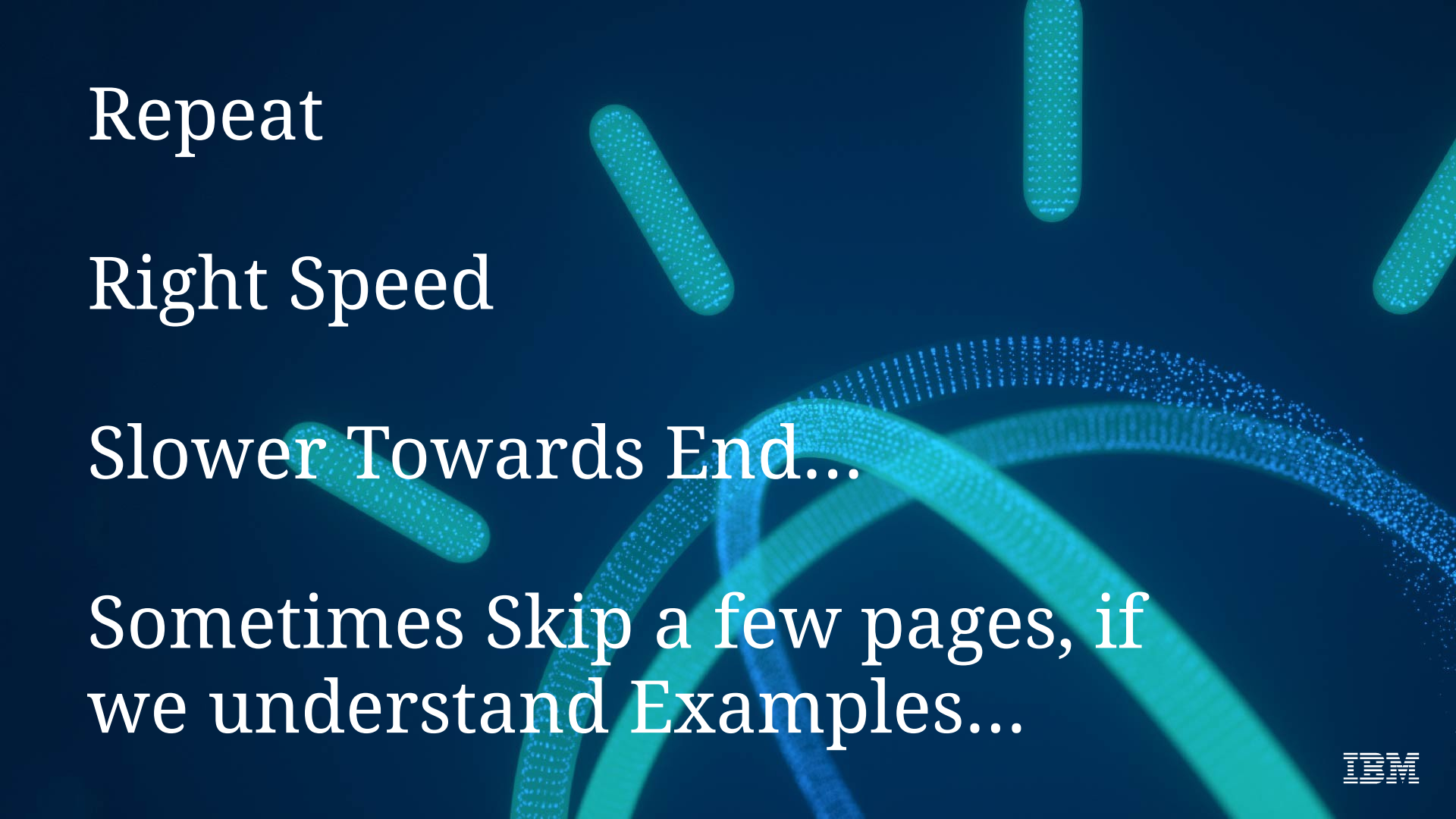
crease) in height may be co
nple 3. Using this method
orking.

Repeat

Right Speed

Slower Towards End…

Sometimes Skip a few pages, if we understand Examples…



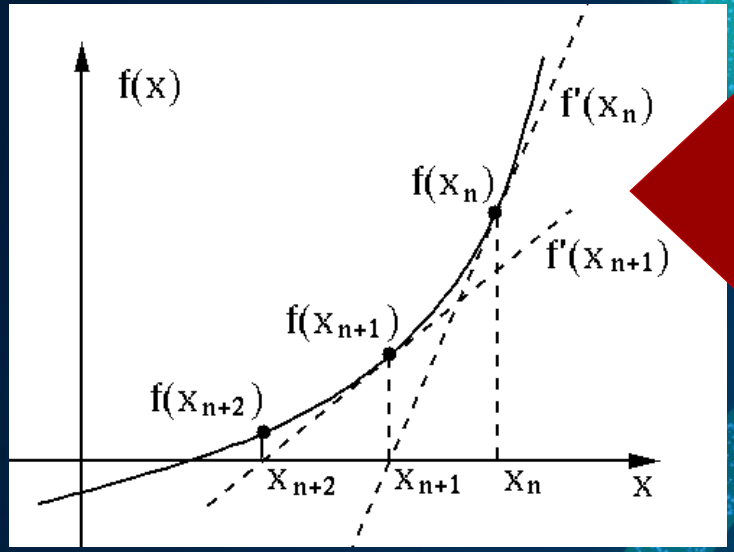
Repeat, **Epochs**

Right Speed, **Learning Rate**

Slower Towards End… **Decay**

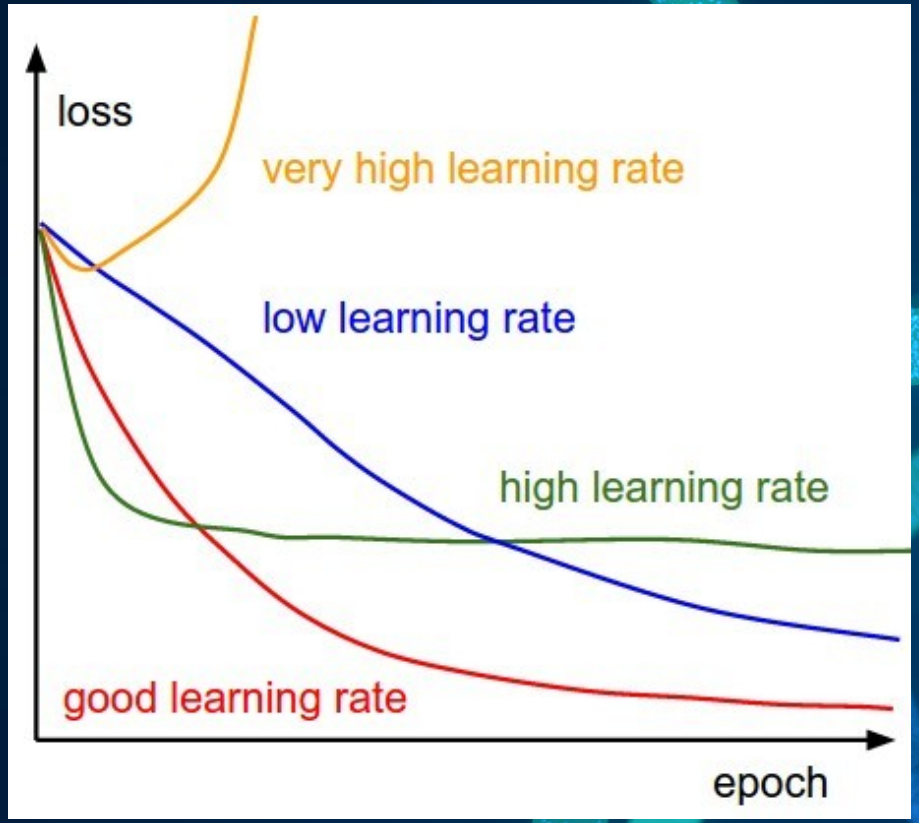
Sometimes Skip a few pages **Momentum**

Right Speed, **Learning Rate**



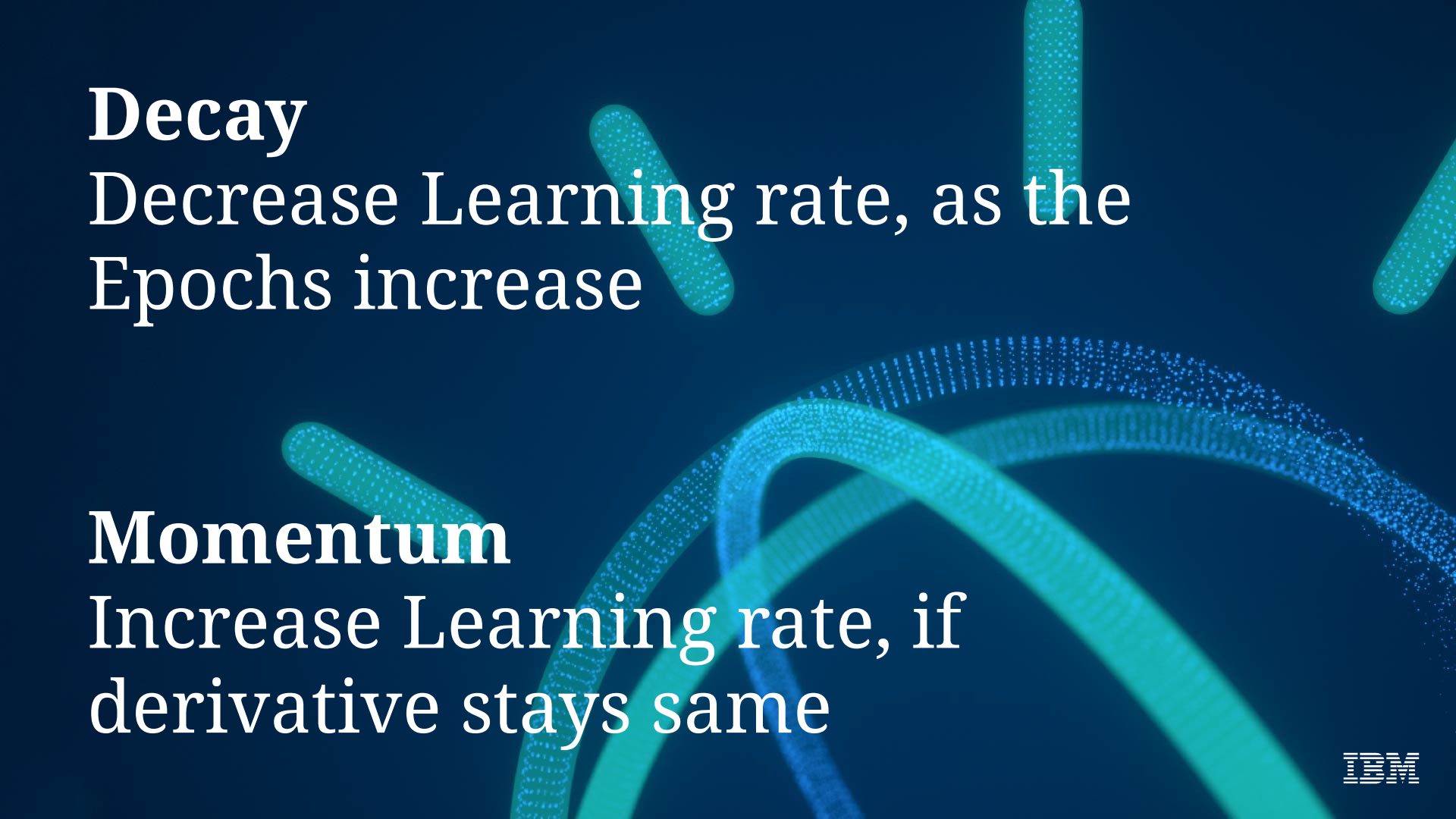
The size of the step, towards the right direction

Right Speed, **Learning Rate**



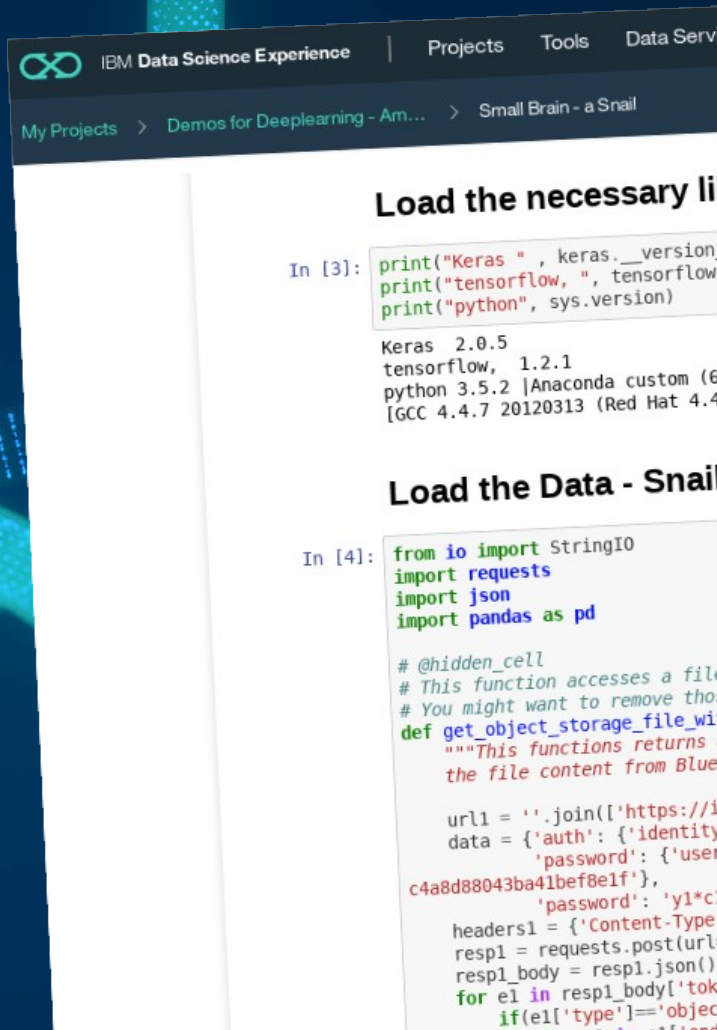
Decay
Decrease Learning rate, as the Epochs increase

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

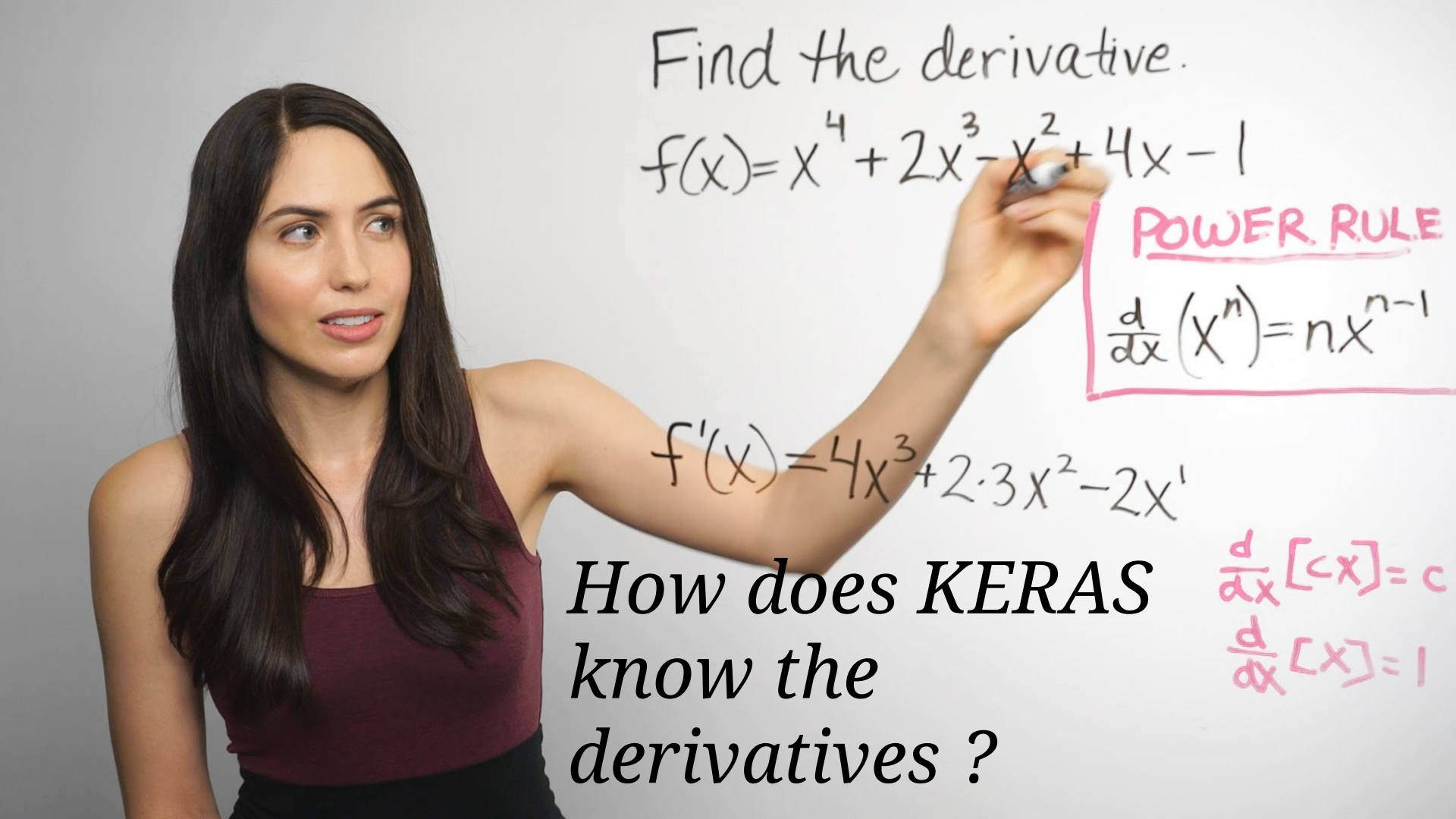
Net binnen

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
- Over-fitting
- Won't learn at all
- Design not Optimal
- Not Fast Enough (GPU)
- “ERROR: Cannot use operator on 232x3 and 12x3”
- …

IBM

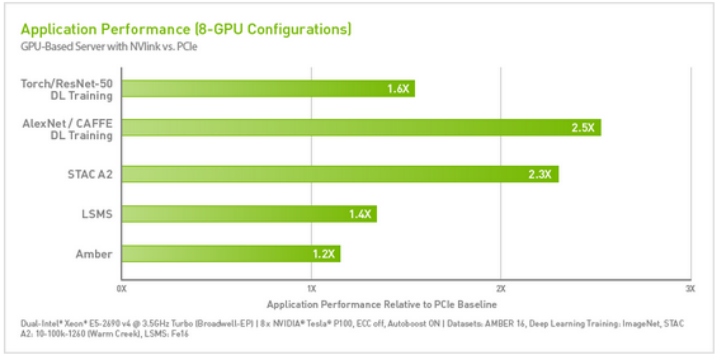
IBM & NVIDIA

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

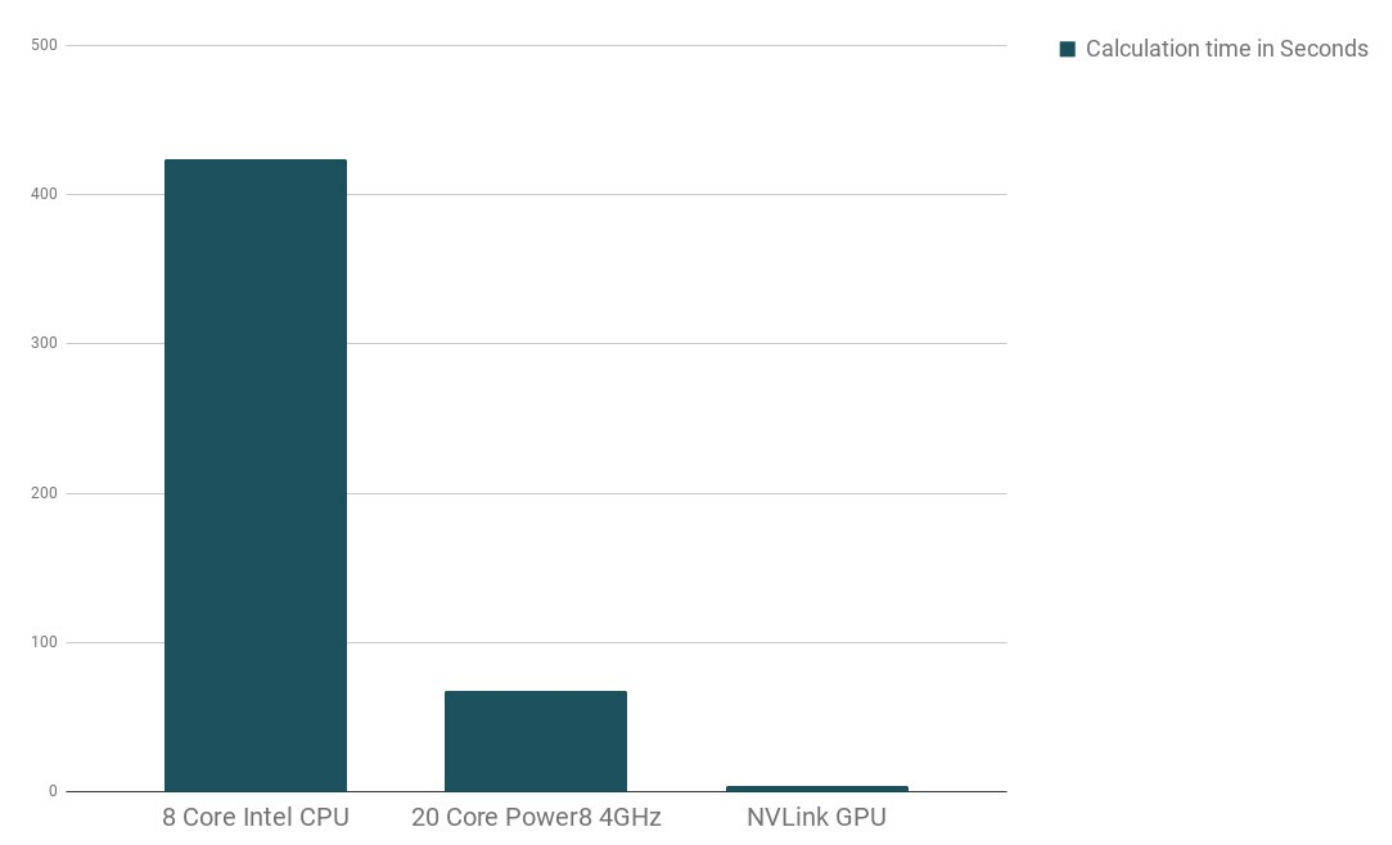
Posted on SEPTEMBER 8, 2016 by IAN BUCK

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 is here behind us in A'dam
NVLink on Power



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 to teach

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.

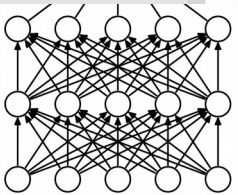
Dropout consists in randomly setting a fraction `rate` of input units to 0 at each update during training time, which helps prevent overfitting.

Arguments

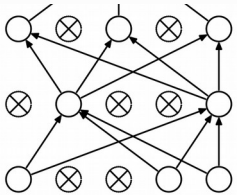
- **rate**: float between 0 and 1. Fraction of the input units to drop.
- **noise_shape**: 1D integer tensor representing the shape of the binary drop
  multiplied with the input. For instance, if your inputs have shape
  `(batch_size, timesteps, features)` and you want the dropout m
  timesteps, you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

References

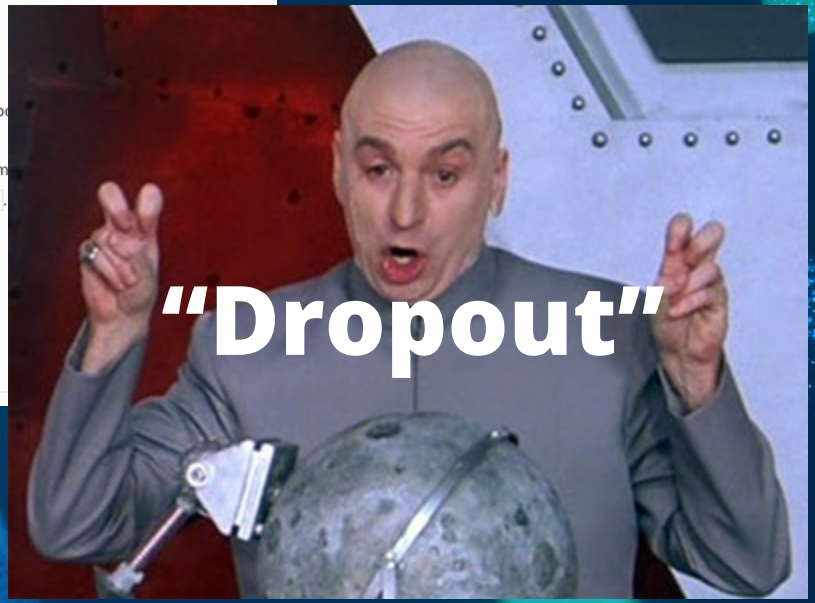
- Dropout: A Simple Way to Prevent Neural Networks from Overfitting



(a) Standard Neural Net

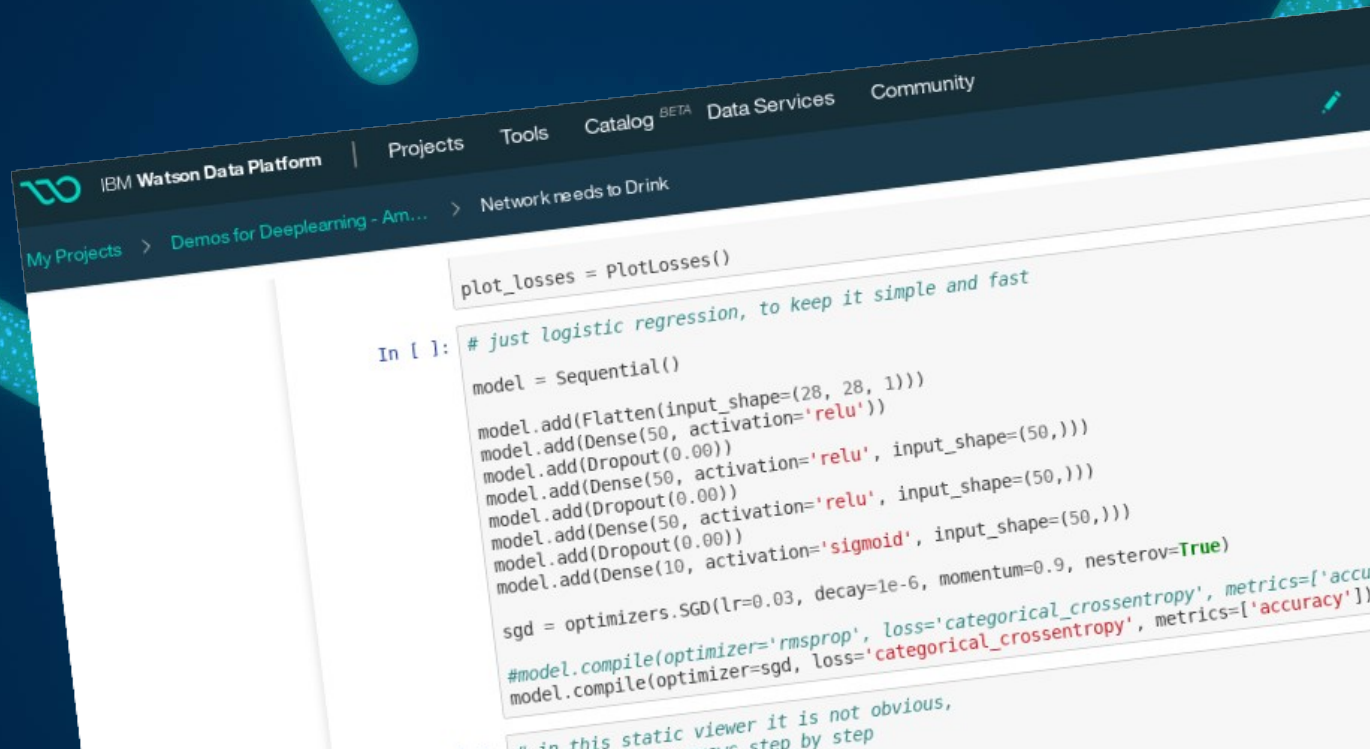


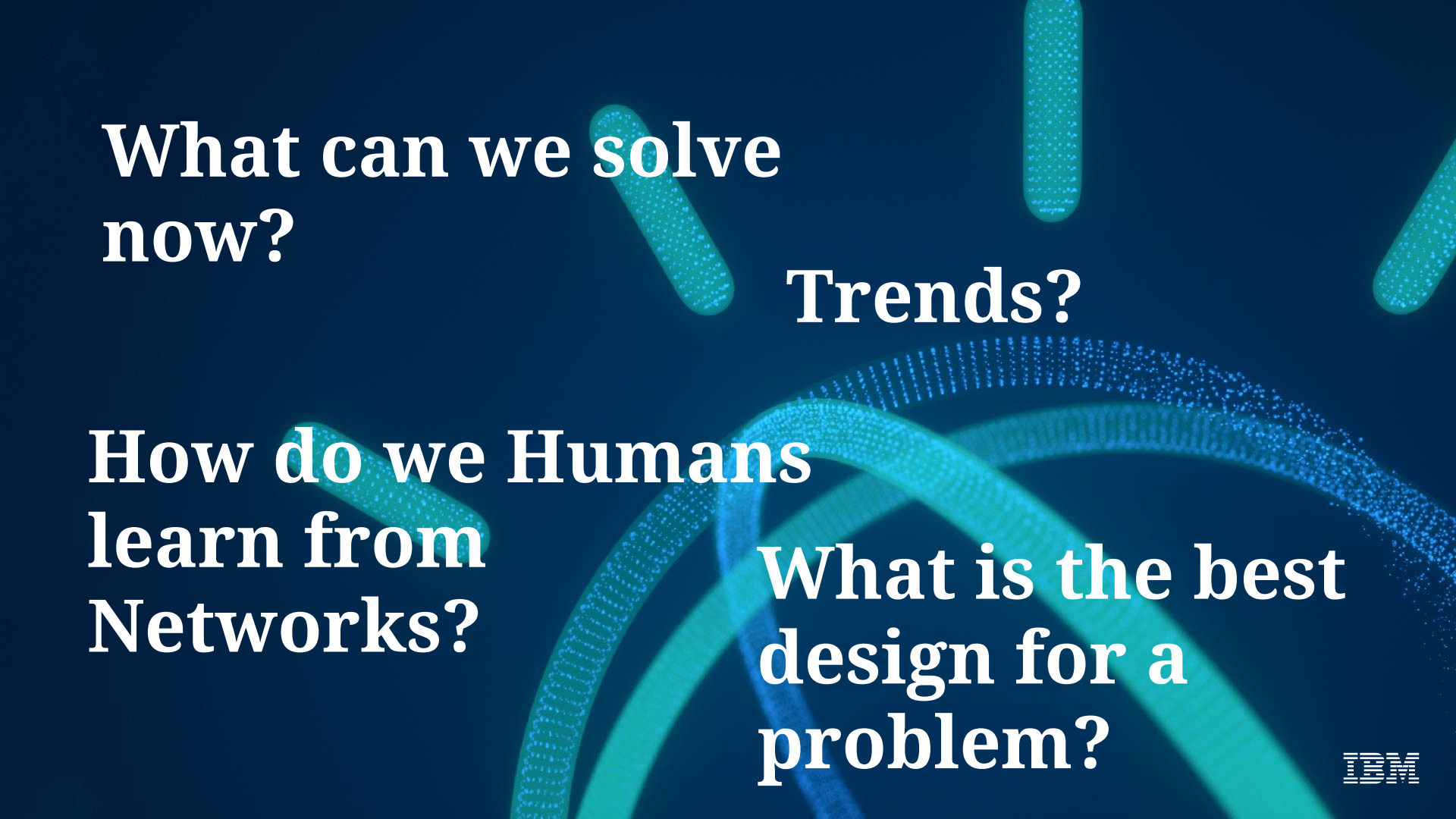
(b) After applying dropout.



IBM

DROP OUT DEMO




What can we solve now?
Trends?
How do we Humans learn from Networks?
What is the best design for a problem?
IBM