Hortonworks

Deep Learning on HDP

Dhruv Kumar – dkumar@hortonworks.com
Solutions Engineer, Hortonworks

2015
Version 1.0

MIT Technology Review

10 BREAKTHROUGH TECHNOLOGIES 2013

Deep Learning

With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart. →

Temporary Social Media

Messages that quickly self-destruct could enhance the privacy of online communications and make people freer to be spontaneous. →

Prenatal DNA Sequencing

Reading the DNA of fetuses will be the next frontier of the genomic revolution. But do you really want to know about the genetic problems or musical aptitude of your unborn child? →

Additive Manufacturing

Skeptical about 3-D printing? GE, the world's largest manufacturer, is on the verge of using the technology to make jet parts. →

Baxter: The Blue-Collar Robot

Rodney Brooks's newest creation is easy to interact with, but the complex innovations behind the robot show just how hard it is to get along with people. →

Memory Implants

A maverick neuroscientist believes he has deciphered the code by which the brain forms long-term memories. Next: testing a prosthetic implant for people suffering from long-term memory loss. →

Smart Watches

The designers of the Pebble watch realized that a mobile phone is more useful if you don't have to take it out of your pocket. →

Ultra-Efficient Solar Power

Doubling the efficiency of a solar cell would completely change the economics of renewable energy. Nanotechnology just might make it possible. →

Big Data from Cheap Phones

Collecting and analyzing information from simple cell phones can provide surprising insights into how people move about and behave – and even help us understand the spread of diseases. →

Supergrids

A new high-power circuit breaker could finally make highly efficient DC power grids practical. →

The New York Times

Scientists See Promise in Deep-Learning Programs

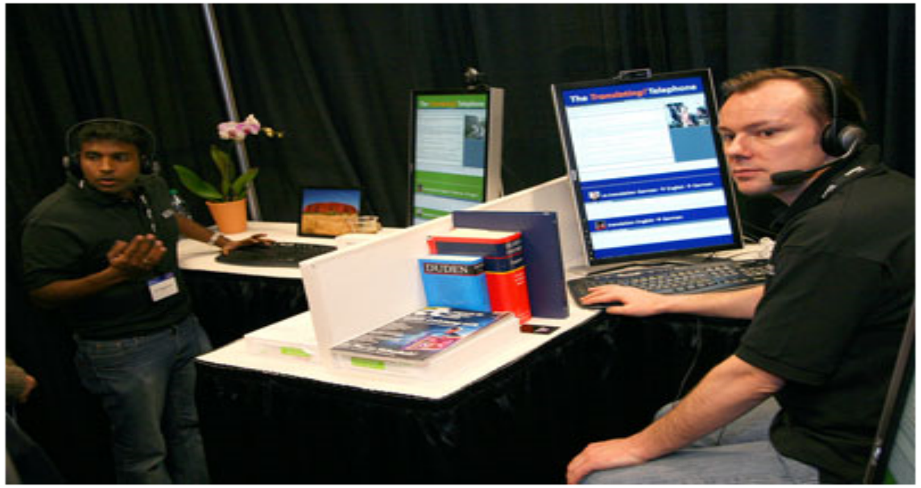
John Markoff
November 23, 2012

Rich Rashid in **Tianjin**, October, 25, 2012

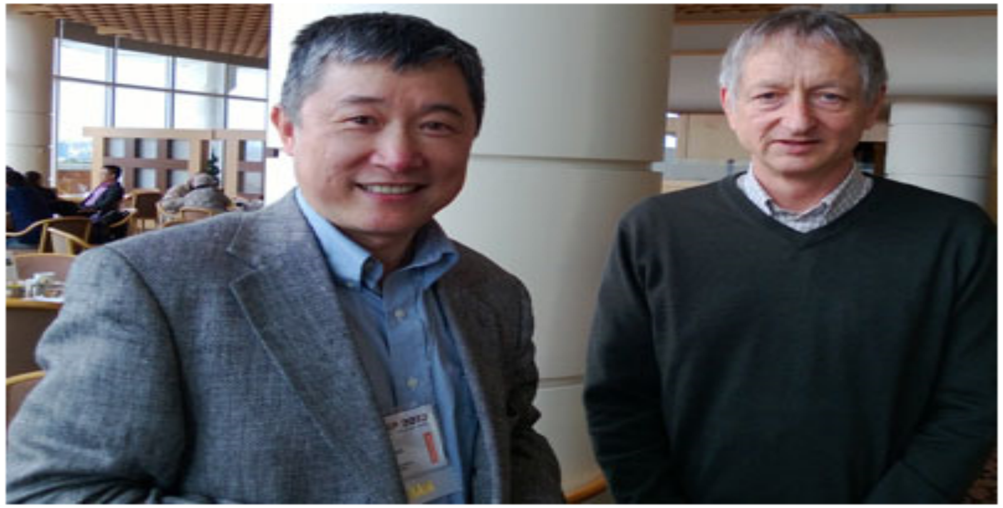
Enabling Cross-Lingual Conversations in Real Time

Microsoft Research
May 27, 2014 5:58 PM PT

The success of the team's progress to date was on display May 27, in a talk by Microsoft CEO Satya Nadella in Rancho Palos Verdes, Calif., during the Code Conference. During Nadella's conversation with Kara Swisher and Walt Mossberg of the Re/code tech website relating to a new
e asked
tage. Pall, the
ident of
first time
app, with Pall
rman-
Diana



The path to the Skype Translator gained momentum with an encounter in the autumn of 2010. Seide and colleague Kit Thambiratnam had developed a system they called The Translating! Telephone for live speech-to-text and speech-to-speech trans calls.



Li Deng (*left*) and Geoff Hinton.

A core development that enables Skype translation came from Redmond researcher Li Deng. He invited Geoff Hinton, a professor at the University of Toronto, to visit Redmond in 2009 to work on new neural-network learning methods, based on a couple of seminal papers from Hinton and his collaborators in 2006 that had brought new

Where is it?

Hortonworks

MIT Technology Review

September 20, 2013

COMPUTING NEWS

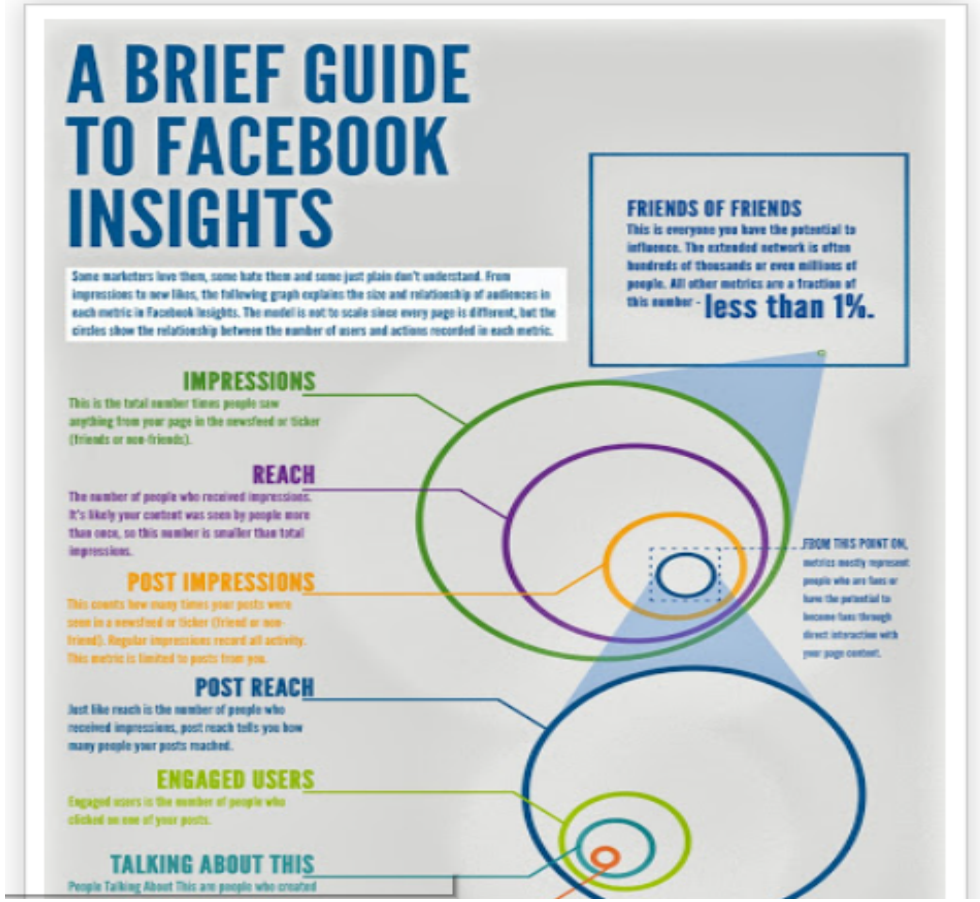
8 COMMENTS

Facebook Launches Advanced AI Effort to Find Meaning in Your Posts

A technique called deep learning could help Facebook understand its users and their data better.

By Tom Simonite on September 20, 2013

……Facebook's foray into deep learning sees it following its competitors Google and Microsoft, which have used the approach to impressive effect in the past year. Google has hired and acquired leading talent in the field (see "10 Breakthrough Technologies 2013: Deep Learning"), and last year created software that taught itself to recognize cats and other objects by reviewing stills from YouTube videos. The underlying deep learning technology was later used to slash the error rate of Google's voice recognition services (see "Google's Virtual Brain Goes to Work") ….Researchers at Microsoft have used deep learning to build a system that translates speech from English to Mandarin Chinese in real time (see "Microsoft Brings Star Trek's Voice Translator to Life"). Chinese Web giant Baidu also recently established a Silicon Valley research lab to work on deep learning.



MIT Technology Review

BUSINESS NEWS

8 COMMENTS

Is Google Cornering the Market on Deep Learning?

A cutting-edge corner of science is being wooed by Silicon Valley, to the dismay of some academics.

By Antonio Regalado on January 29, 2014

How much are a dozen deep-learning researchers worth? Apparently, more than $400 million.

This week, Google reportedly paid that much to

BloombergBusinessweek
Technology

Acquisitions

The Race to Buy the Human Brains Behind Deep Learning Machines

By Ashlee Vance | January 27, 2014

intelligence projects. “DeepMind is bona fide in terms of its research capabilities and depth,” says Peter Lee, who heads Microsoft Research.

According to Lee, Microsoft, Facebook (FB), and Google find themselves in a battle for deep learning talent. Microsoft has gone from four full-time deep learning experts to 70 in the past three years. “We would have more if the talent was there to be had,” he says. “Last year, the cost of a top, world-class deep learning expert was about the same as a top NFL quarterback prospect. The cost of that talent is pretty remarkable.”

   Hortonworks

Deep Learning's Role in the Age of Robots

BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM

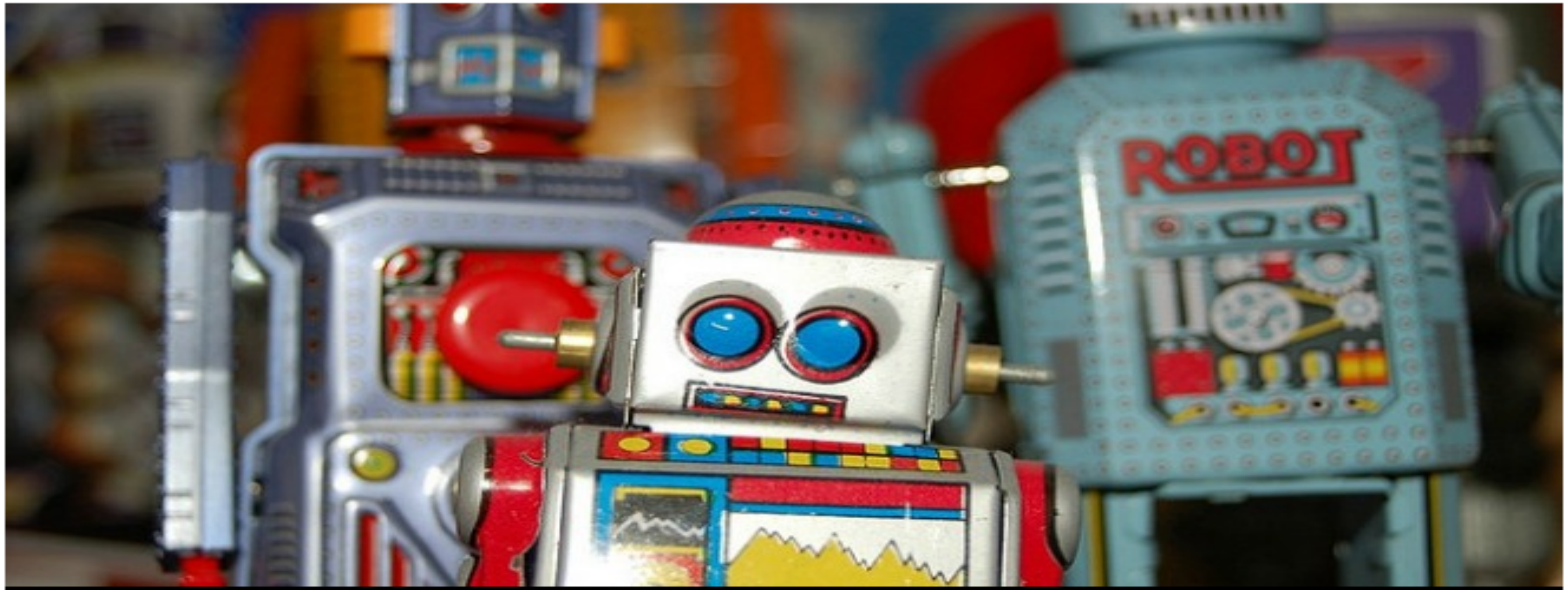


Image: jeffedoe/Flickr

Can robots see as well as humans? That's a question the biggest companies around are trying to answer.

Hortonworks

MIT Technology Review

COMPUTING NEWS

4 COMMENTS

Chinese Search Giant Baidu Hires Man Behind the "Google Brain"

Leading AI researcher Andrew Ng, previously associated with Google, will lead a new effort by China's Baidu to create software that understands the world.

By Tom Simonite on May 16, 2014

Baidu has long been referred to as "China's Google" because it dominates Web search in the country. Today the comparison grew more apt: Baidu has opened a new artificial-intelligence research lab in Silicon Valley that will be overseen by

MIT Technology Review

NEWS & ANALYSIS | FEATURES | VIEWS | MULTIMEDIA | DISCUSSIONS | TOPICS

POPULAR: 10 BREAKTHROUGH TECHNOLOGIES

China's Baidu Bets on **Deep Learning**

MIT Technology Review - 3 days ago

Deep learning makes it possible for machines to process large amounts of date using simulated networks of simple neurons, crudely modeled ...

Baidu snatches Google's **deep-learning** visionary, Andrew ...

VentureBeat - 3 days ago

Hortonworks

GIGAOM EVENTS RESEARCH JOBS

APPLE CLEANTECH CLOUD DATA EUROPE MEDIA MOBILE

artificial intelligence / machine-learning / natural language processing

DARPA is working on its own deep-learning project for natural-language processing

by Derrick Harris MAY. 2, 2014 - 10:49 AM PDT

2 Comments +1

SUMMARY: *The Defense Advanced Research Projects Agency, or DARPA, is building a set of technologies to help it better understand human language so it can analyze speech and text sources and alert analysts of potentially useful information.*

photo: DARPA

When it comes to large organizations working on artificial intelligence systems for understanding language, there's Google, Microsoft, Yahoo and … the Defense Advanced Research Projects Agency. The agency, better known as DARPA, is working on a project it calls Deep Exploration and Filtering of Text, or DEFT, in order to analyze textual data at a scale beyond what humans could do by themselves.

Hortonworks

Forbes / Tech

DEC 29, 2014 @ 8:37 AM 79,012 VIEWS

Tech 2015: Deep Learning And Machine Intelligence Will Eat The World

"The word is spreading in all corners of the tech industry that the biggest part of big data, the unstructured part, possesses learnable patterns that we now have the computing power and algorithmic leverage to discern…This change marks a true disruption, and there are fortunes to be made. There are also tremendous social consequences to consider that require as much creativity and investment as the more immediately lucrative deep learning startups that are popping up all over…"

SCIENCE

Scientists See Promise in Deep-Learning Programs

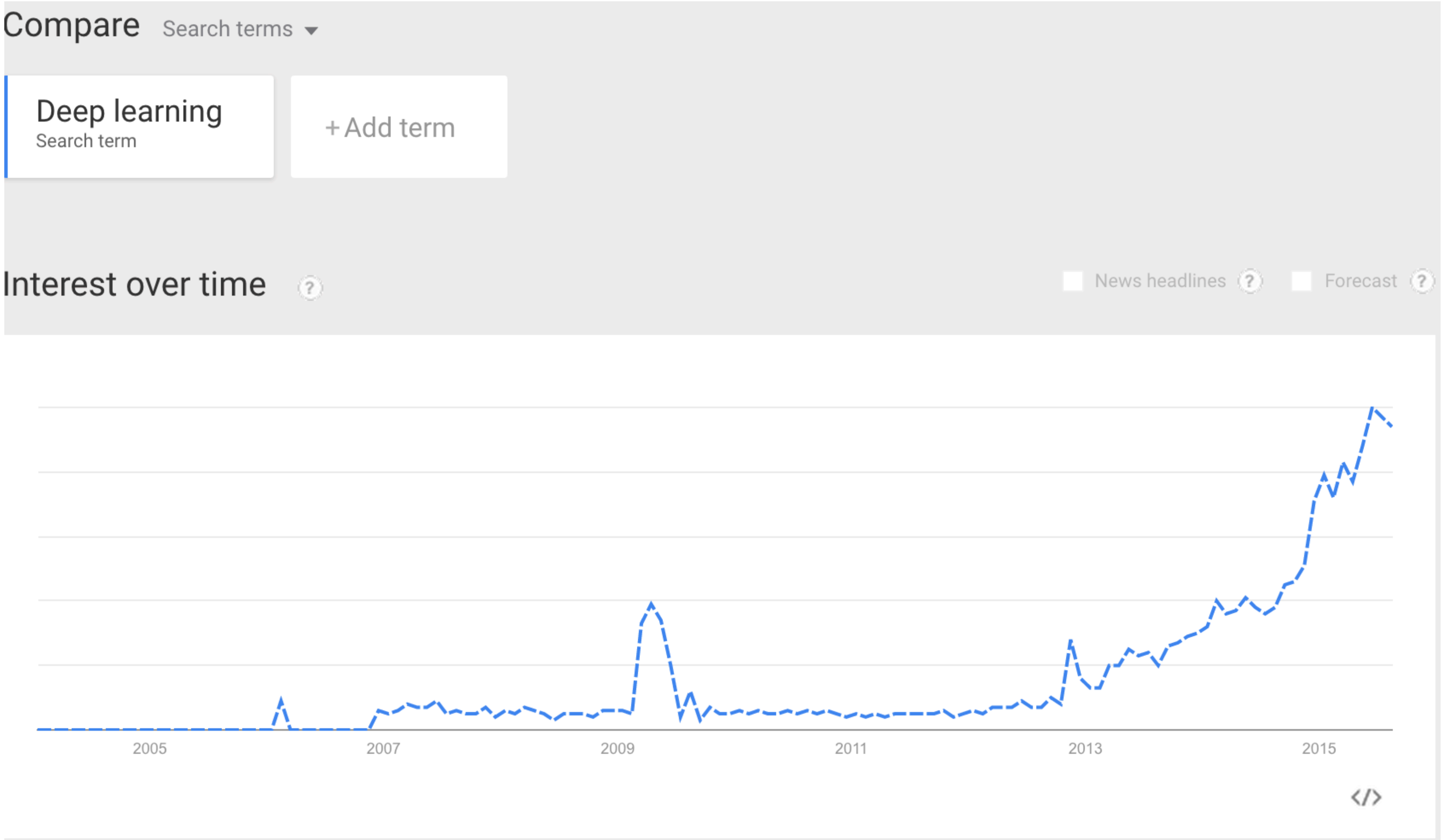
By JOHN MARKOFF NOV. 23, 2012

"Using an artificial intelligence technique inspired by theories about how the brain recognizes patterns, technology companies are reporting startling gains in fields as diverse as computer vision, speech recognition and the identification of promising new molecules for designing drugs.

The advances have led to widespread enthusiasm among researchers who design software to perform human activities like seeing, listening and thinking. They offer the promise of machines that converse with humans and perform tasks like driving cars and working in factories, raising the specter of automated robots that could replace human workers."

Hortonworks

Google Trends for "Deep Learning" keyword



Hortonworks

Enterprise use cases

General use case	Industry
Sound	
Voice recognition	UX/UI, Automotive, Security, IoT
Voice search	Handset maker, Telecoms
Sentiment analysis	CRM
Flaw detection (engine noise)	Automotive, Aviation
Fraud detection (latent audio artifacts)	Finance, Credit Cards
Time Series	
Log analysis/Risk detection	Data centers, Security, Finance
Enterprise resource planning	Manufacturing, Auto., Supply chain
Predictive analysis using sensor data	IoT, Smart home, Hardware manufact.
Business and Economic analytics	Finance, Accounting, Government
Recommendation engine	E-commerce, Media, Social Networks
Text	
Sentiment Analysis	CRM, Social media, Reputation mgt.
Augmented search, Theme detection	Finance
Threat detection	Social media, Govt.
Fraud detection	Insurance, Finance
Image	
Facial recognition	
Image search	Social media
Machine vision	Automotive, aviation
Photo clustering	Telecom, Handset makers
Video	
Motion detection	Gaming, UX, UI
Real-time threat detection	Security, Airports

Hortonworks

Deep Learning

- One of the many pattern recognition techniques in Data Science
- Excels at rich media applications:
  - Image recognition
  - Speech translation
  - Voice recognition
- Loosely inspired by human brain models
- Synonymous with Artificial Neural Networks, Multi Layer Networks

In this workshop

- Fundamentals of Deep Learning
- Implementation and Libraries in Real Life
- Demo!

Hortonworks

So, 1. **what exactly is deep learning** ?

And, 2. **why is it generally better** than other methods on image, speech and certain other types of data?

So, 1. **what exactly is deep learning** ?

And, 2. **why is it generally better** than other methods on image, speech and certain other types of data?

The short answers

1. 'Deep Learning' means using a neural network
   with several layers of nodes between input and output

2. the series of layers between input & output do
feature identification and processing in a series of stages,
just as our brains seem to.

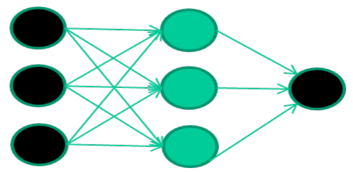
but:

3. multilayer neural networks have been around for 25 years. What's actually new?

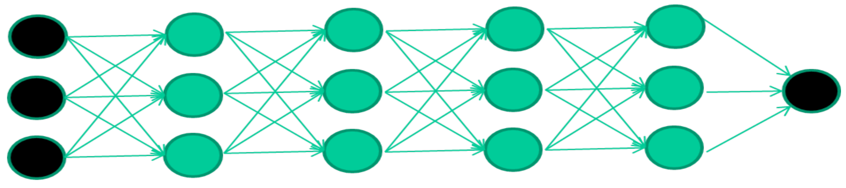
but:

3. multilayer neural networks have been around for 25 years. What's actually new?

we have always had good algorithms for learning the weights in networks with 1 hidden layer



but these algorithms are not good at learning the weights for networks with more hidden layers



what's new is: **<u>algorithms for training many-layer networks</u>**

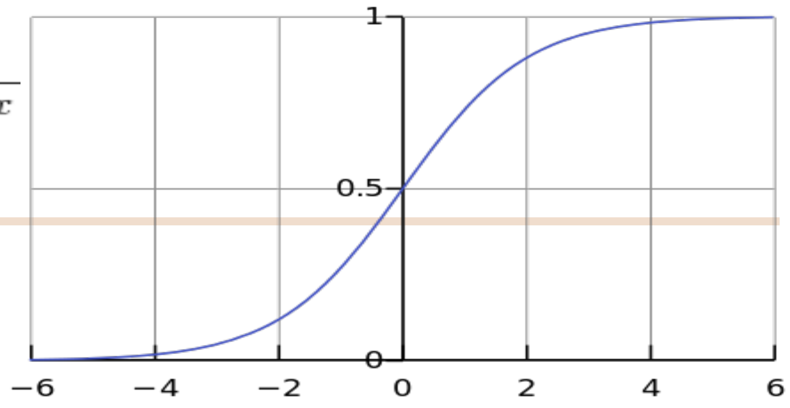
Hortonworks

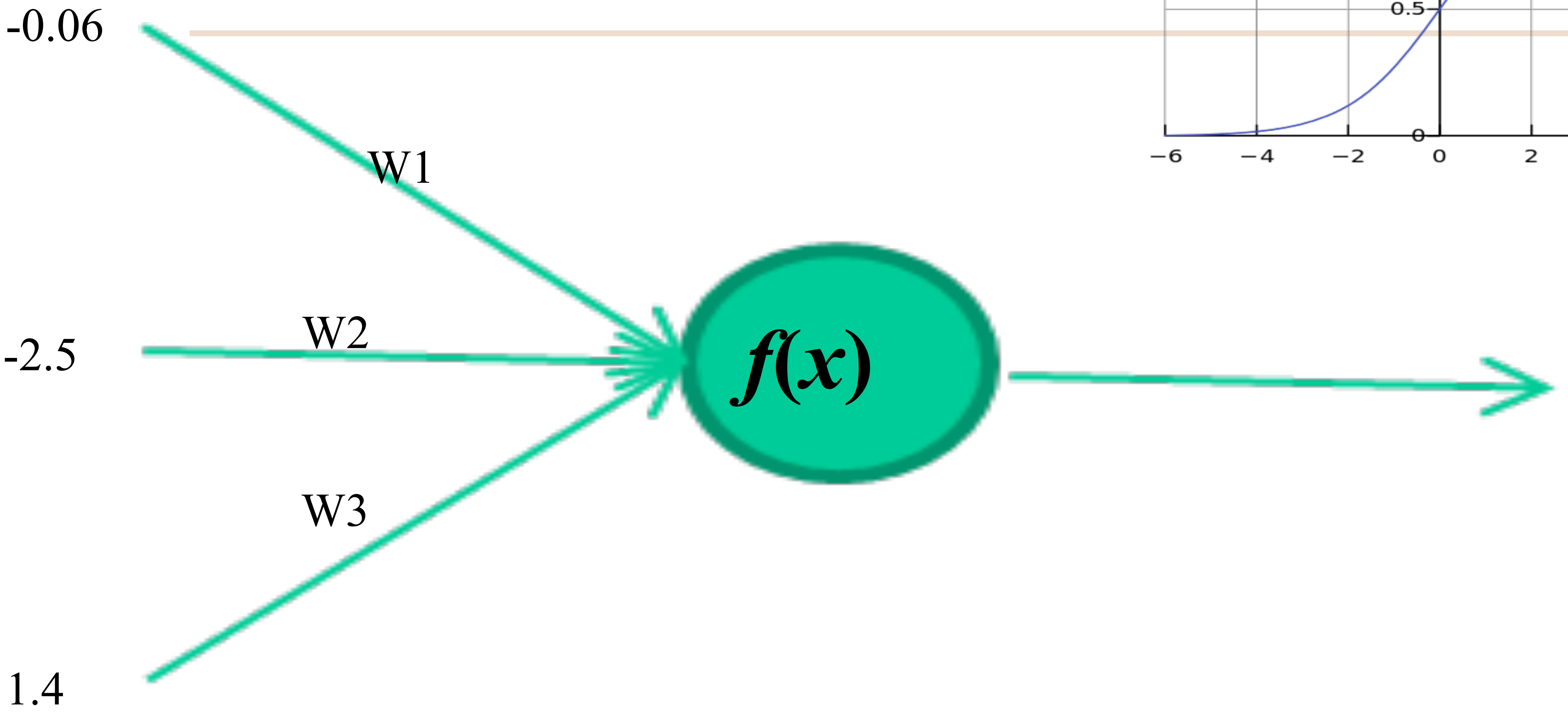
Longer answers

1. reminder/quick-explanation of how neural network weights are learned;
2. the idea of **unsupervised feature learning** (why ‘intermediate features’ are important for difficult classification tasks, and how NNs seem to naturally learn them)
3. The ‘breakthrough’ – the simple trick for training Deep neural networks

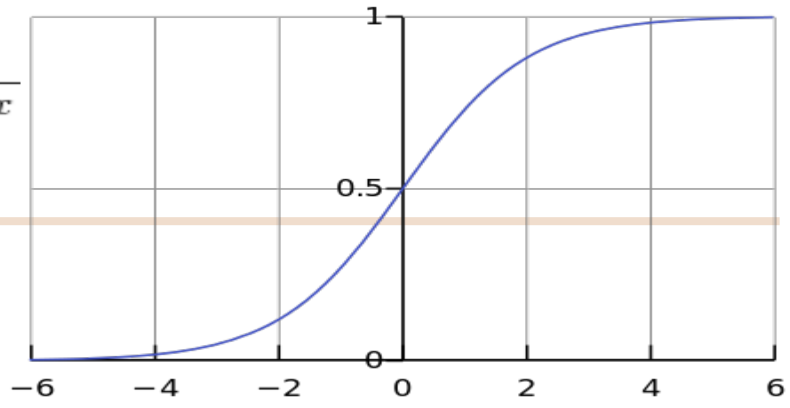
Neuron Quick Look

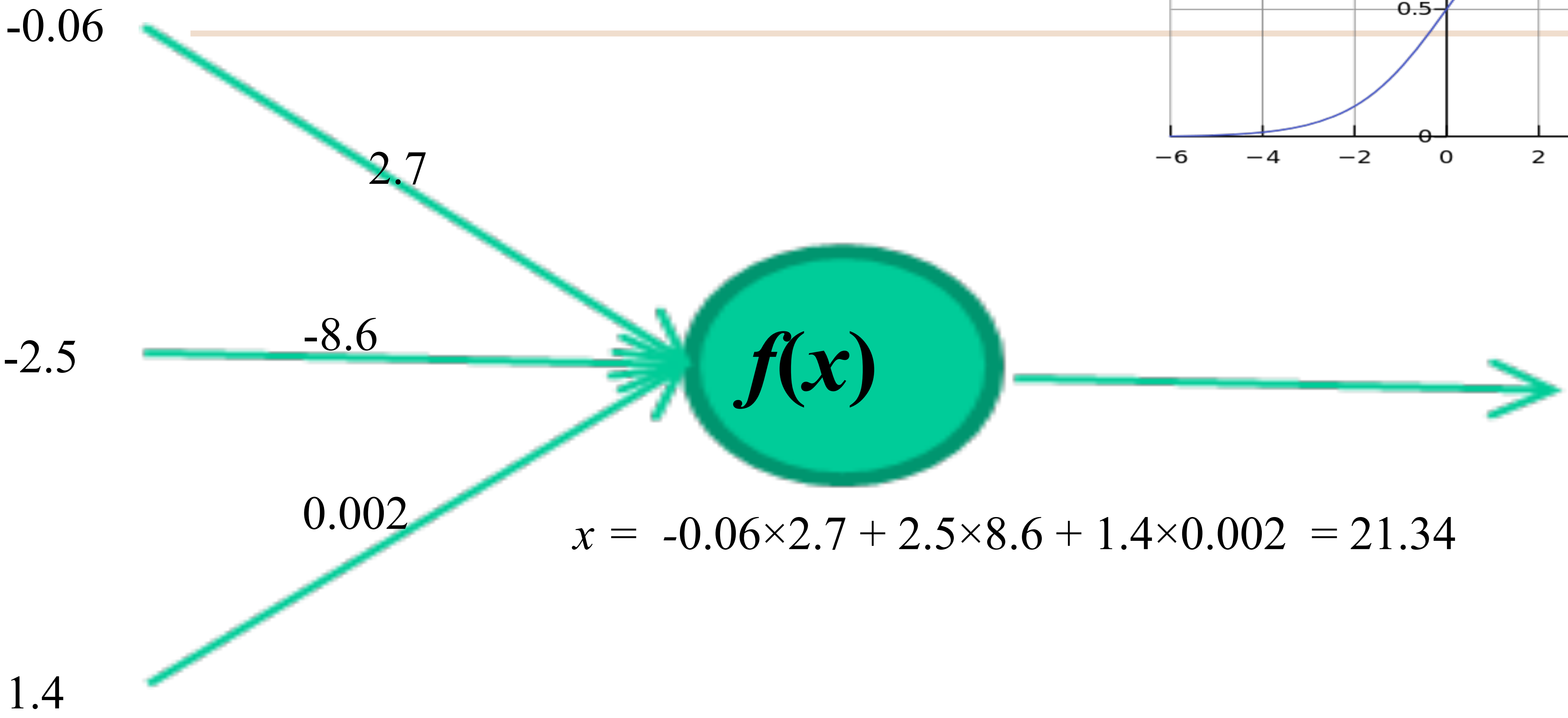
$$f(x) = \frac{1}{1 + e^{-x}}$$





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





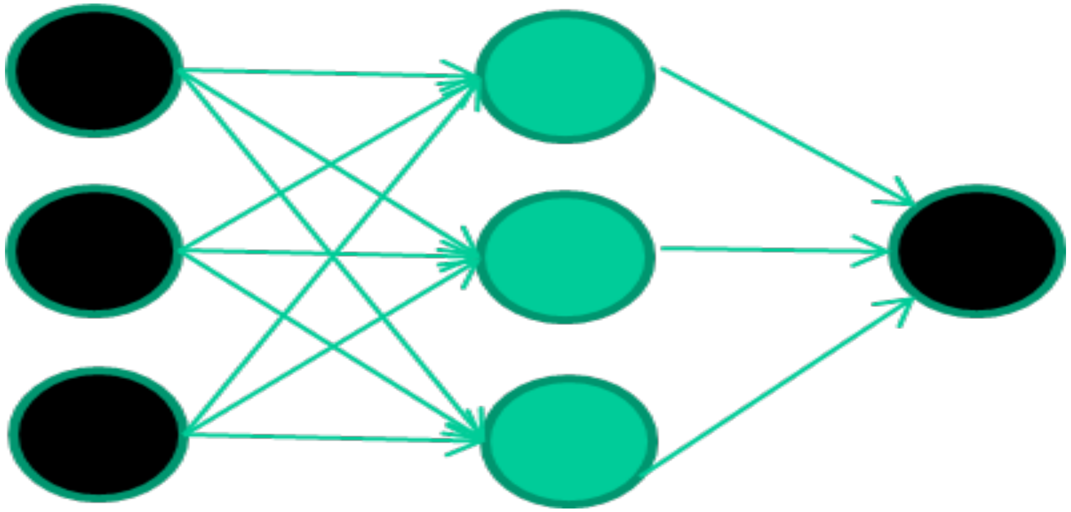
$$x = -0.06\times2.7 + 2.5\times8.6 + 1.4\times0.002 = 21.34$$

Hortonworks

A dataset

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0

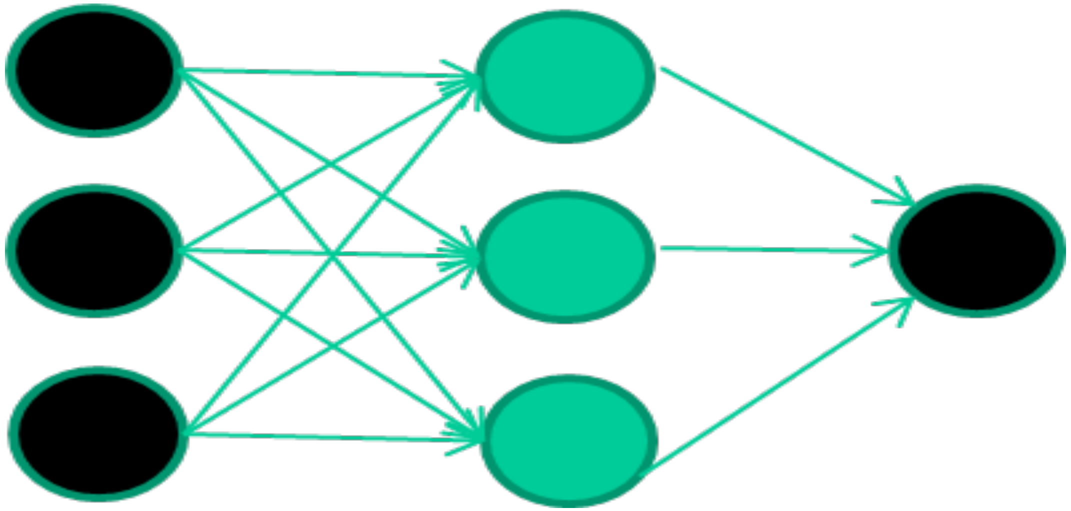
etc …



Training the neural network

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0

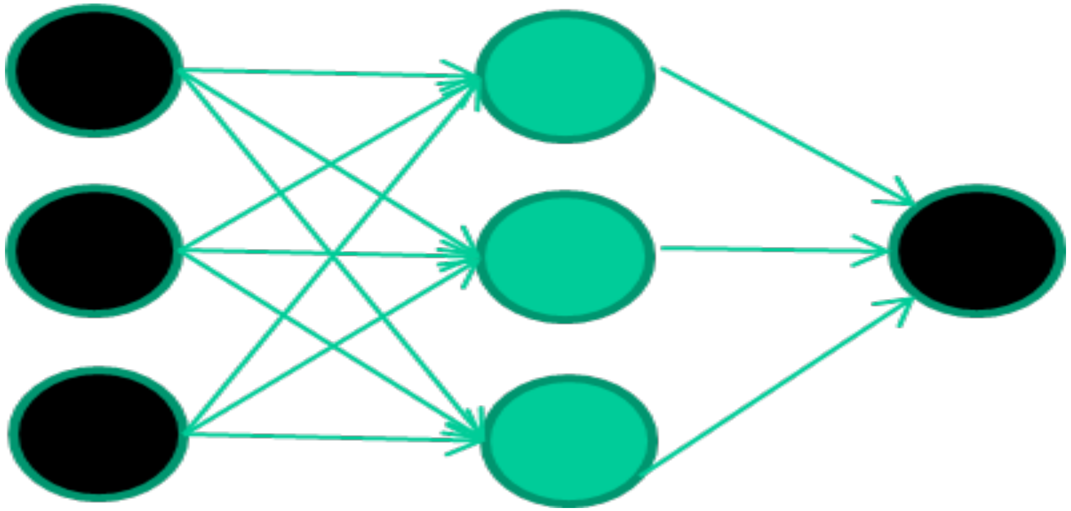
etc …



Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

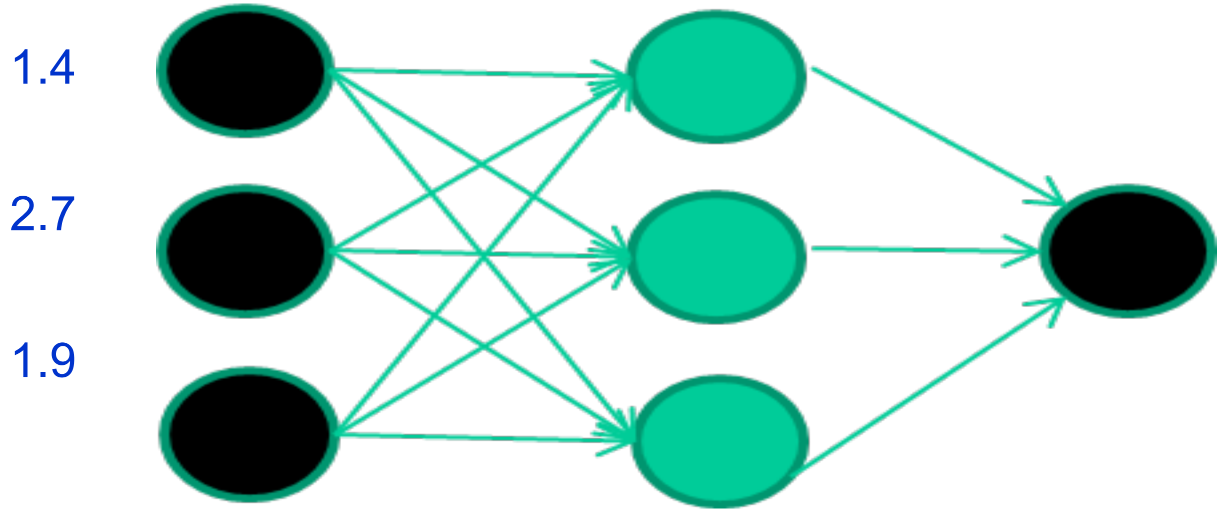
Initialise with random weights



Present a training pattern

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

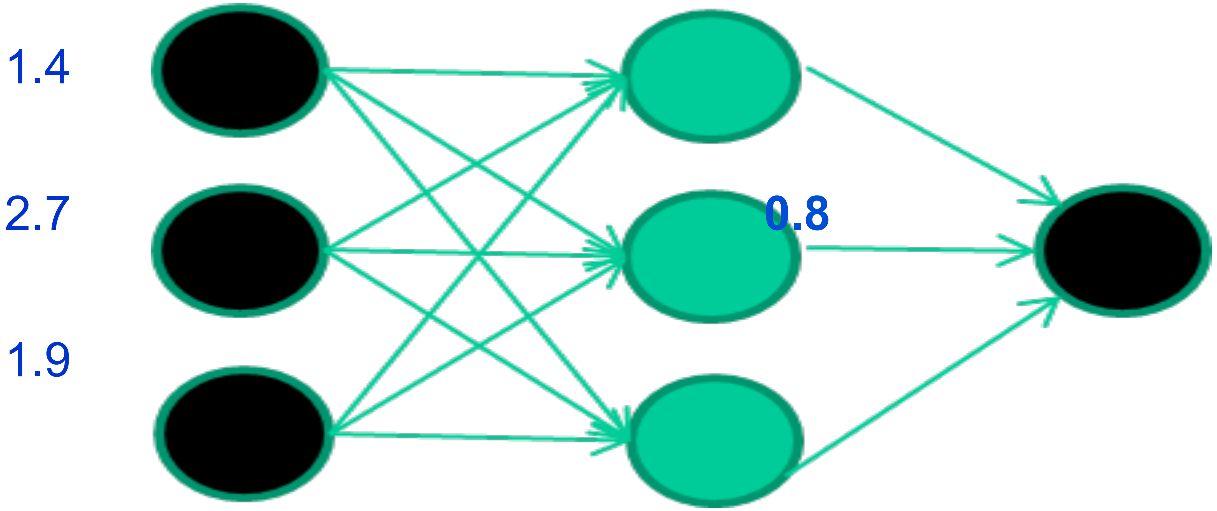


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

Feed it through to get output

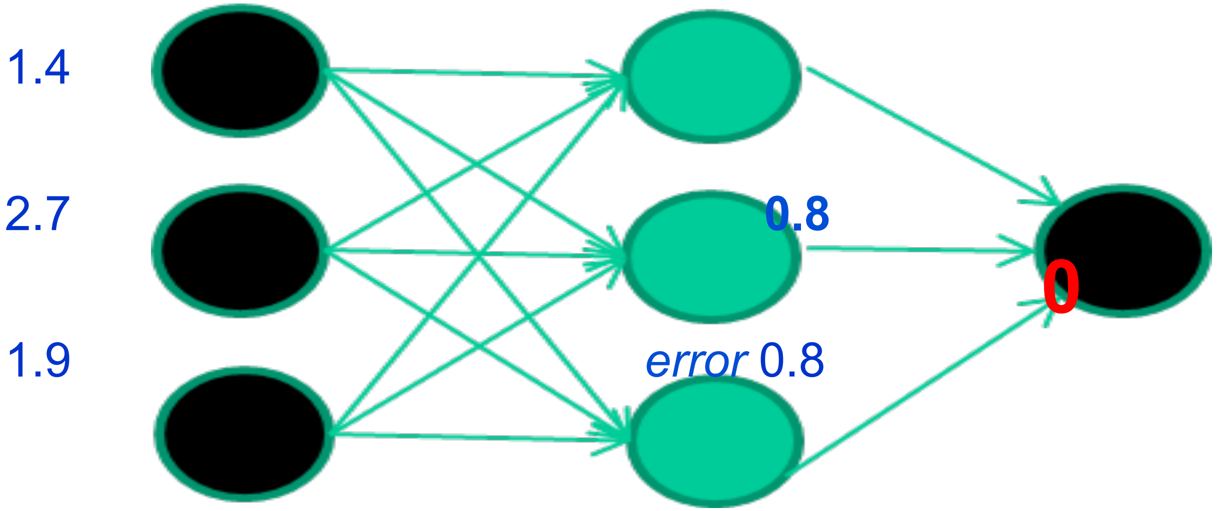


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

Compare with target output

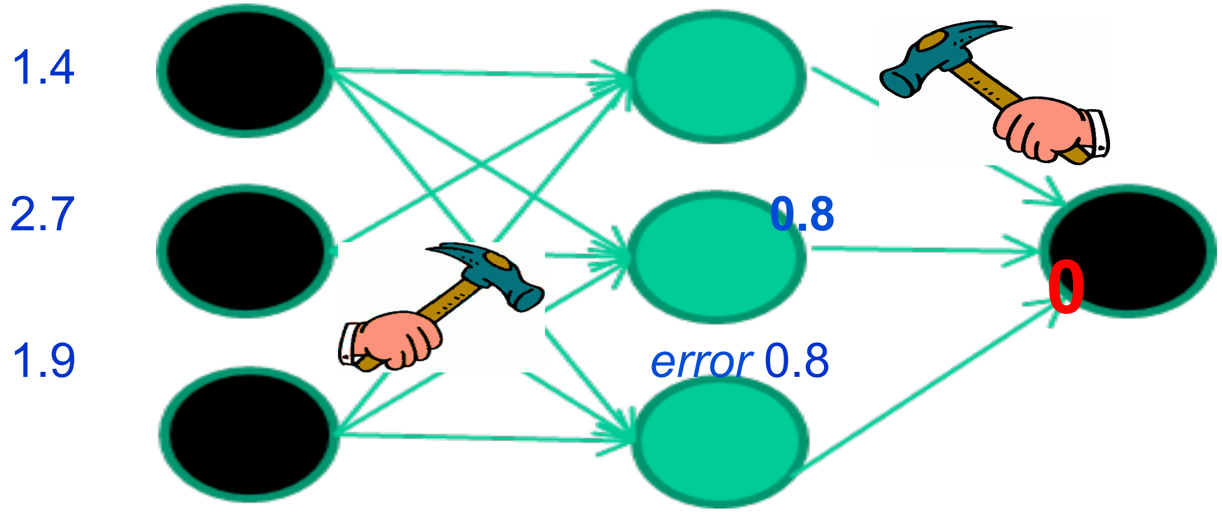


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

Adjust weights based on error

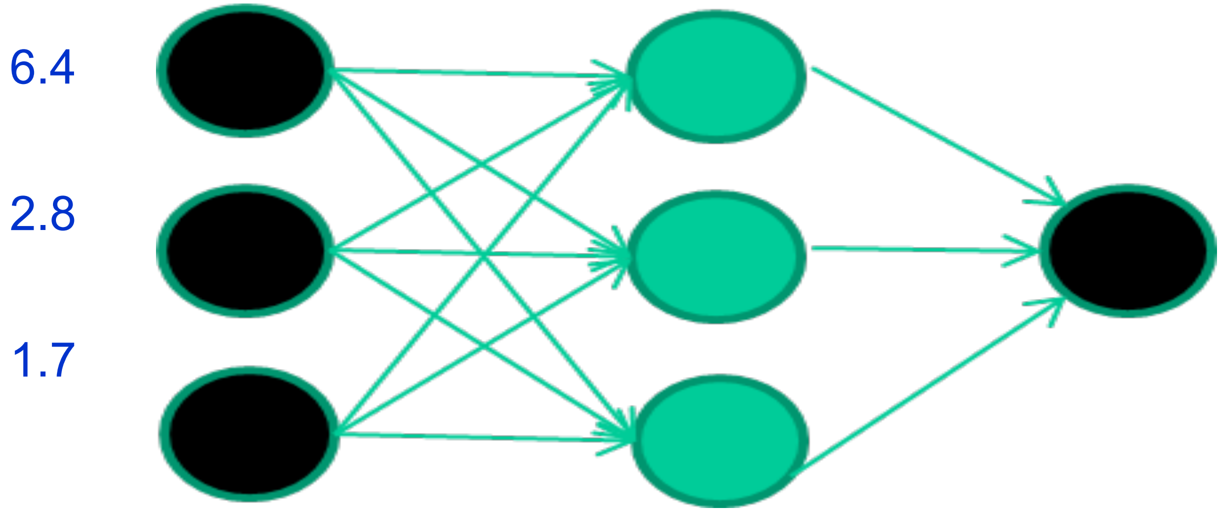


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

Present a training pattern

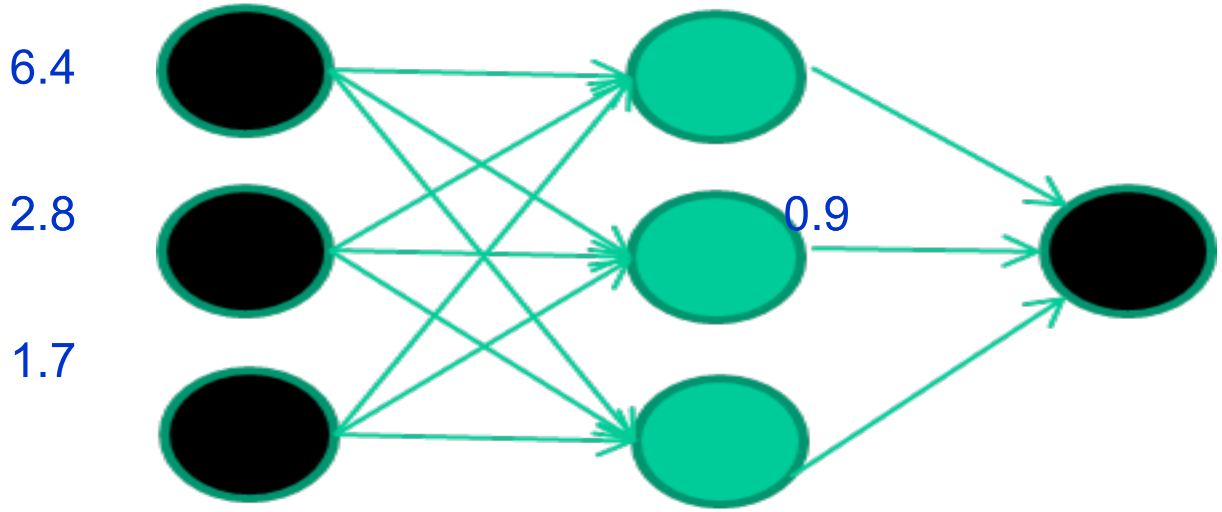


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

Feed it through to get output

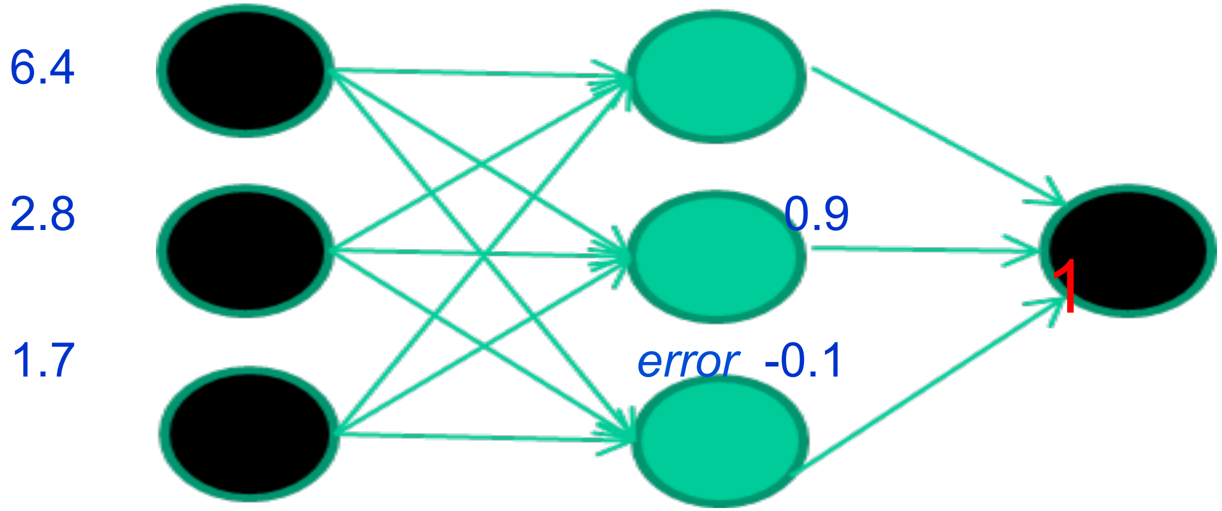


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

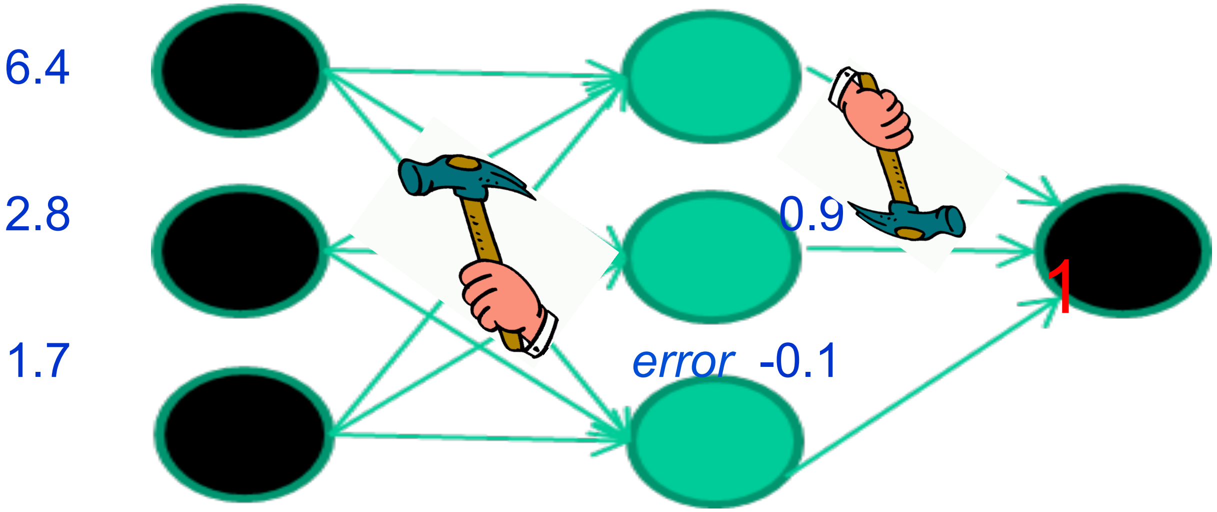
Compare with target output



Adjust weights based on error

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

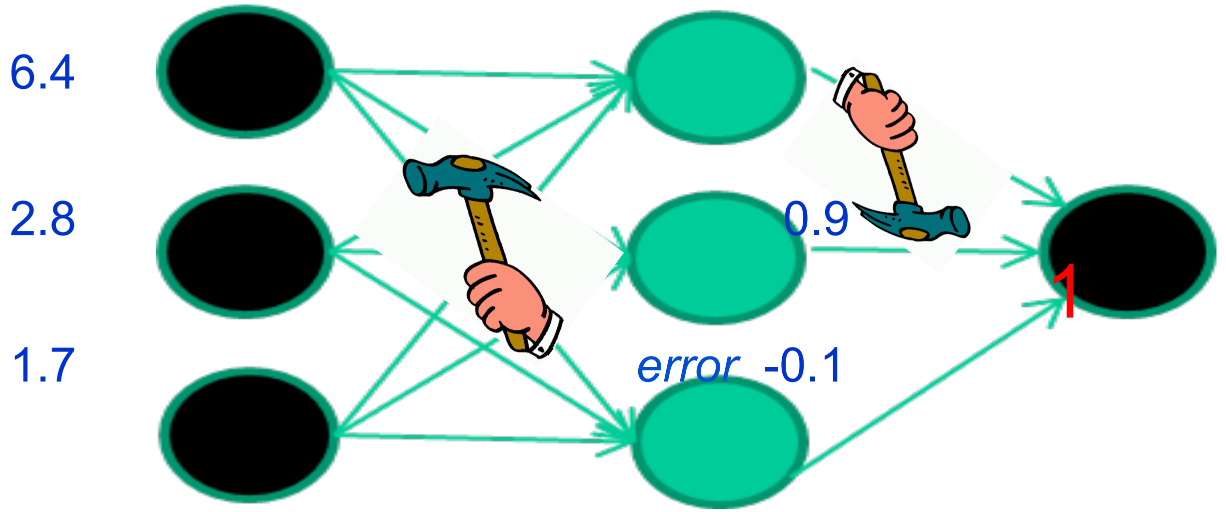


Hortonworks

Training data

Fields			***class***
1.4	2.7	1.9	0
3.8	3.4	3.2	0
6.4	2.8	1.7	1
4.1	0.1	0.2	0
etc …			

And so on ….



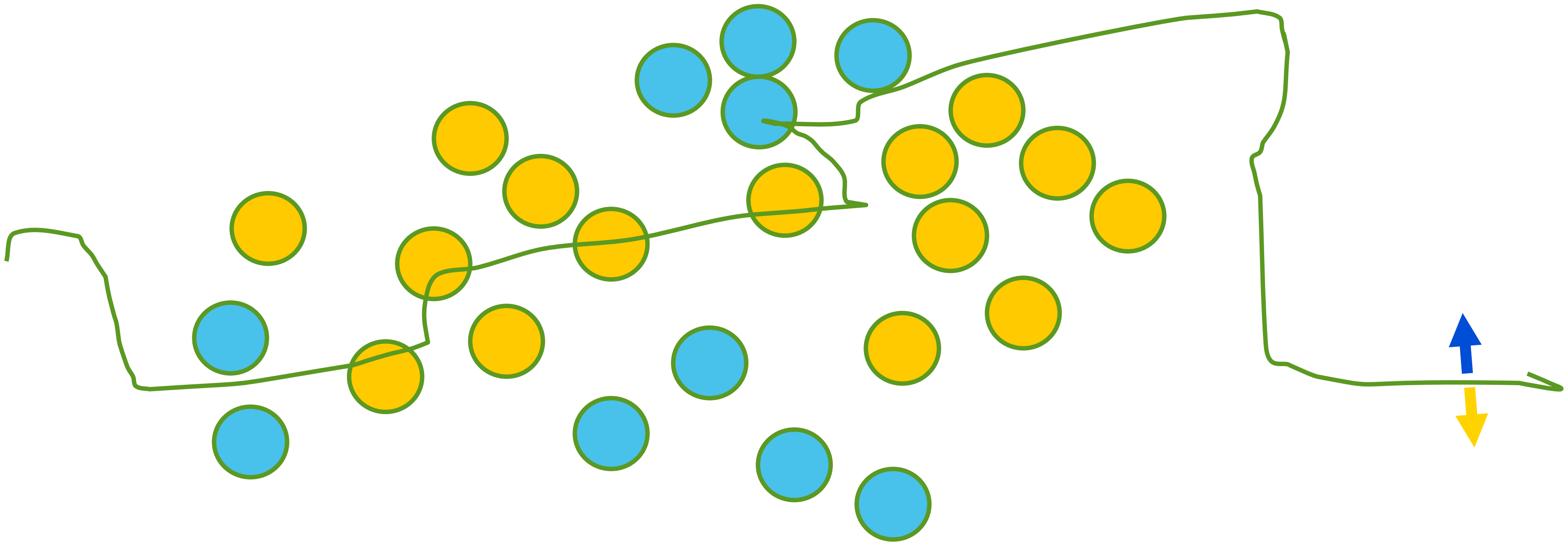
Repeat this thousands, maybe millions of times – each time taking a random training instance, and making slight weight adjustments

Algorithms for weight adjustment are designed to make changes that will reduce the error

ortonworks

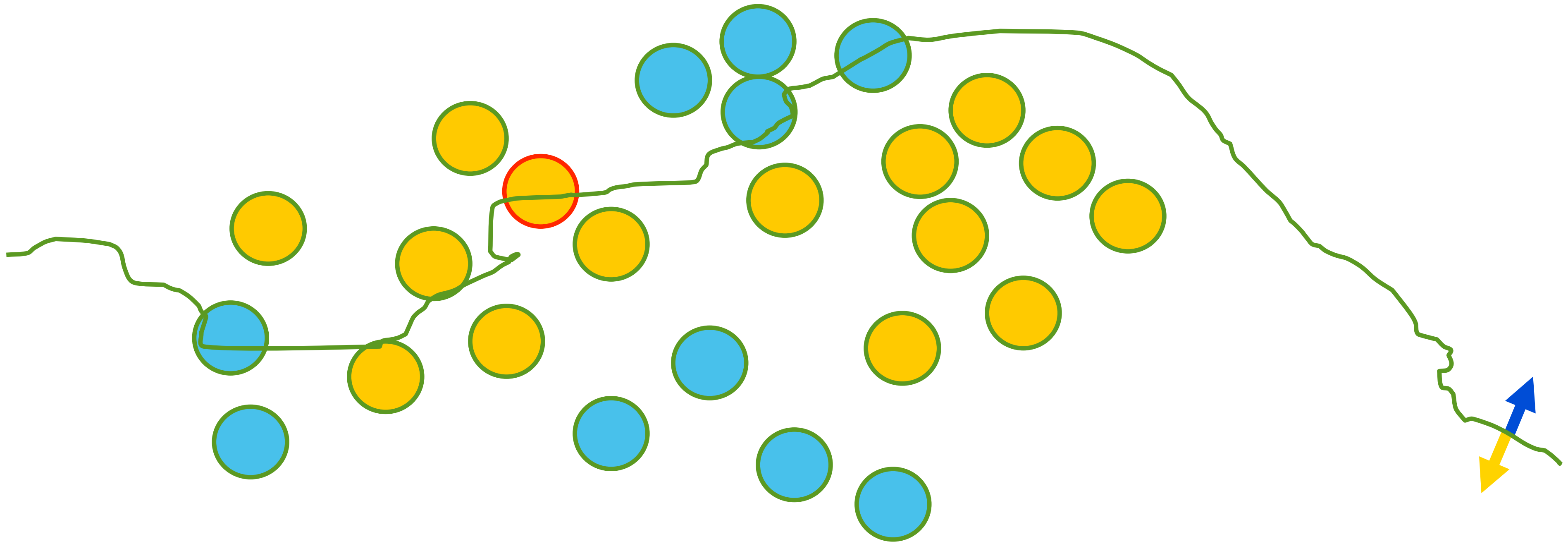
The decision boundary perspective…

Initial random weights



The decision boundary perspective…

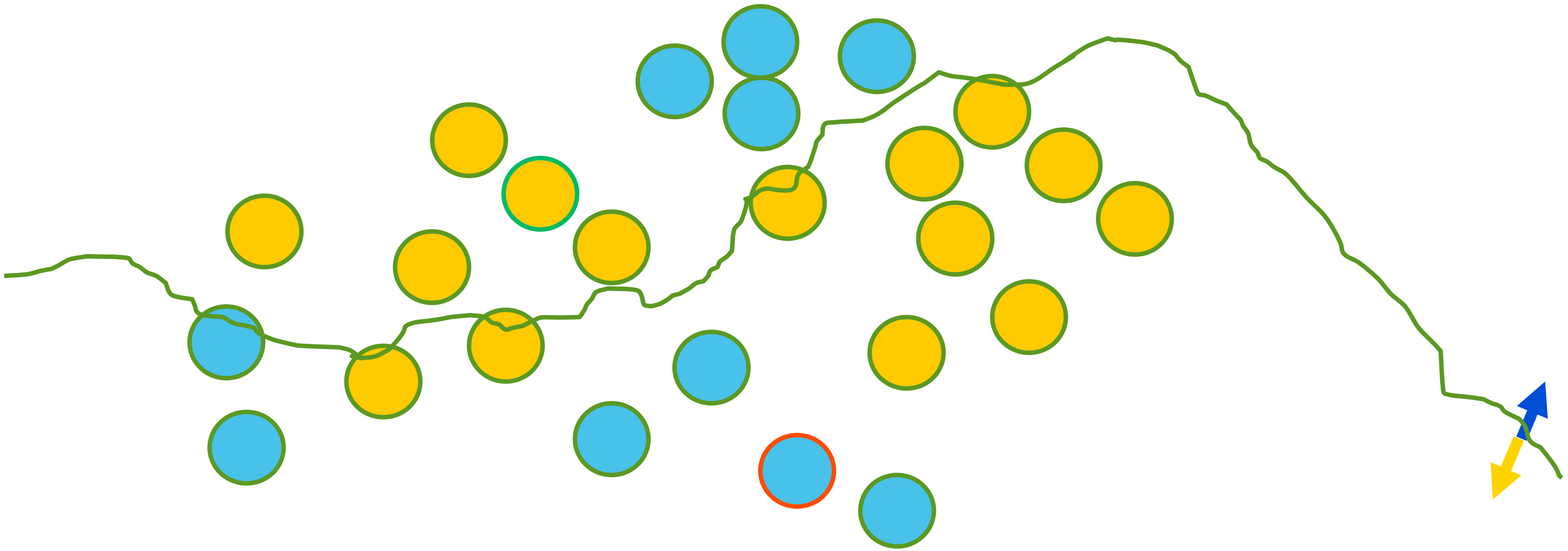
Present a training instance / adjust the weights



Hortonworks

The decision boundary perspective…

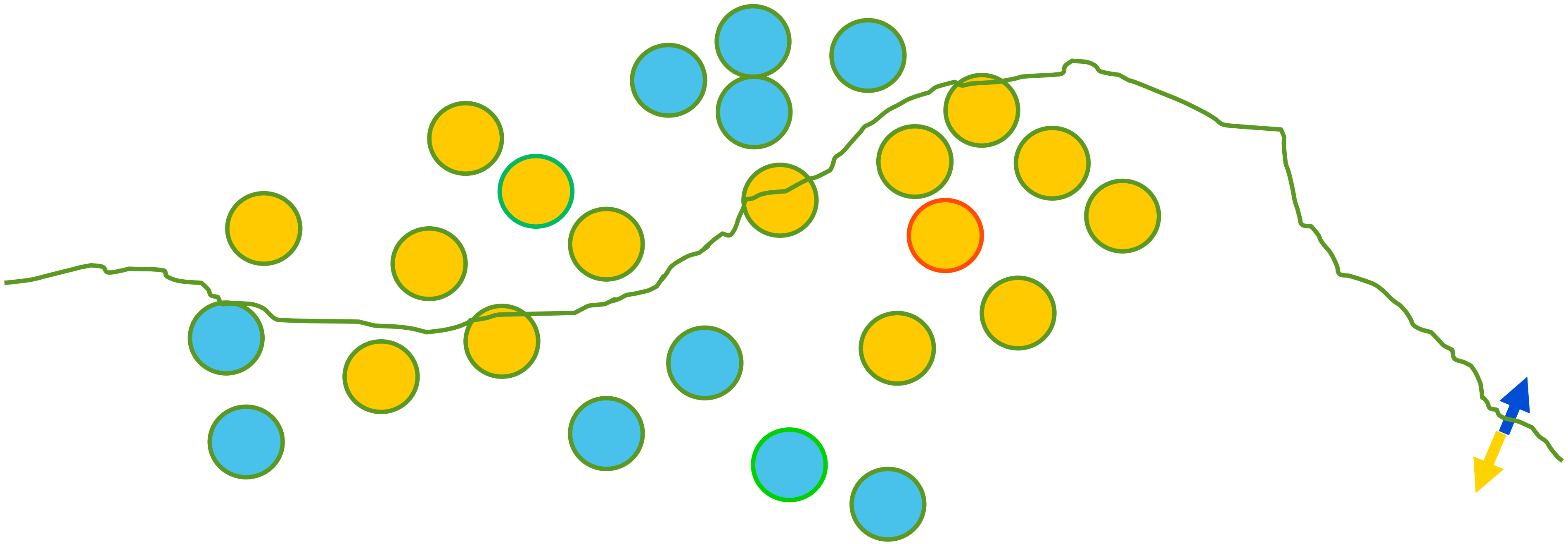
Present a training instance / adjust the weights



Hortonworks

The decision boundary perspective…

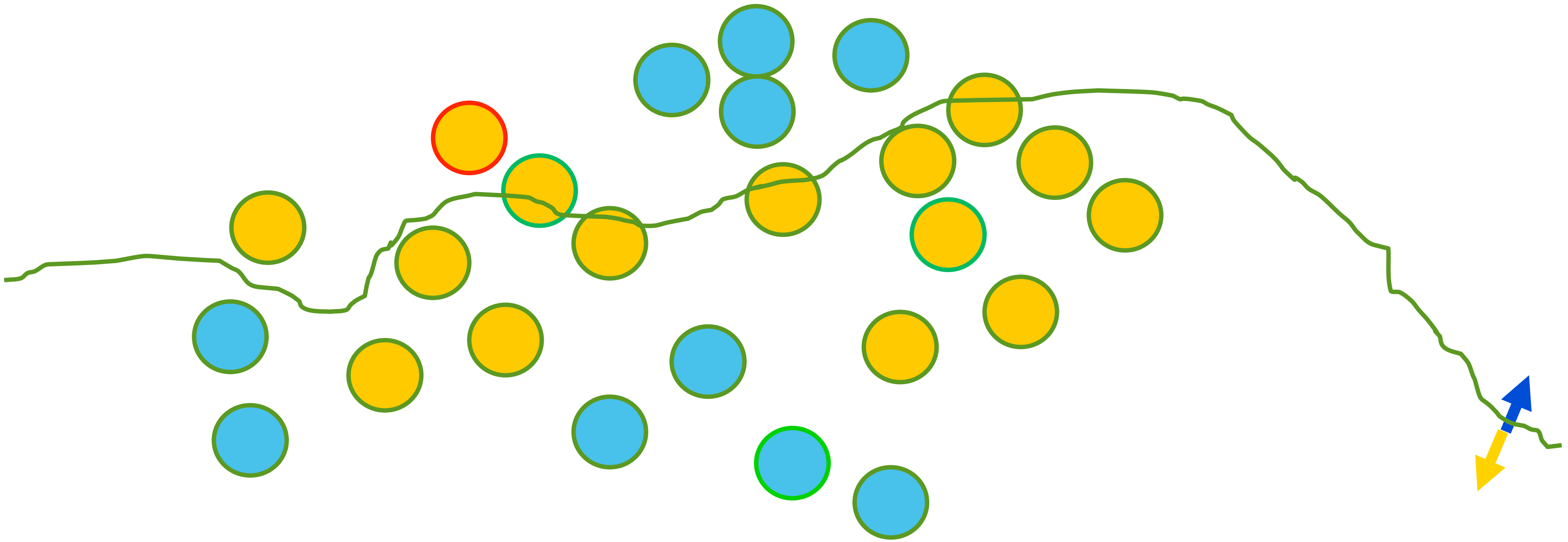
Present a training instance / adjust the weights



Hortonworks

The decision boundary perspective…

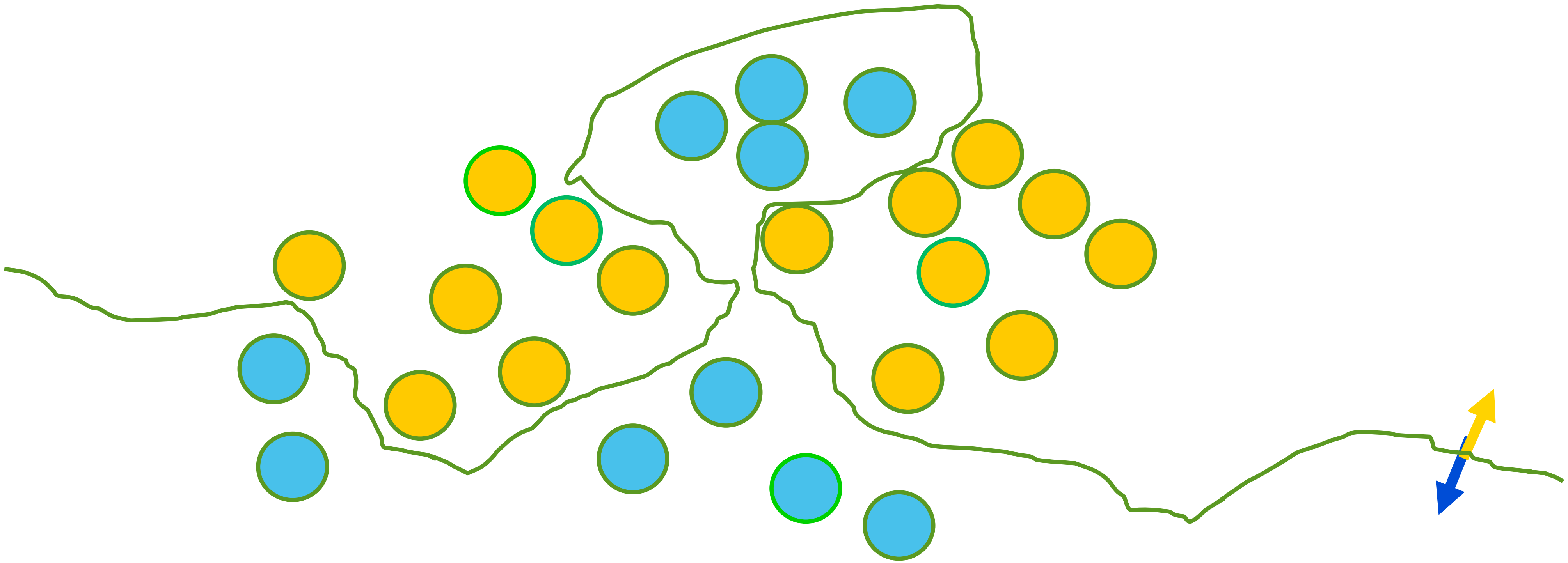
Present a training instance / adjust the weights



Hortonworks

The decision boundary perspective…

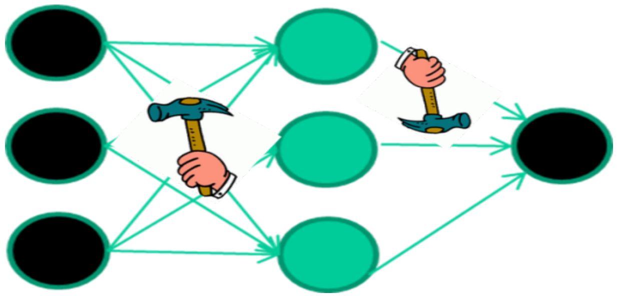
Eventually ….



Hortonworks

In essence

- weight-learning algorithms for NNs are dumb
- they work by making thousands and thousands of tiny adjustments, each making the network do better at the most recent pattern, but perhaps a little worse on many others
- but, by dumb luck, eventually this tends to be good enough to learn effective classifiers for many real applications



Hortonworks

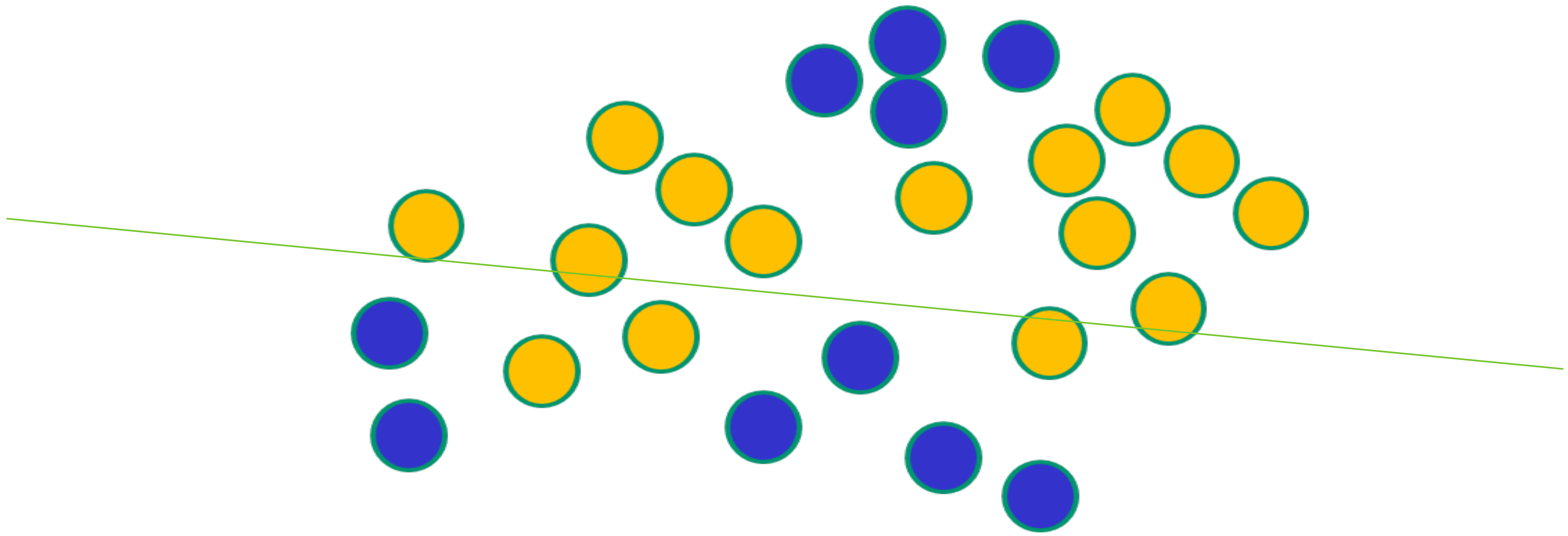
Some other points

Detail of a standard NN weight learning algorithm – **later**

If $f(x)$ is non-linear, a network with 1 hidden layer can, in theory, learn perfectly any classification problem. A set of weights exists that can produce the targets from the inputs. The problem is finding them.

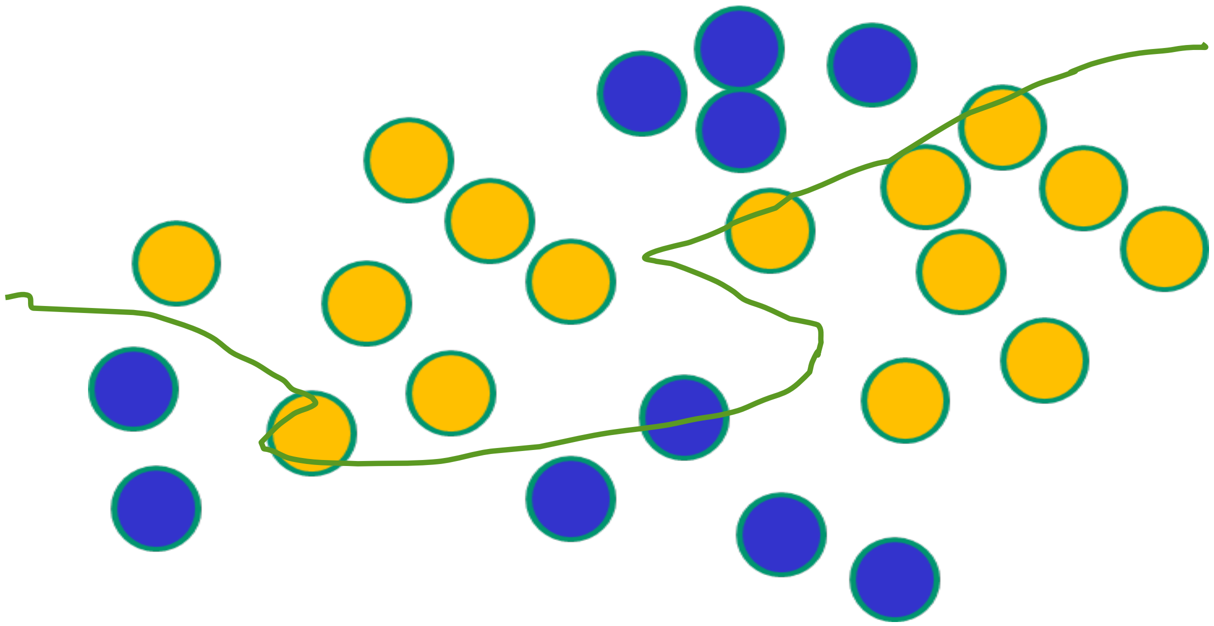
Some other ‘by the way’ points

If $f(x)$ is linear, the NN can **only** draw straight decision boundaries (even if there are many layers of units)



Some other 'by the way' points

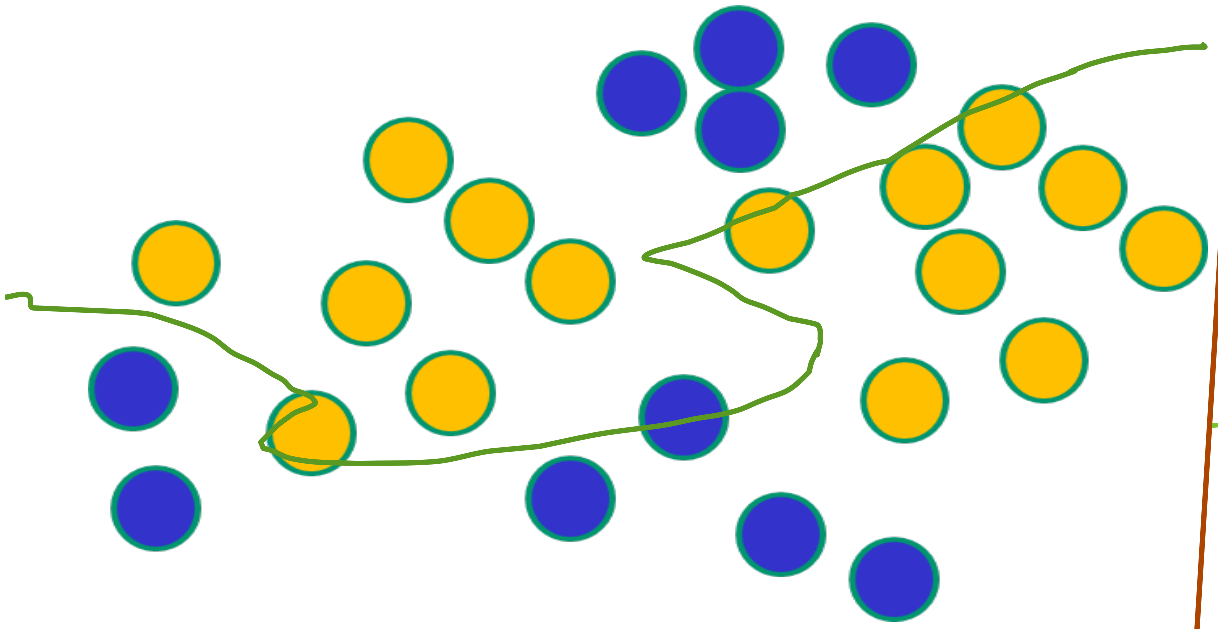
NNs use nonlinear *f*(x) so they can draw complex boundaries, but keep the data unchanged



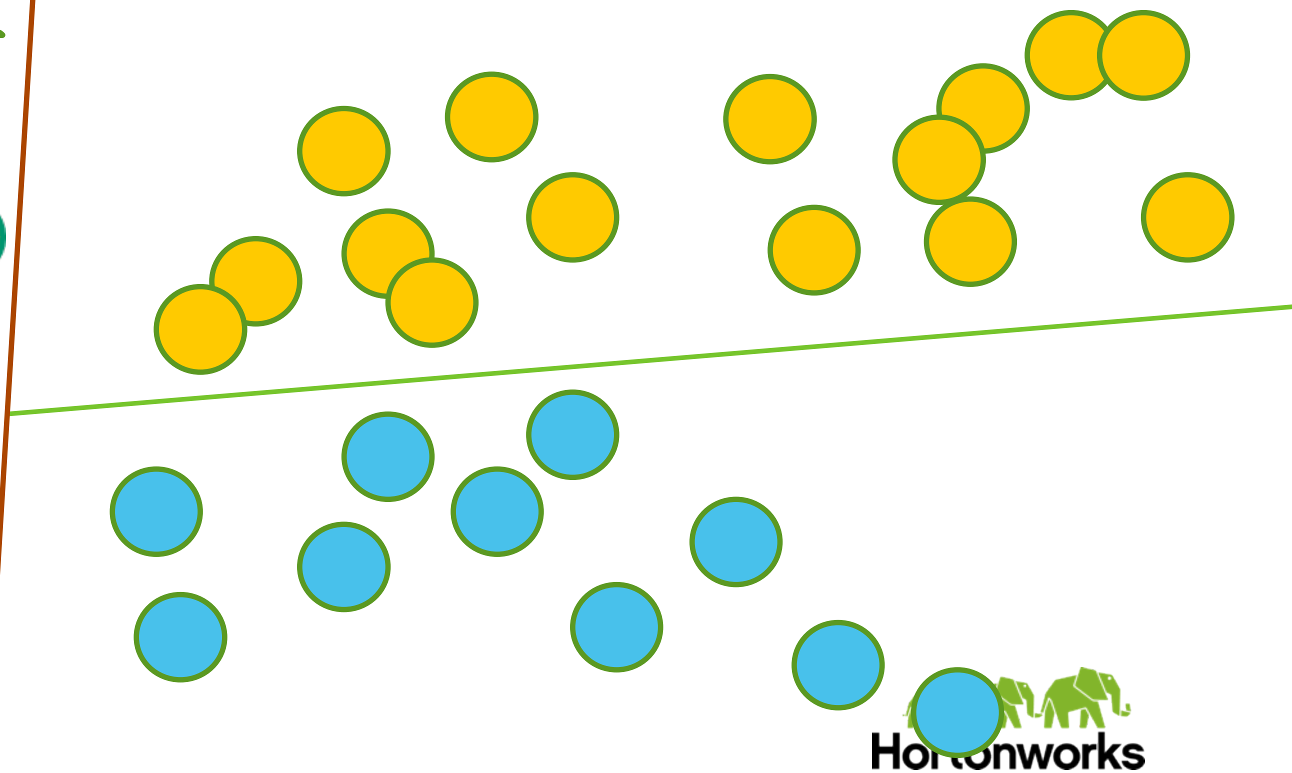
Hortonworks

Some other 'by the way' points

NNs use nonlinear *f*(x) so they can draw complex boundaries, but keep the data unchanged



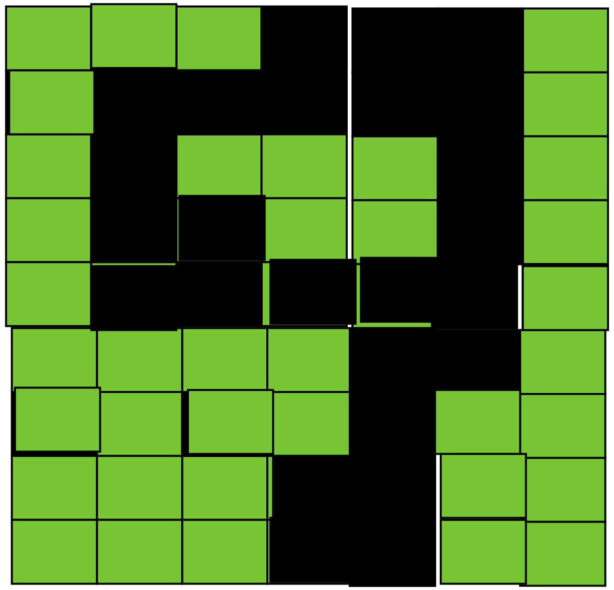
SVMs only draw straight lines, but they transform the data first in a way that makes that OK

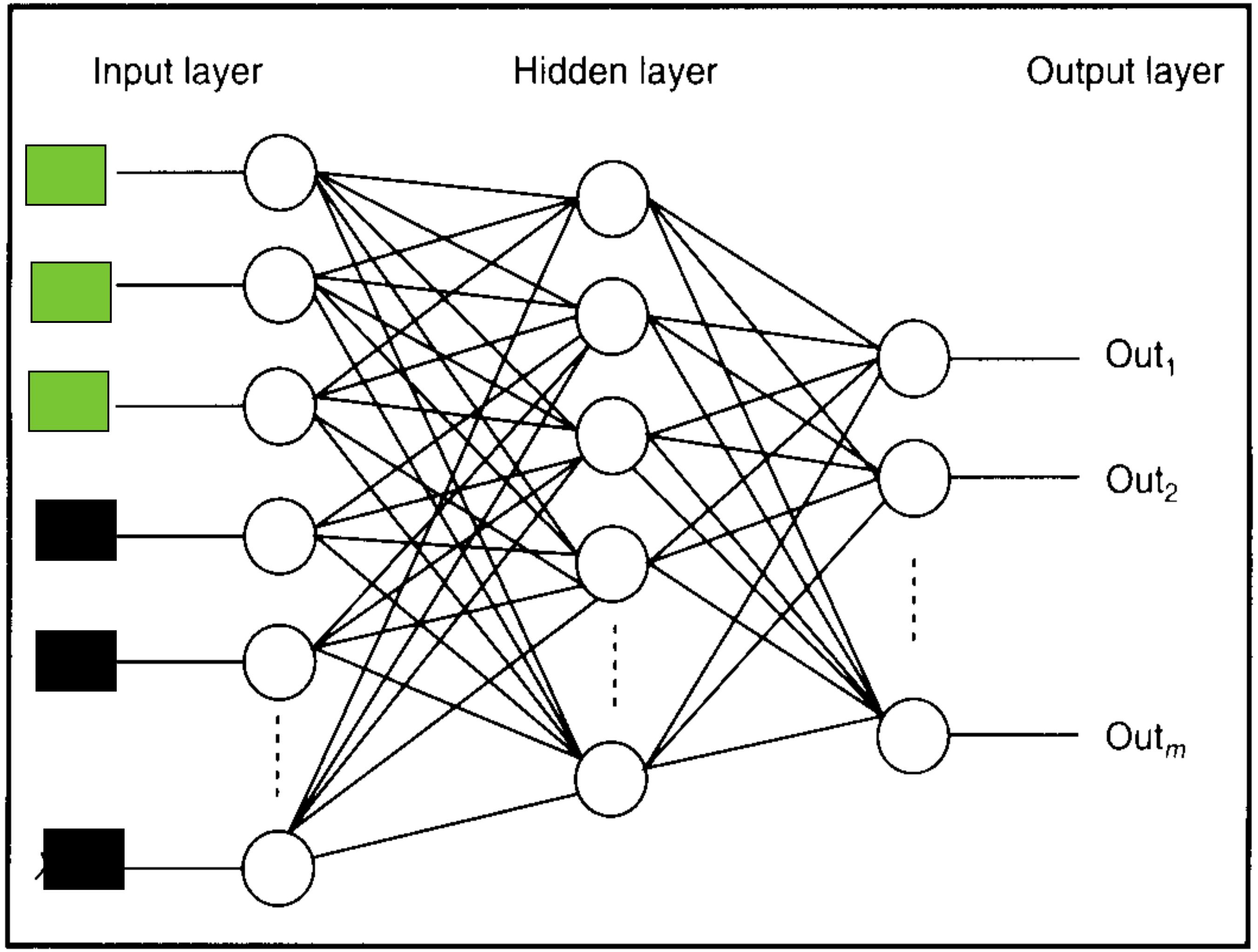


Hortonworks

Feature detectors

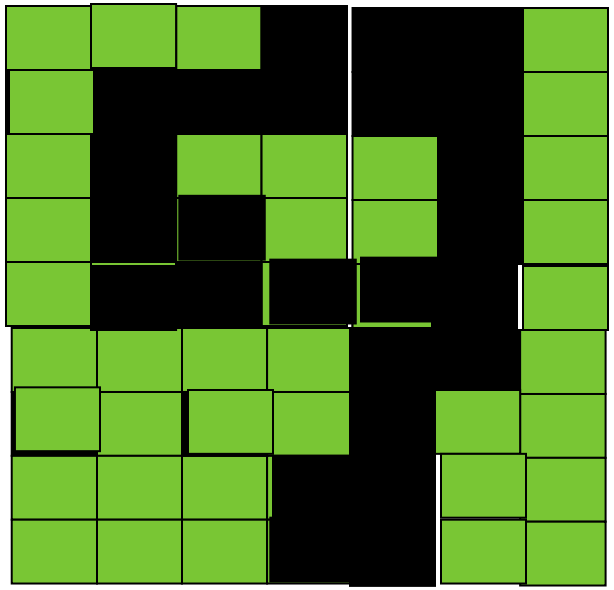
Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

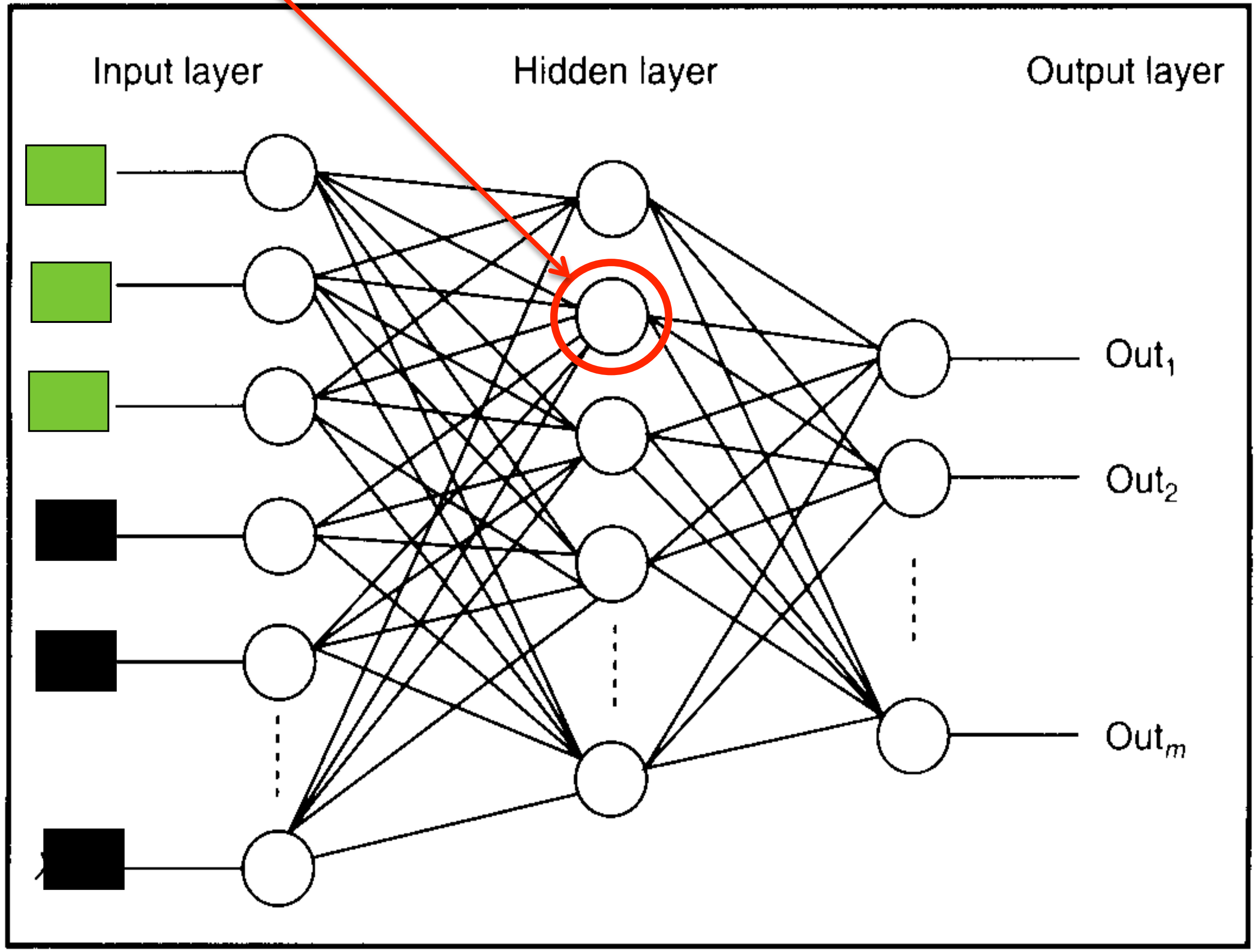




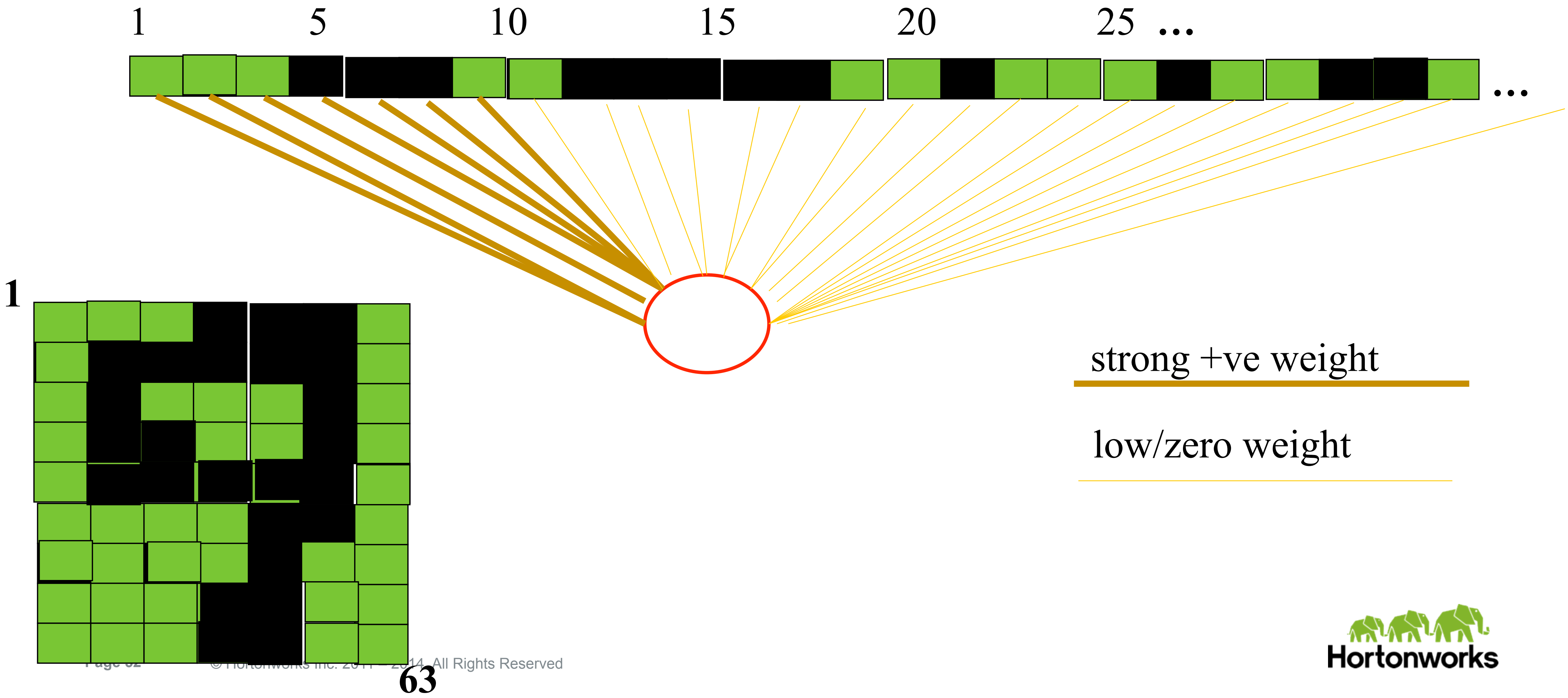
What’s this unit doing?

Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*



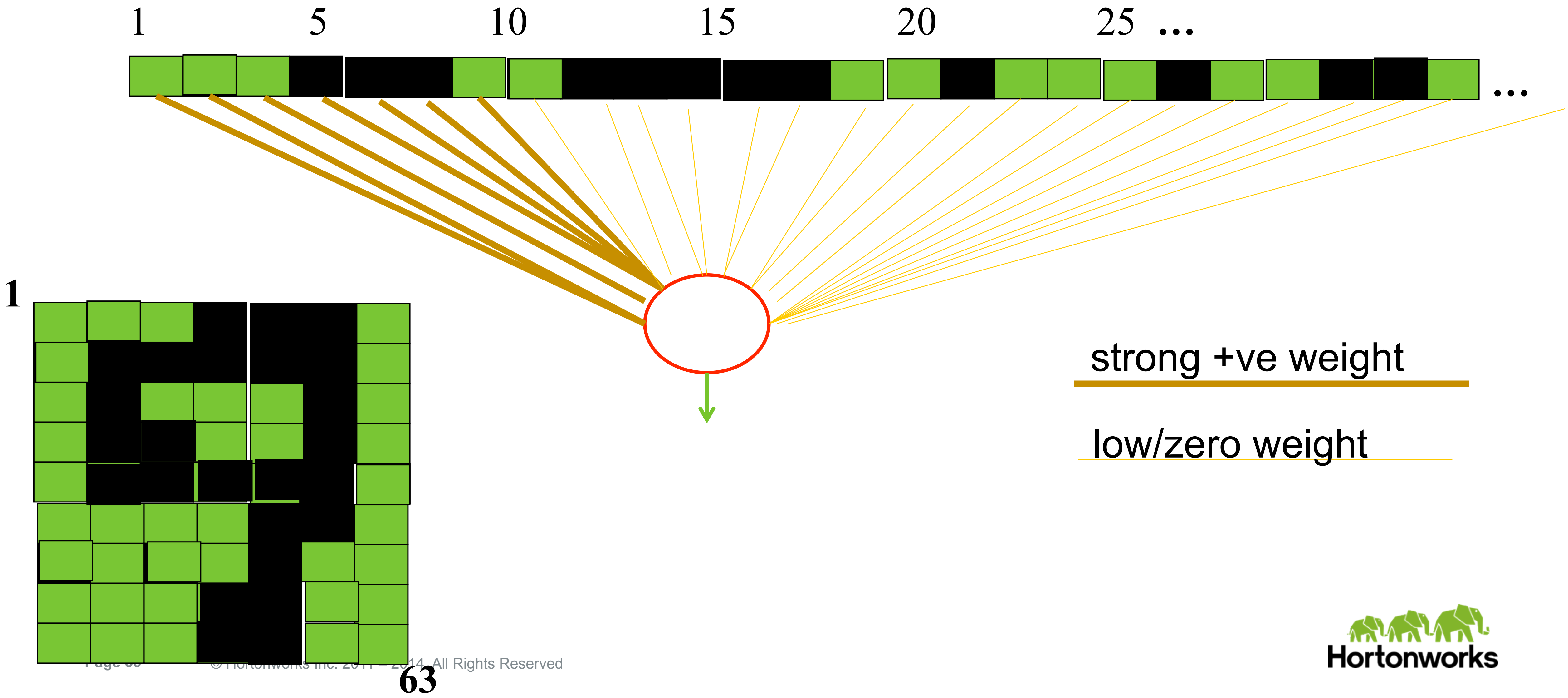


Hidden layer units become *self-organised feature detectors*



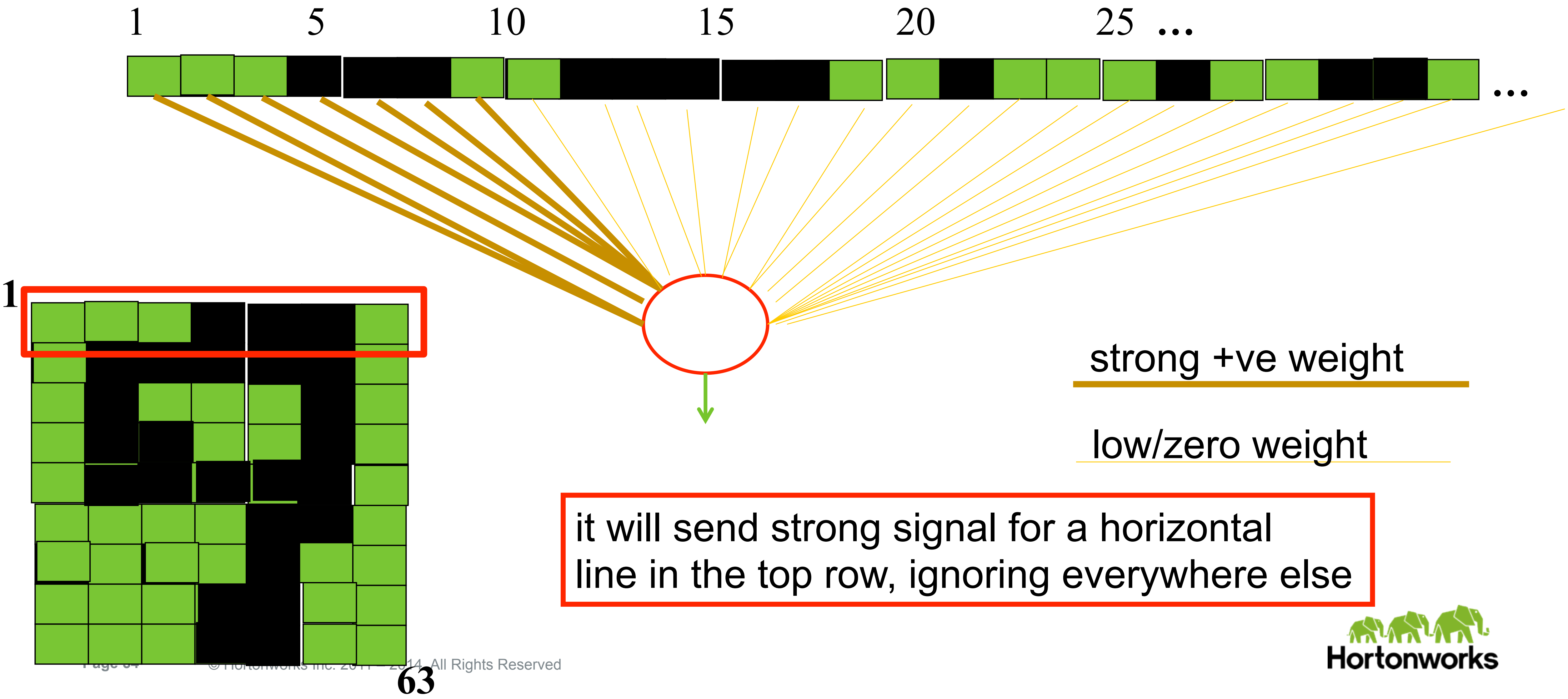
Hortonworks

What does this unit detect?



Hortonworks

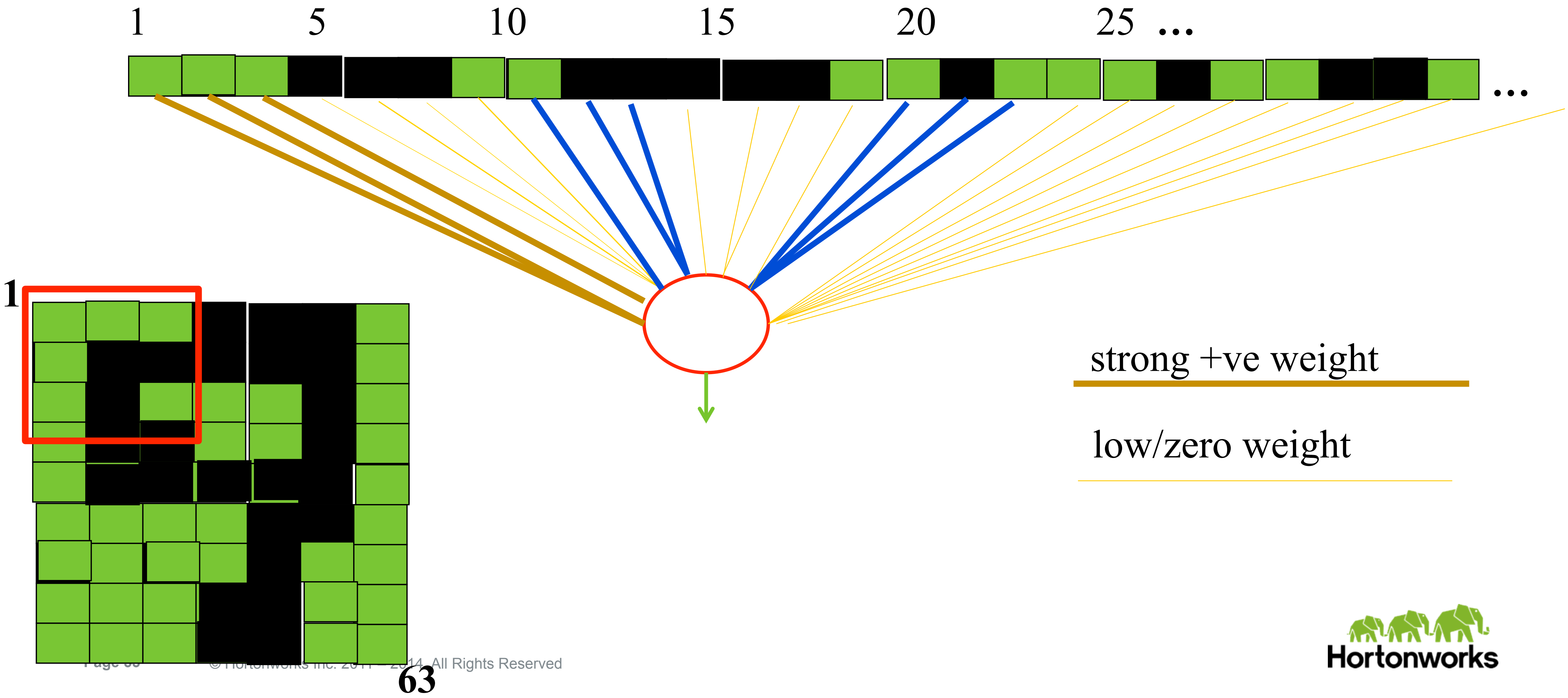
What does this unit detect?



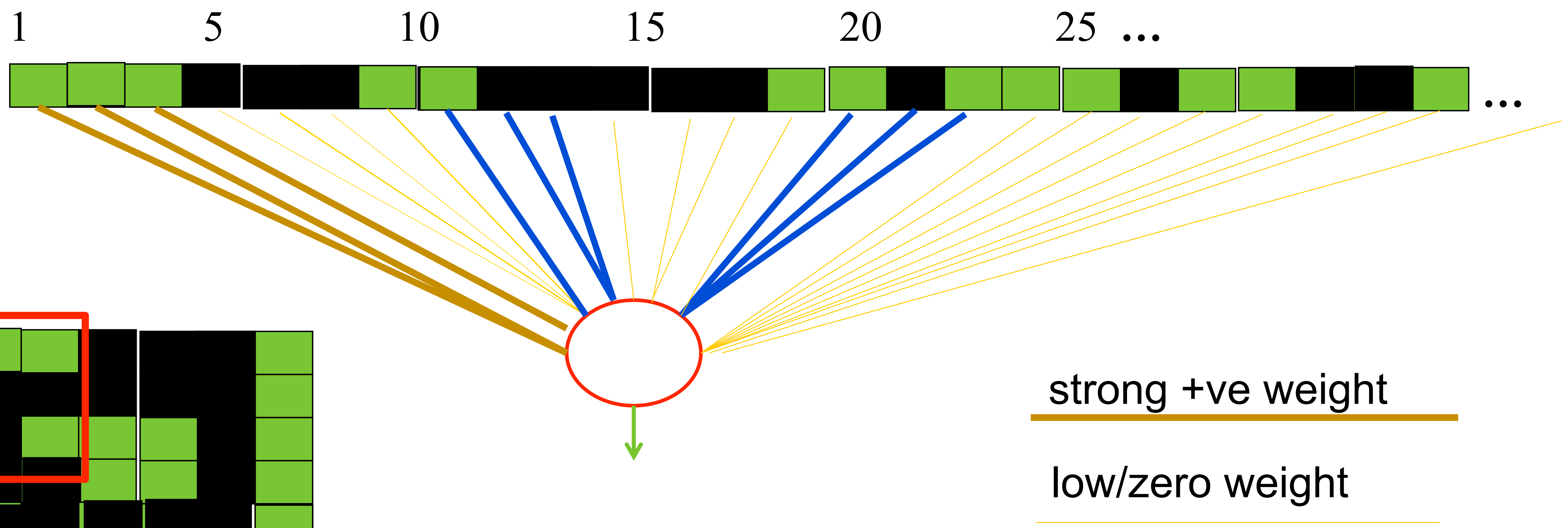
it will send strong signal for a horizontal line in the top row, ignoring everywhere else

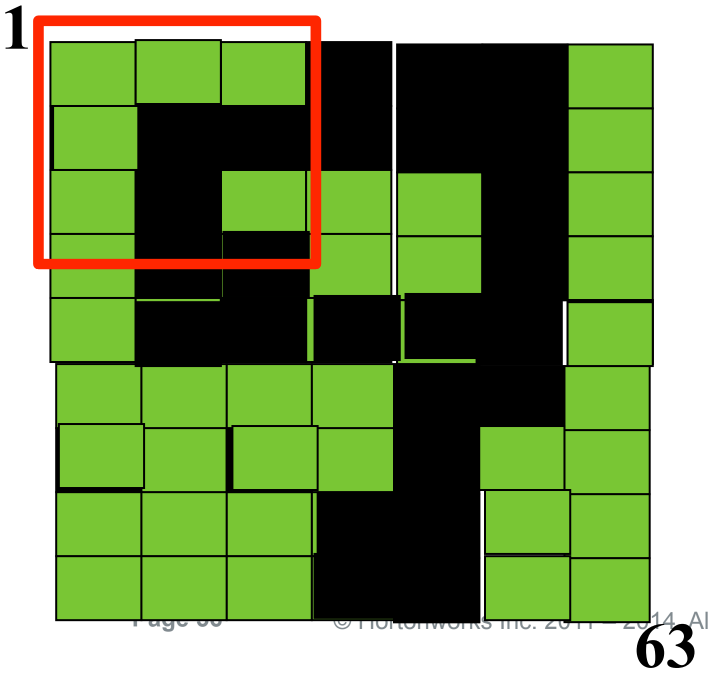
Hortonworks

What does this unit detect?



What does this unit detect?





Strong signal for a dark area in the top left corner

Hortonworks

What features might you expect a good NN to learn, when trained with data like this?

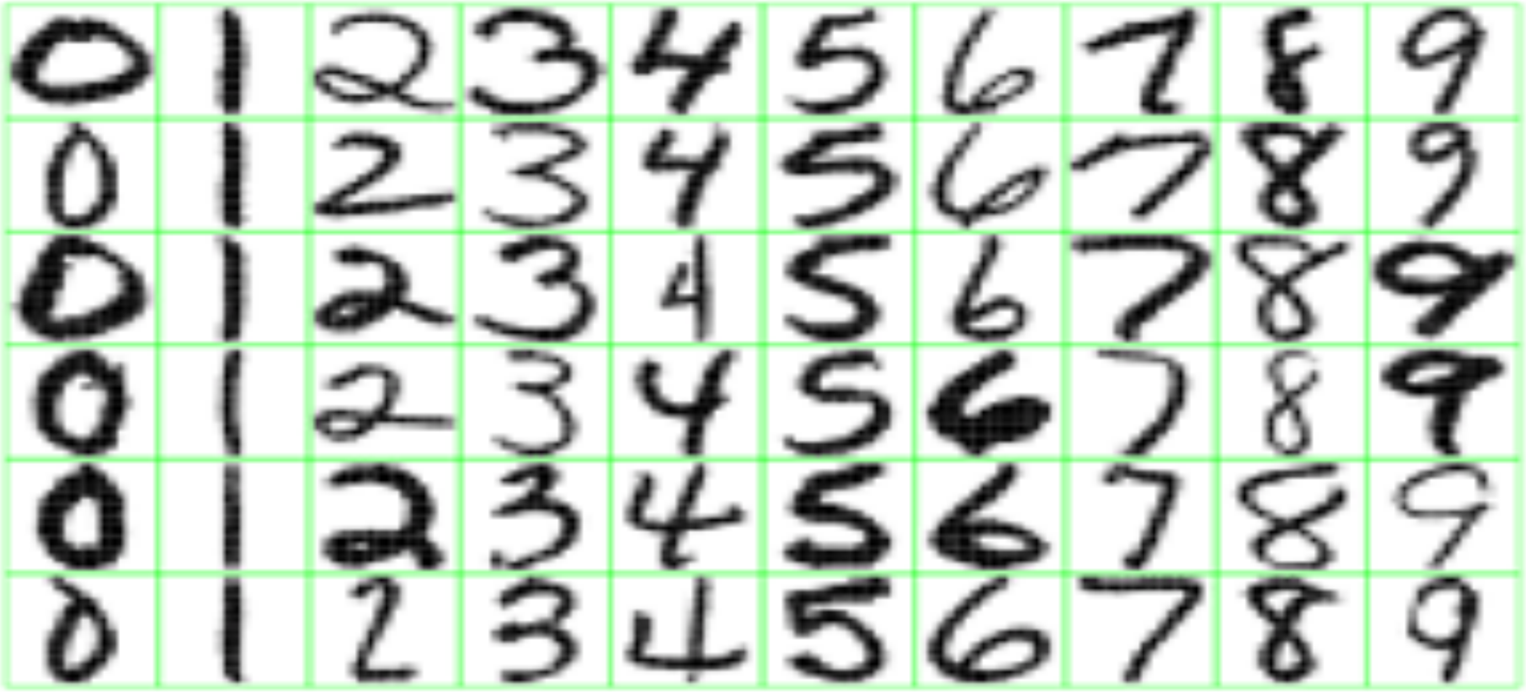


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

Hortonworks

1

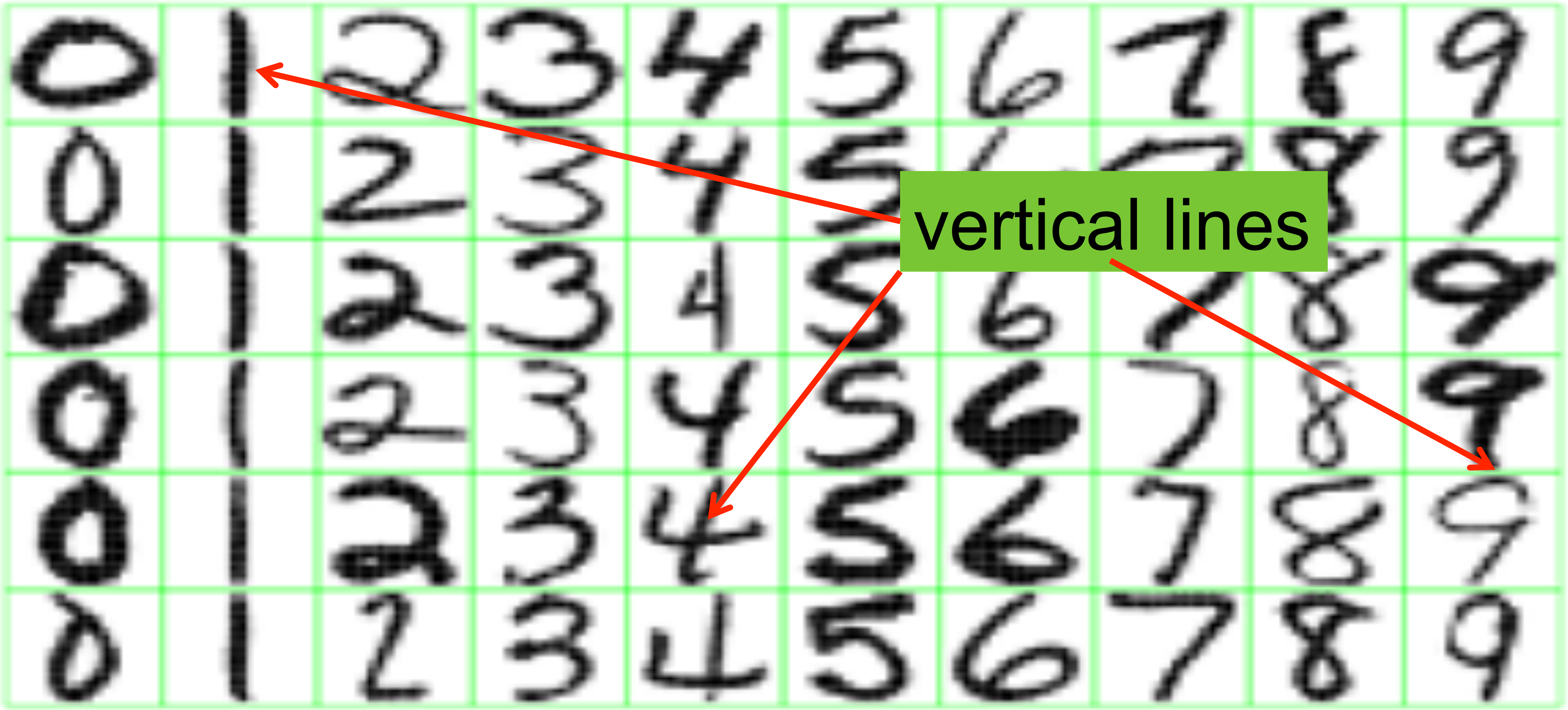


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

1

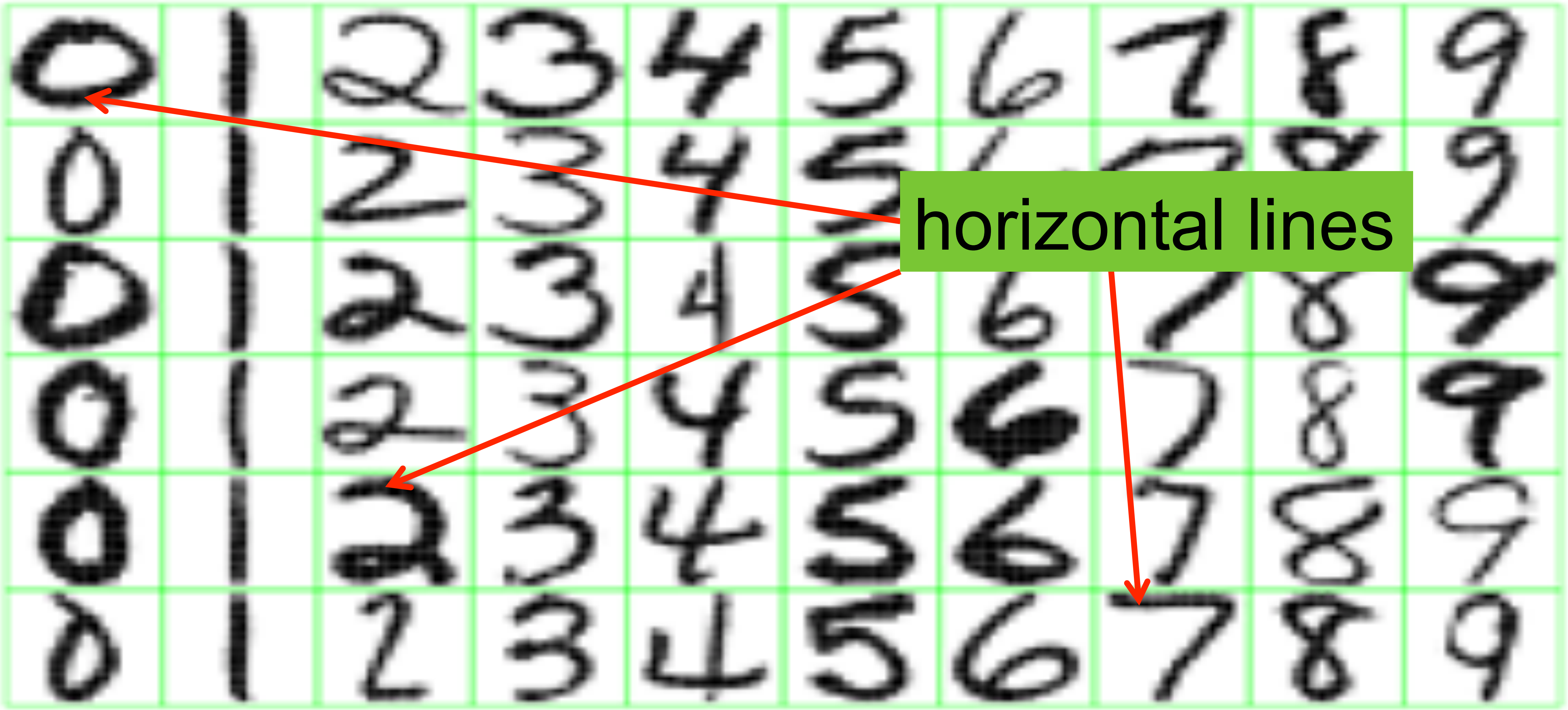
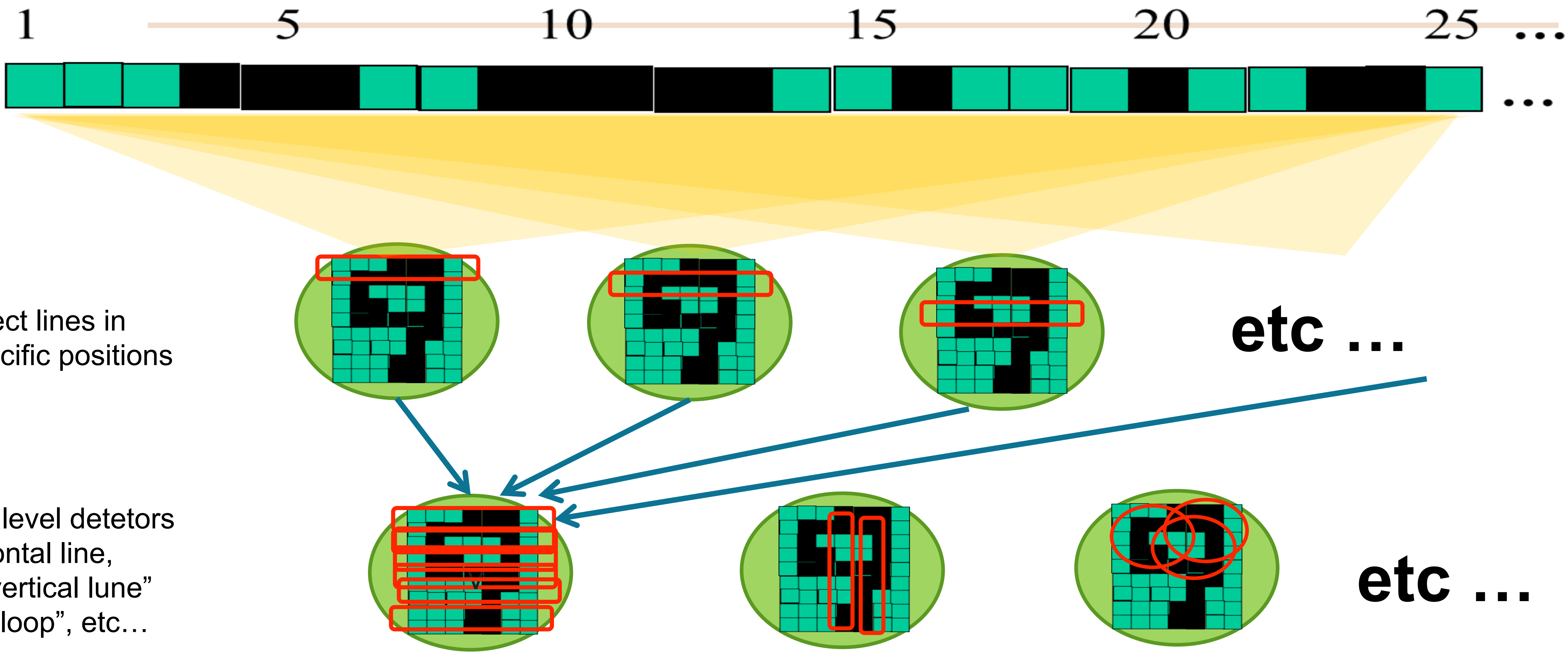


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

successive layers can learn higher-level features …

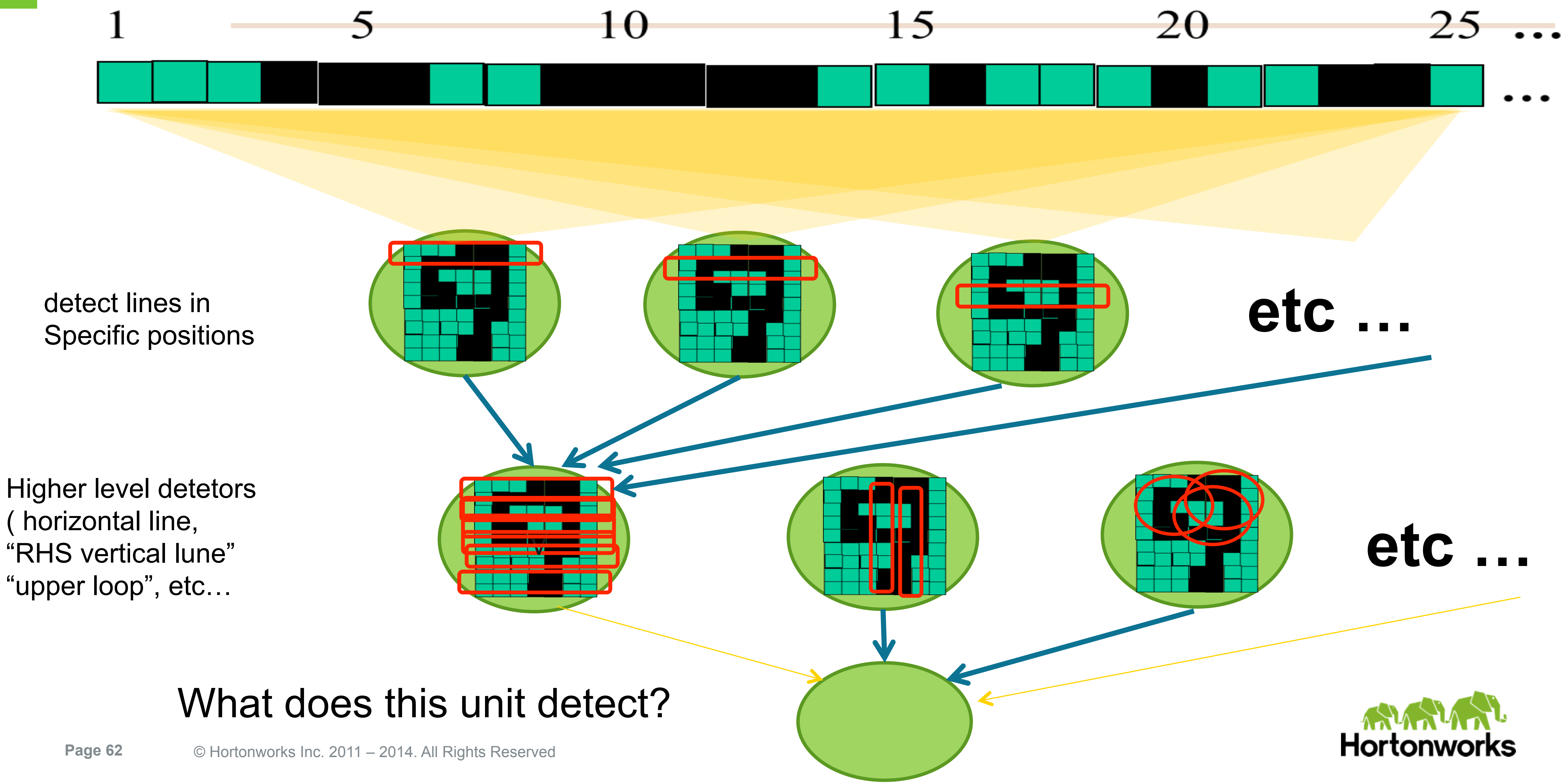


detect lines in
Specific positions

Higher level detetors
(horizontal line,
“RHS vertical lune”
“upper loop”, etc…

Hortonworks

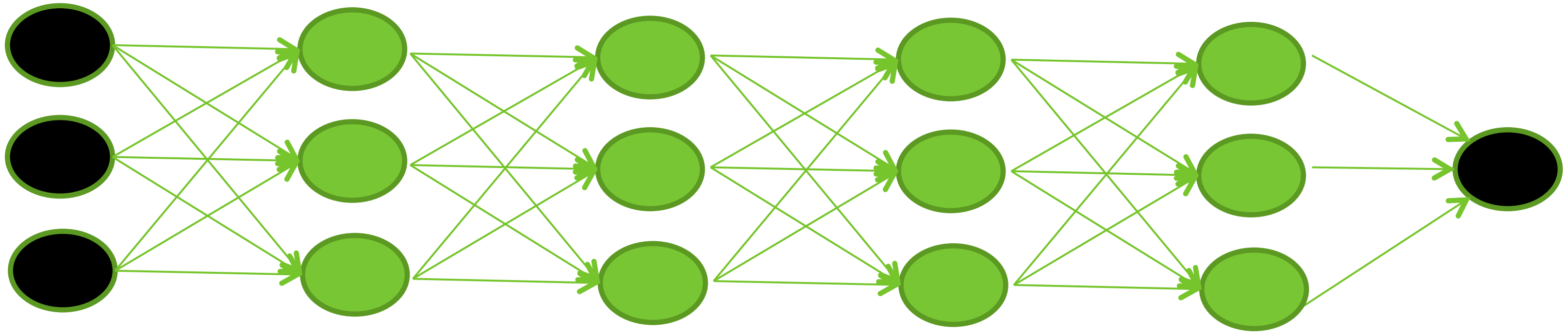
successive layers can learn higher-level features …



What does this unit detect?

Hortonworks

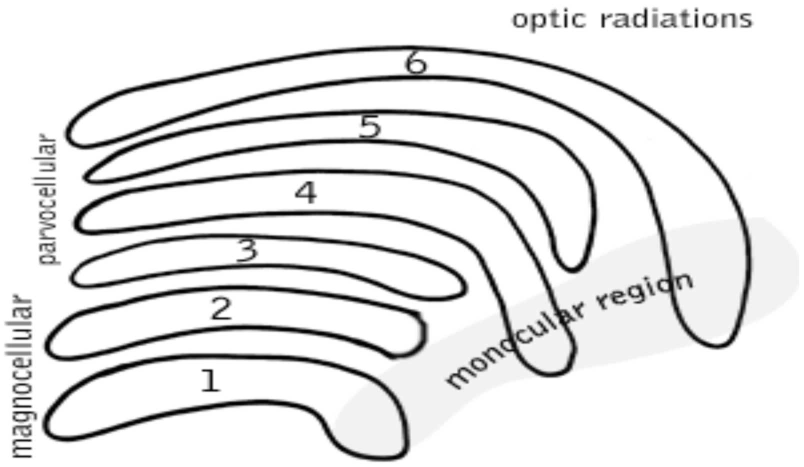
So: *multiple layers make sense*

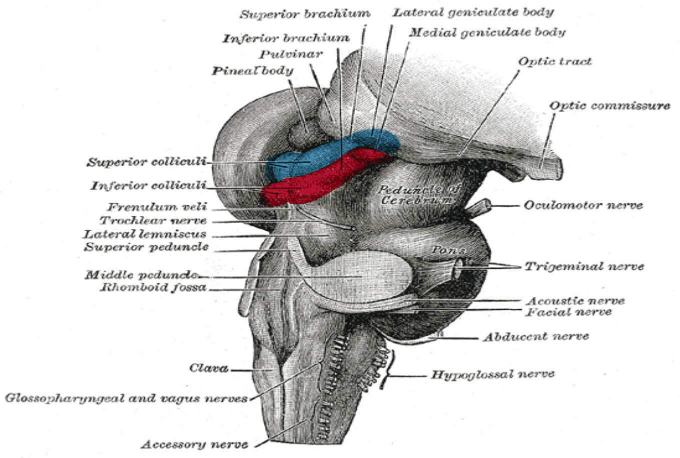


Hortonworks

So: *multiple layers make sense*

Your brain works that way

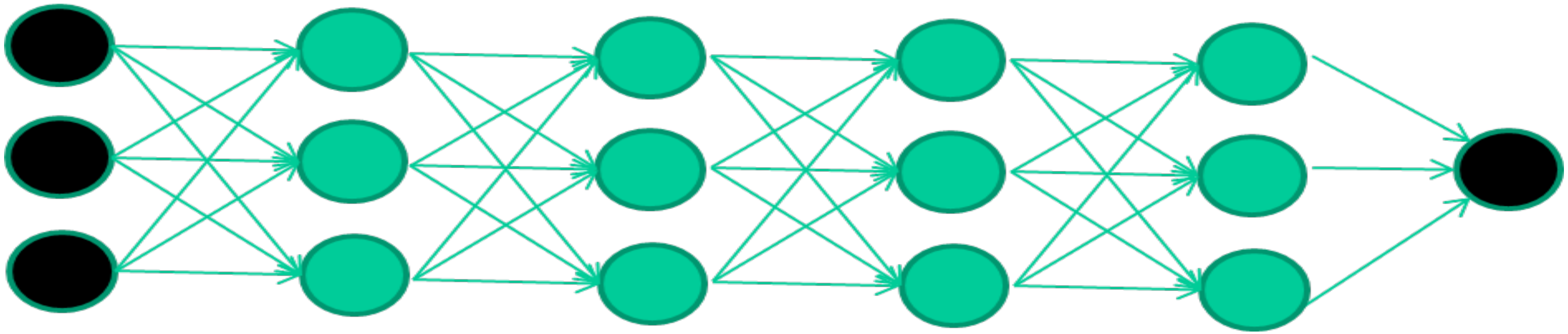




Hortonworks

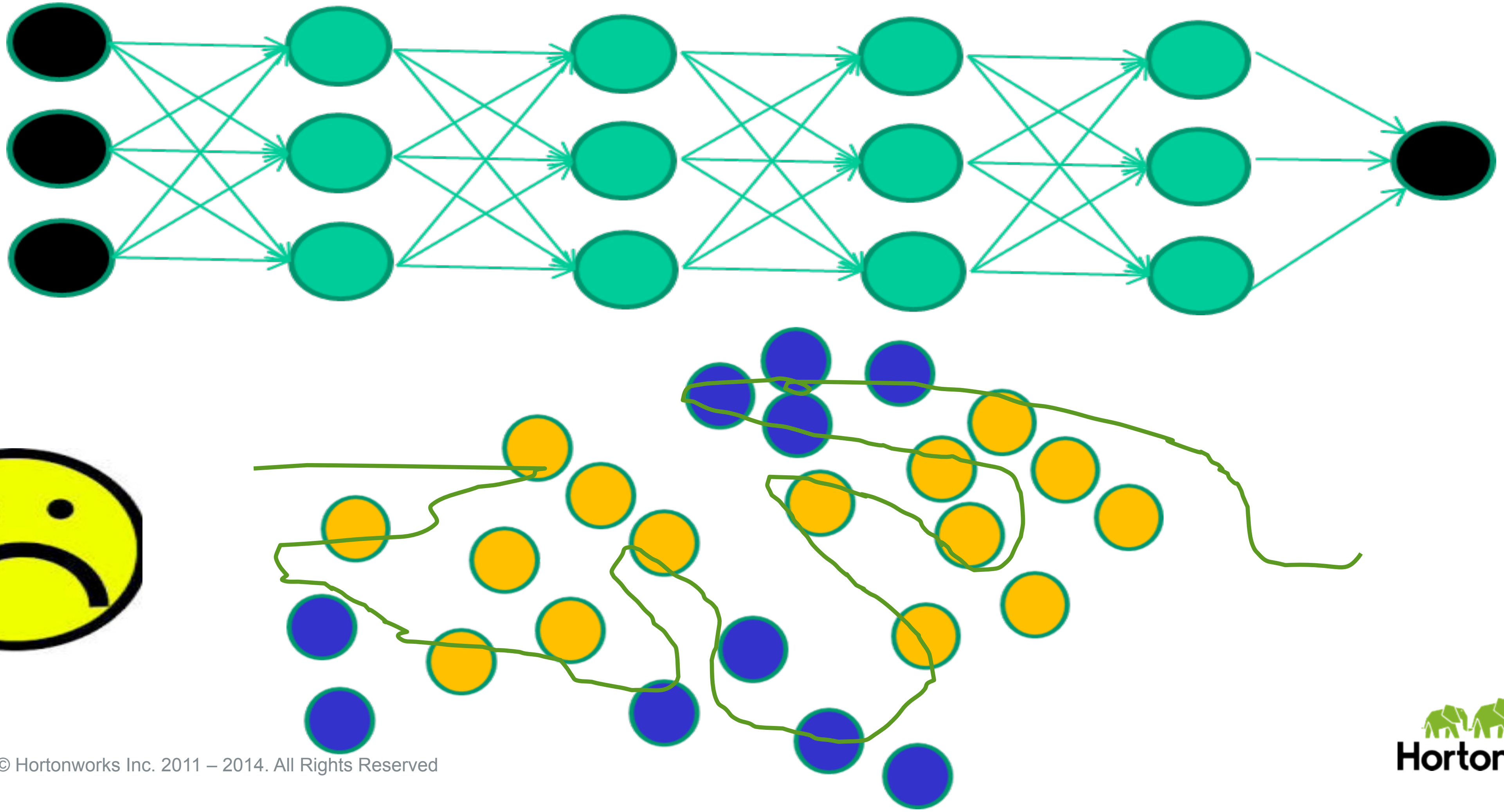
So: *multiple layers make sense*

Many-layer neural network architectures should be capable of learning the true underlying features and ‘feature logic’, and therefore generalise very well …



Hortonworks

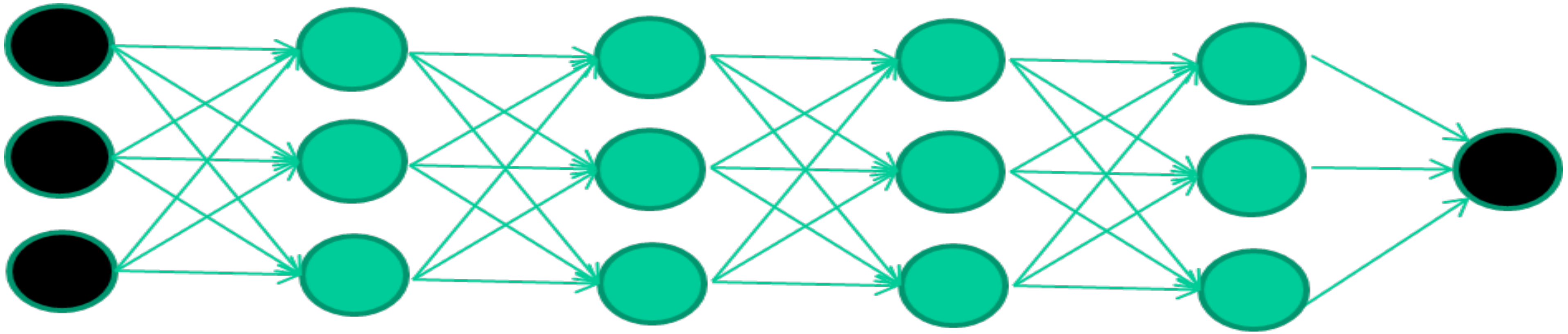
But, until very recently, weight-learning algorithms simply did not work on multi-layer architectures



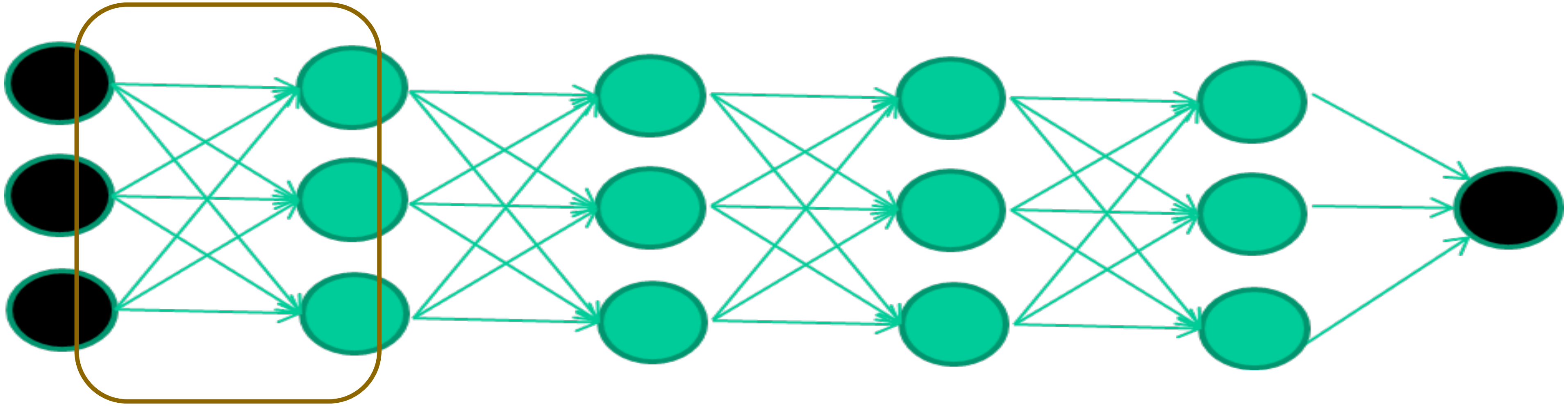
Hortonworks

Along came deep learning …

The new way to train multi-layer NNs…



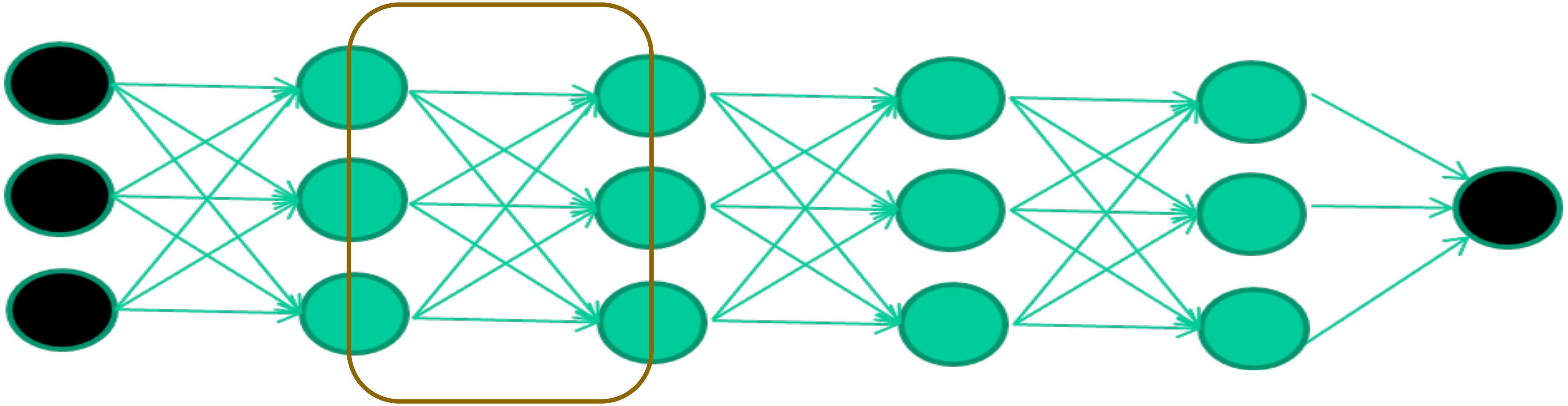
The new way to train multi-layer NNs…



Train **this** layer first

Hortonworks

The new way to train multi-layer NNs…

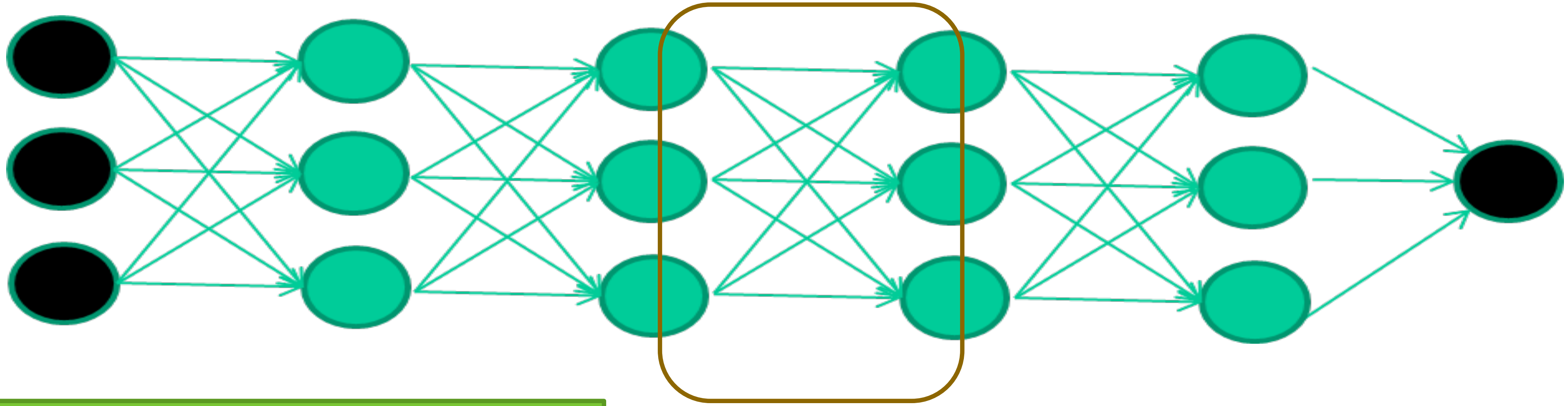


Train **this** layer first

then **this** layer

Hortonworks

The new way to train multi-layer NNs…

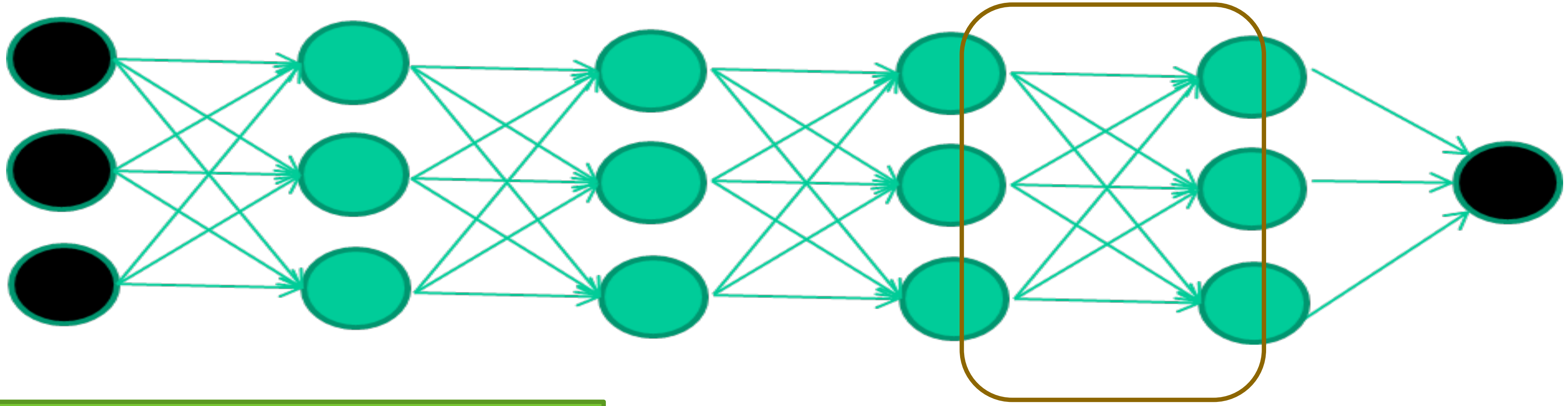


Train **this** layer first

then **this** layer

then **this** layer

The new way to train multi-layer NNs…



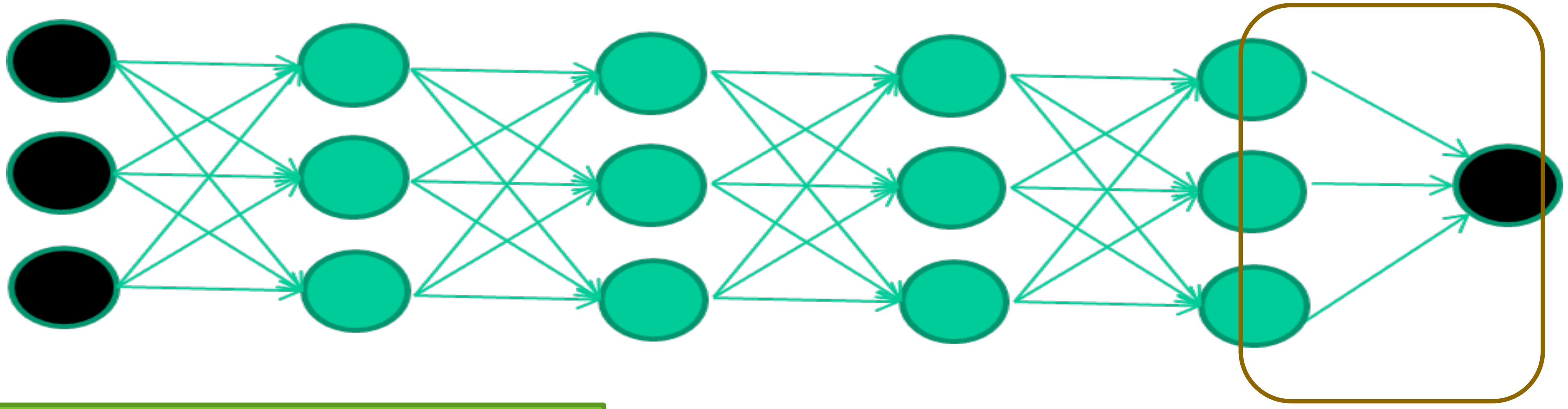
Train **this** layer first

then **this** layer

then **this** layer

then **this** layer

The new way to train multi-layer NNs…



Train **this** layer first

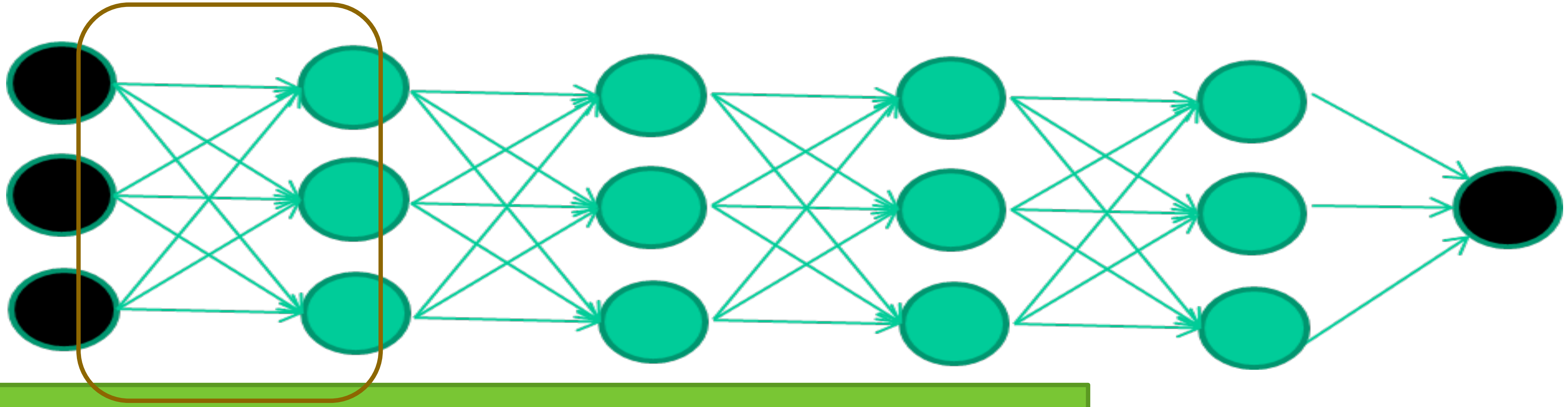
then **this** layer

then **this** layer

then **this** layer

finally **this** layer

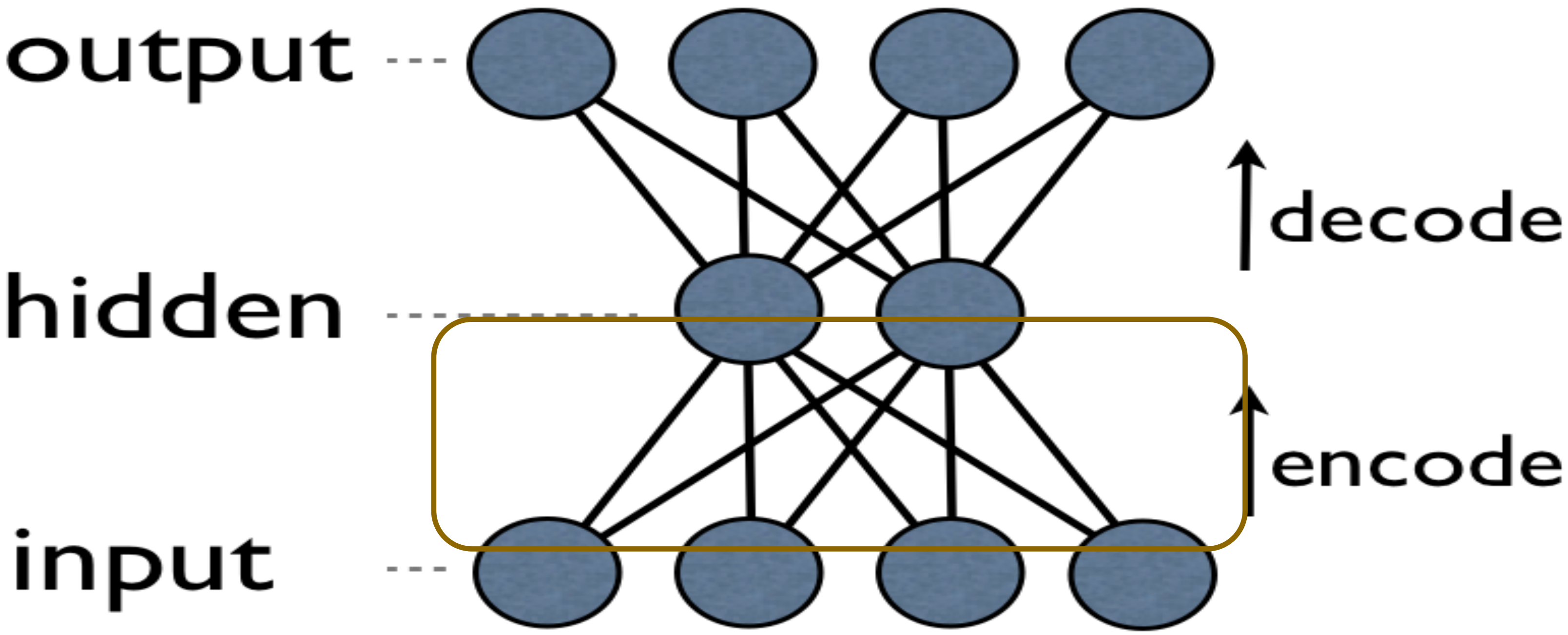
The new way to train multi-layer NNs…



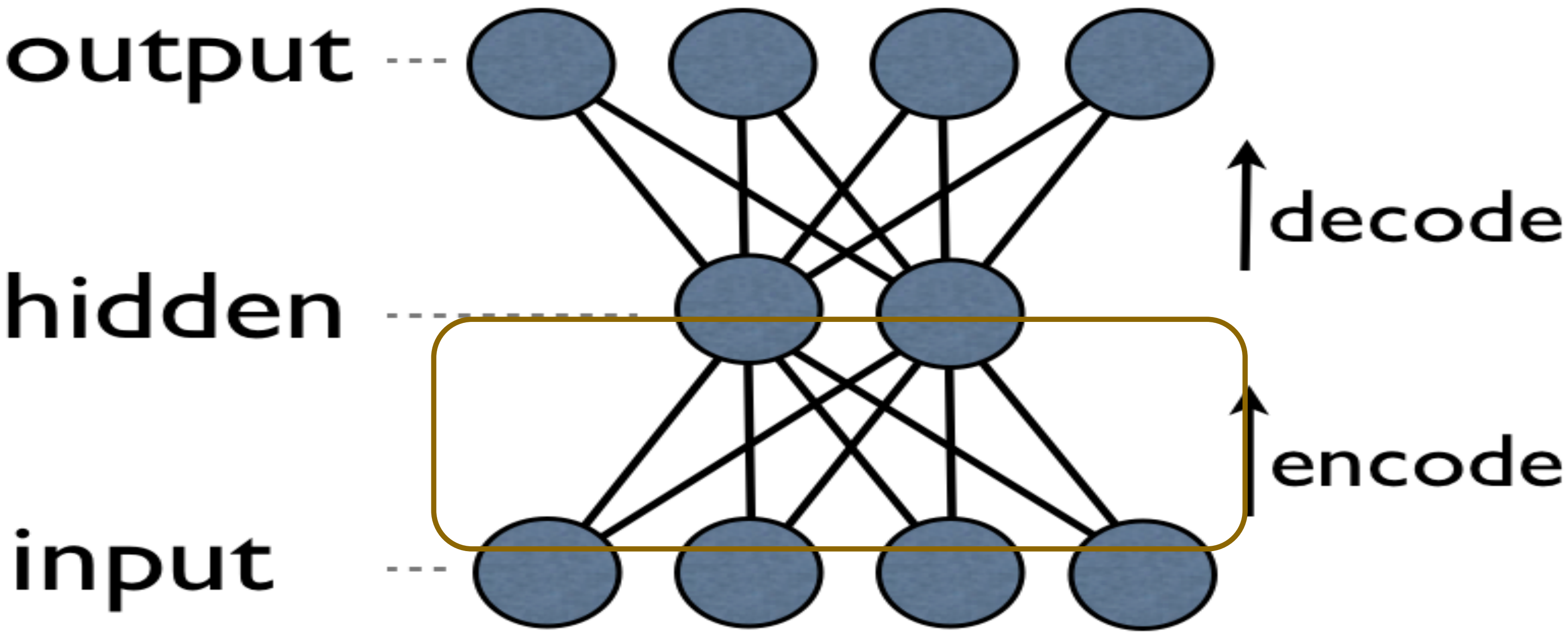
EACH of the (non-output) layers is trained to be an ***auto-encoder***

Basically, it is forced to learn good features that describe what comes from the previous layer

an auto-encoder is trained, with an absolutely standard weight-adjustment algorithm to reproduce the input



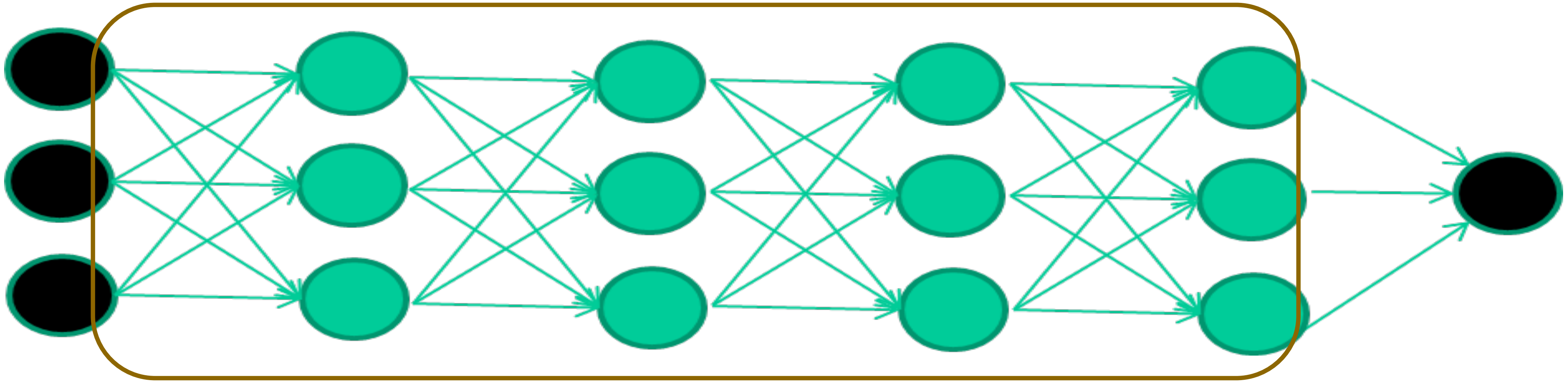
an auto-encoder is trained, with an absolutely standard weight-adjustment algorithm to reproduce the input



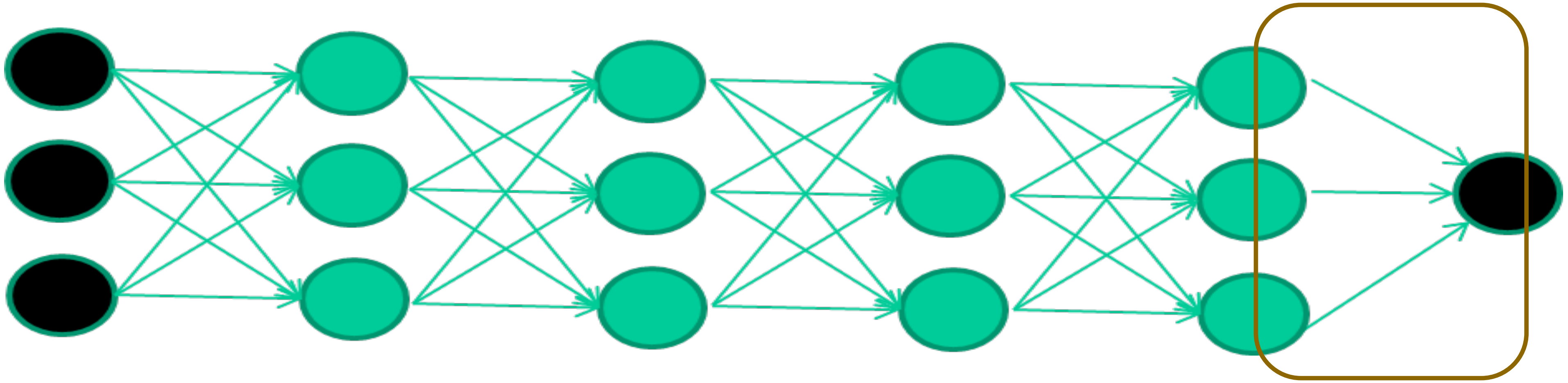
By making this happen with (many) fewer units than the inputs, this forces the 'hidden layer' units to become good feature detectors

Hortonworks

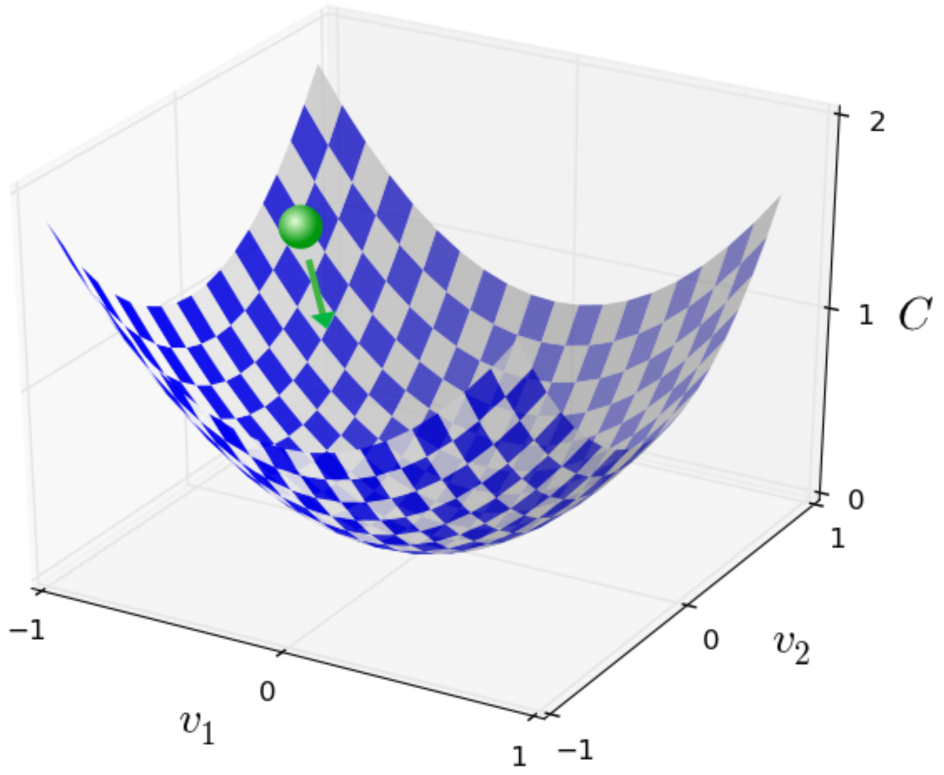
intermediate layers are each trained to be auto encoders (or similar)



Final layer trained to predict class based on outputs from previous layers



But, how does one *train*?



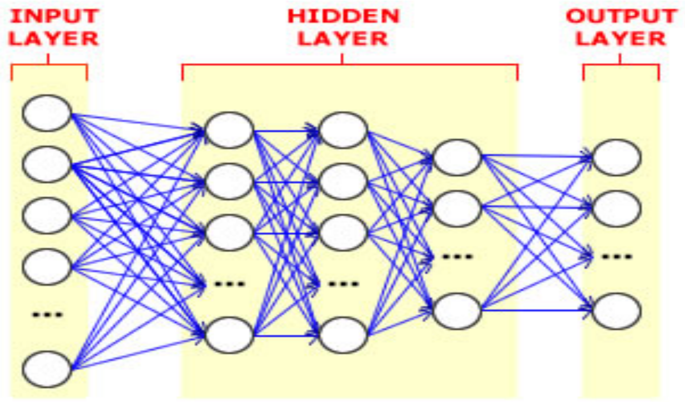
Overall, the NN is trying to minimize a cost function while adjusting the weights. To find the minima, **Gradient Descent** is used.

If the training set is very large, GD can be too slow since each input is evaluated in the cost function. So, one can sample a subset of input for computing GD. If sampling is done at random, it is called **Stochastic Gradient Descent**.

The implementation of this mathematical formulation is done by the **Error Backpropagation Algorithm.**

And that's that

- That's the basic idea
- There are many many types of deep learning,
- Different kinds of autoencoder, variations on architectures and training algorithms, etc…
- Very fast growing area …



Hortonworks

Doing this in real life

Typical Workflow

1. Ingest training data and store it
2. Split data set into: training, testing and validation sets
3. Vectorize and extract features to go into next step
4. Architect multi layer network, initialize
5. Feed data and train
6. Test and Validate
7. Repeat steps 4 and 5 until desired
8. Store model
9. Put model in app, start generalizing on real data.

Hortonworks

So what do you get?

1. Ingest training data and store it using Kafka, Flume, good old web scraping
2. Split data set into: training, testing and validation sets
3. Vectorize and extract features to go into next step
4. Architect multi layer network, initialize
5. Feed data and train
6. Test and Validate
7. Repeat steps 4 and 5 until desired
8. Store model
9. Put model in app, start generalizing on real data.

Steps 2, 3, 4 and 5: Use libraries such as Caffe, Theano, Deeplearning4j, H20

Hortonworks

Distributed Deep Learning on Hadoop

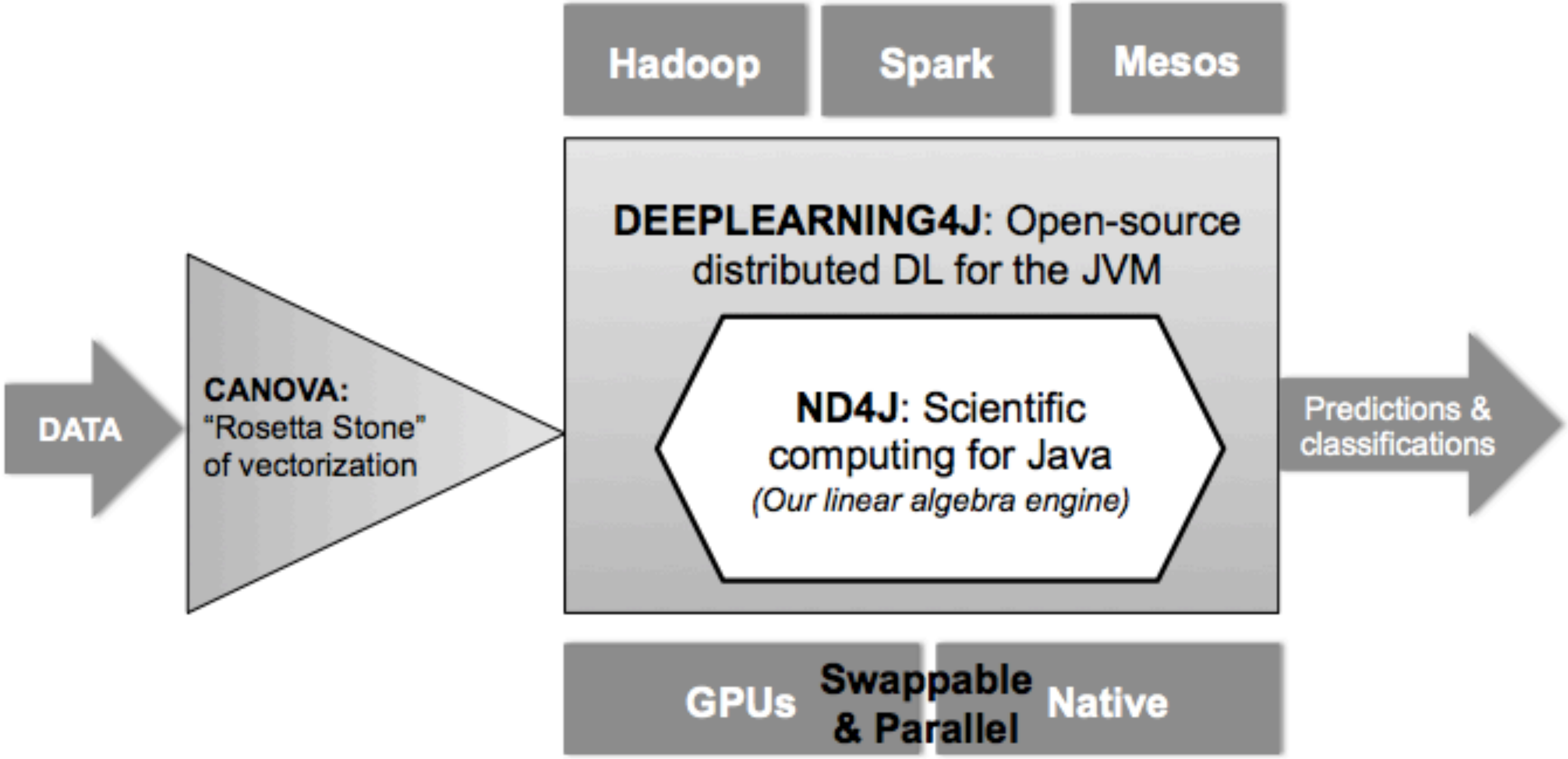
- In ASF: Apache Singa – new incubator project
- Two main partners of Hortonworks: Skymind, and H20
- Skymind is focussed on Deep Learning exclusively (deeplearning4j), H2O includes other ML libraries.
- Both provide scale out on HDP + Spark
- Skymind has GPU Acceleration built in - uses CUDA for doing linear algebra operations
- Both are open source, Apache licensed.

SKYMIND

H_2O

Hortonworks

Deeplearning4j Architecture



DL4J: Canova for Vectorization and Ingest

- Canova uses an input/output format system (similar to how Hadoop uses MapReduce)
- Supports all major types of input data (text, CSV, audio, image and video)
- Can be extended for specialized input formats
- Connects to Kafka

Hortonworks

ND4J:

- N-dimensional vector library
- Scientific computing for JVM
- DL4J uses it to do linear algebra for backpropagation
- Supports GPUs via CUDA and Native via Jblas
- Deploys on Android
- DL4J code remains unchanged whether using GPU or CPU

How to chose a Neural Net in DL4J core?

Data Sector	Use Case	Input	Transform	Neural Net
Text	Sentiment analysis	Word vector	Gaussian Rectified	RNTN or DBN (with moving window)
	Named-entity recognition	Word vector	Gaussian Rectified	RNTN or DBN (with moving window)
	Part-of-speech tagging	Word vector	Gaussian Rectified	RNTN or DBN (with moving window)
	Semantic-role labeling	Word vector	Gaussian Rectified	RNTN or DBN (with moving window)
Document	Topic modeling/ semantic hashing (unsupervised)	Word count probability	Can be Binary	Deep Autoencoder (wrapping a DBN or SDA)
	Document classification (supervised)	TF-IDF (or word count prob.)	Binary	Deep-belief network, Stacked Denoising Autoencoder
Image	Image recognition	Binary	Binary (visible and hidden)	Deep-belief network
		Continuous	Gaussian Rectified	Deep-belief network
	Multi-object recognition			Convolutional Net, RNTN (image vectorization forthcoming)
	Image search/ semantic hashing		Gaussian Rectified	Deep Autoencoder (wrapping a DBN)
Sound	Voice recognition		Gaussian Rectified	Recurrent Net
				Moving window for DBN or ConvNet
Time Series	Predictive analytics		Gaussian Rectified	Recurrent Net
				Moving window for DBN or ConvNet

Hortonworks

Demo!

Hortonworks

Thank You

hortonworks.com

Hortonworks