Introduction to Deep Learning

DL

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

/ MOBILE

Ina Fried

Microsoft's Skype "Star Trek" Language Translator Takes on Tower of Babel

May 27, 2014, 5:48 PM PDT

Remember the universal translator on Star Trek? The gad
to aliens?

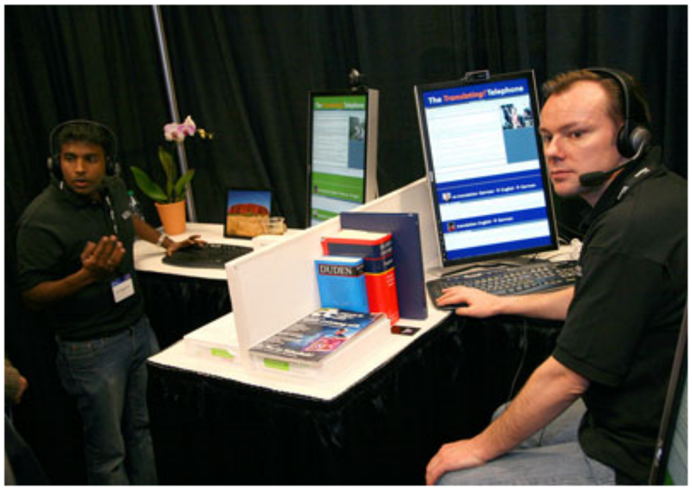
Skype to get 'real-time' translator

Analysts say the translation feature could have wide ranging applications

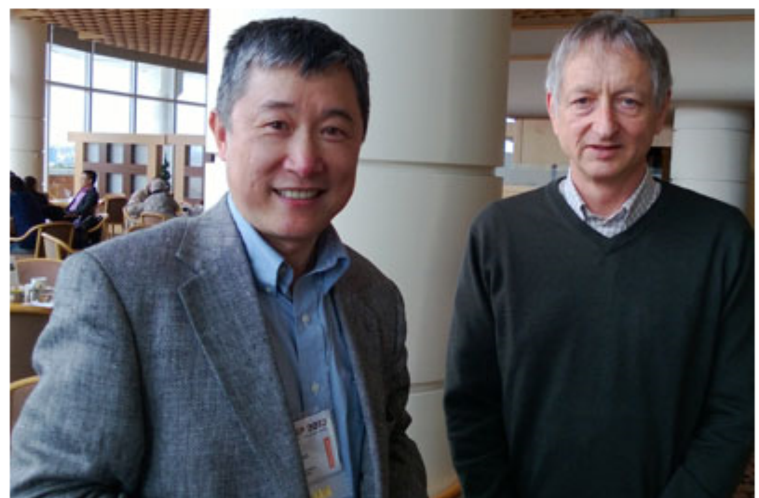
Enabling Cross-Lingual Conversations in Real Time

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

Impact of deep learning in speech technology

MIT Technology Review

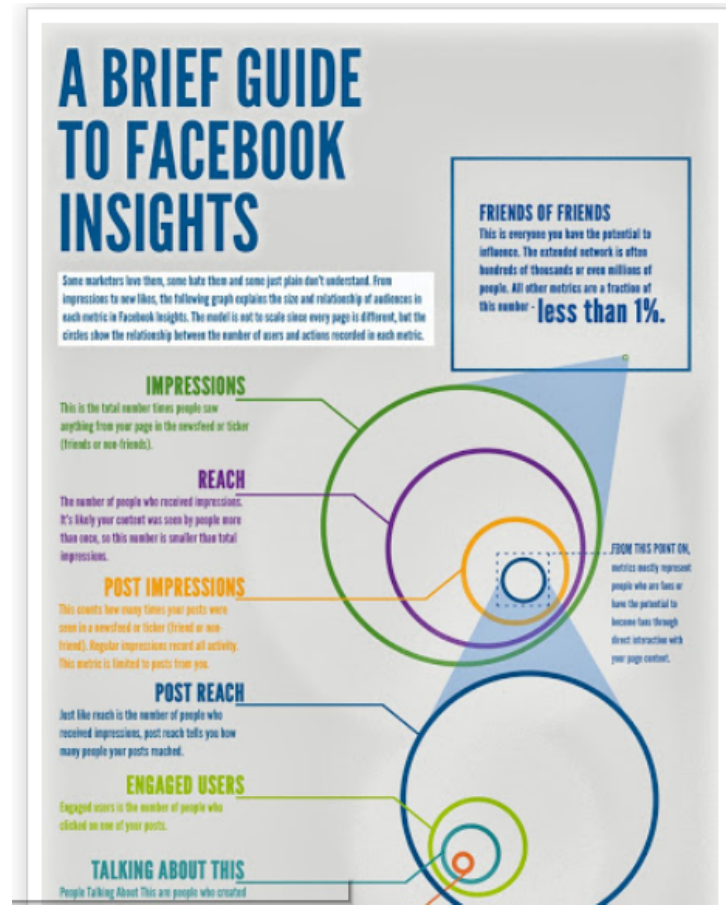
Facebook Launches Advanced AI Effort to Find Meaning in Your Posts

September 20, 2013

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.



Deep Learning's Role in the Age of Robots

BY JULIAN GREEN, JETPAC 05.02.14 2:56 PM

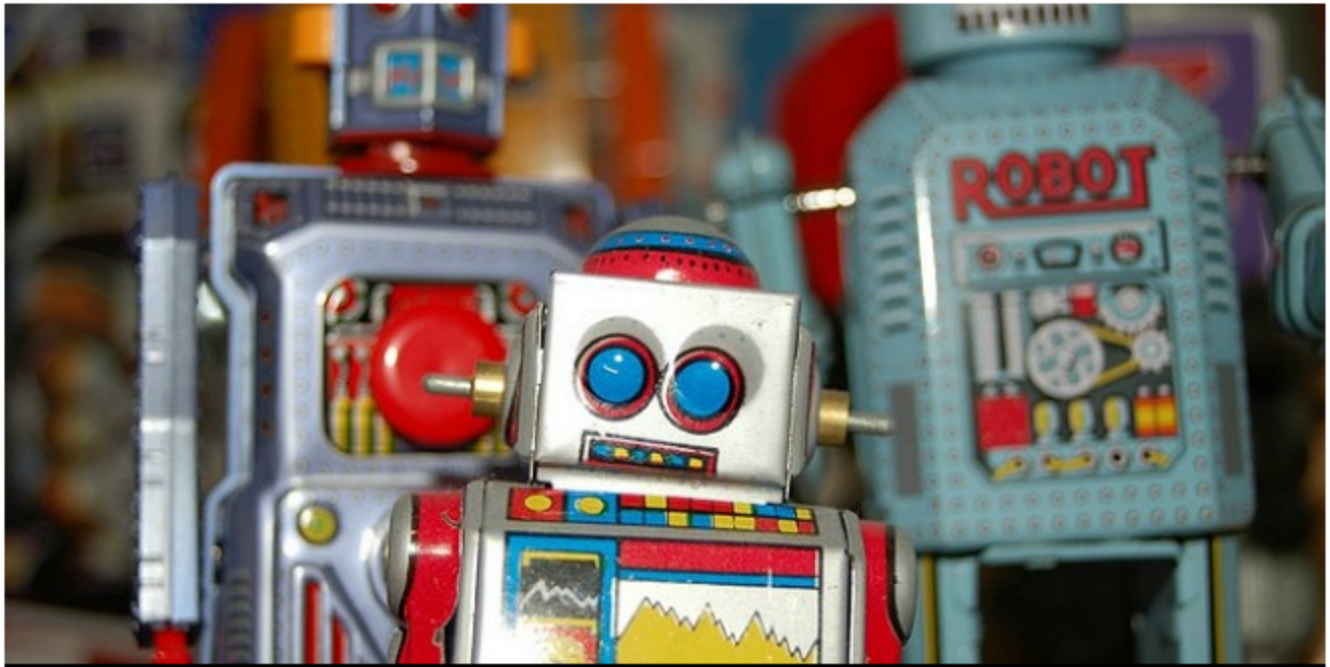


Image: jeffedoe/Flickr

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

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

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

A▼ A▲

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.

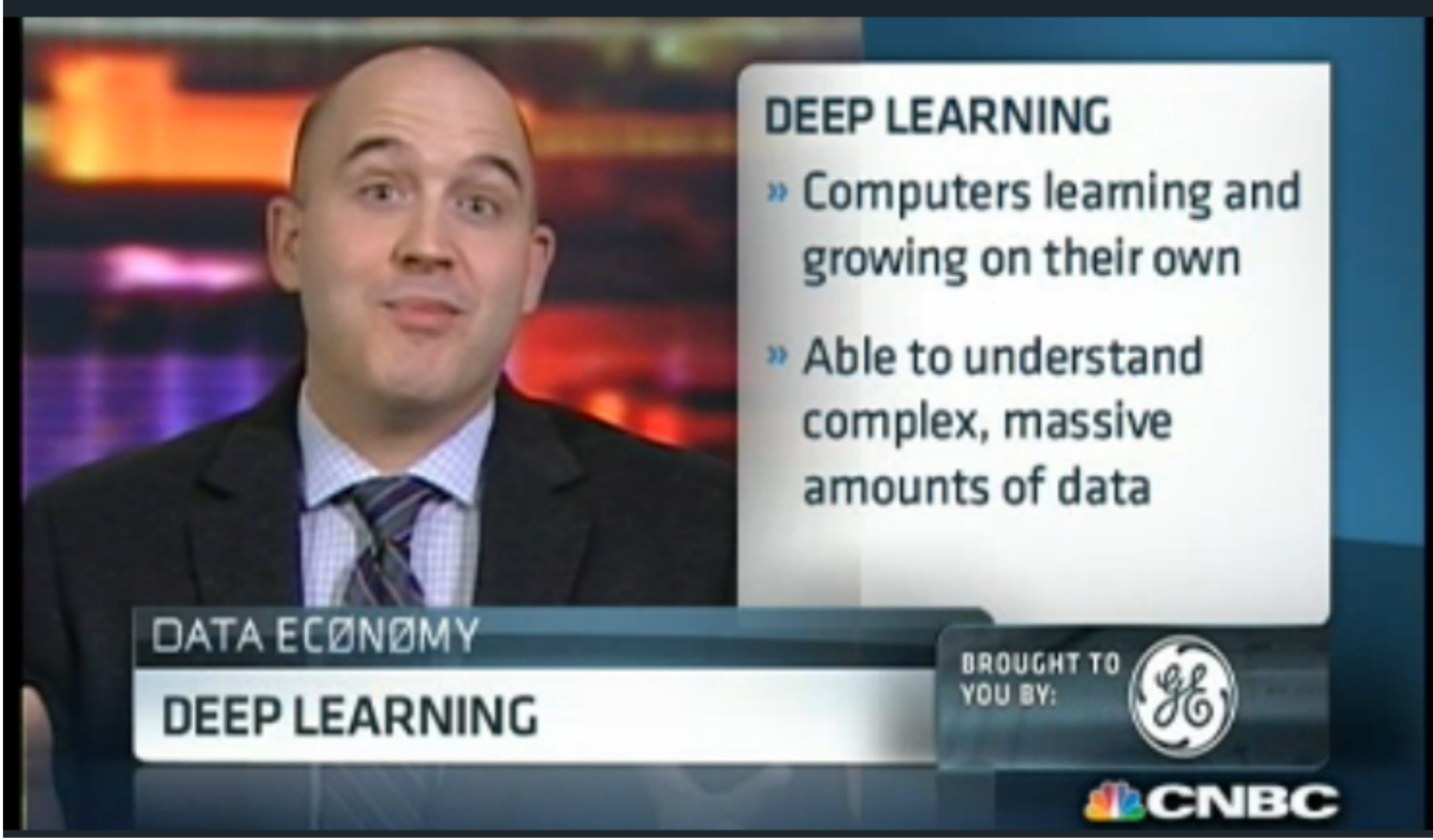
artificial intelligence / machine-learning / open source

A startup called Skymind launches, pushing open source deep learning

by Derrick Harris JUN. 2, 2014 - 10:03 AM PDT

SUMMARY: *Skymind is providing commercial support and services for an open source project called deeplearning4j. It's a collection of of approaches to deep learning that mimic those developed by leading researchers, but tuned for enterprise adoption.*


photo: deviantART / rajasegar



Is **Deep Learning**, the 'holy grail' of big data? - CNBC - Video

video.cnbc.com/gallery/?video=3000192292

Aug 22, 2013

Derrick Harris, GigaOM, explains how "**Deep Learning**" computers are able to process and understand ...

Deep learning

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.**

hmmm… OK, but:

3. **multilayer neural networks have been around for**

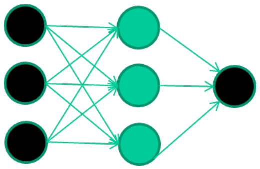
25 years. What's actually new?

hmmm… OK, but:

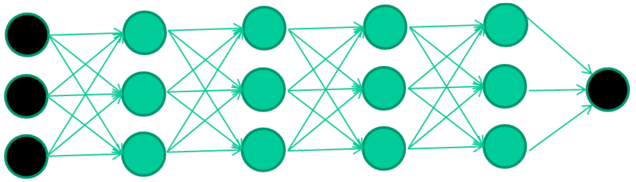
3. **multilayer neural networks have been around for**

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



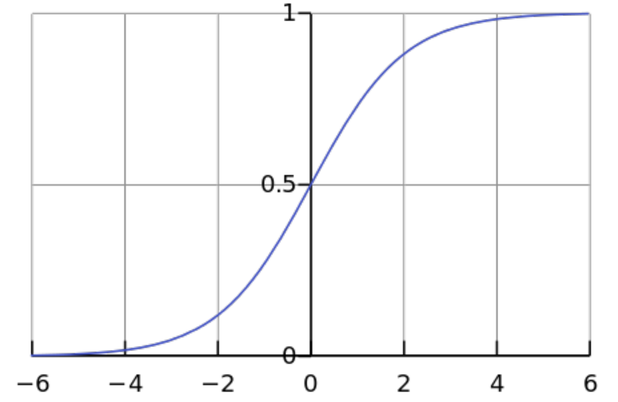
what's new is:

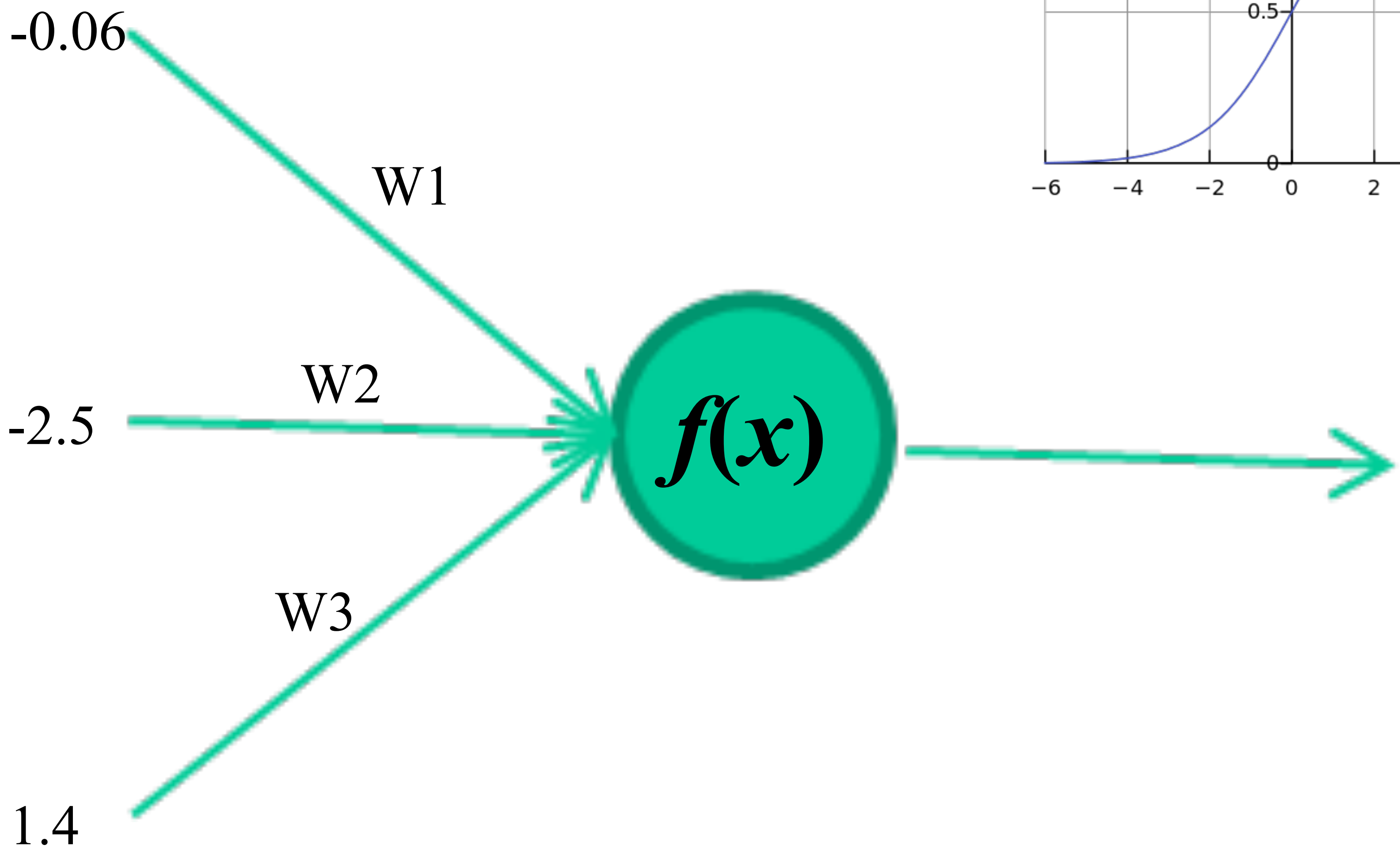
algorithms for training many-later networks

longer answers

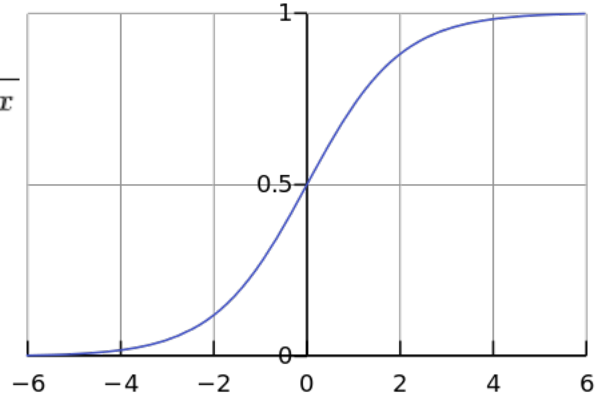
1. reminder/quick-explanation of how neural network weights are learned;
2. the idea of **unsupervised feature learning** (why 'intermediate features' are important for difficult classification tasks, and how NNs seem to naturally learn them)
3. The 'breakthrough' – the simple trick for training Deep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

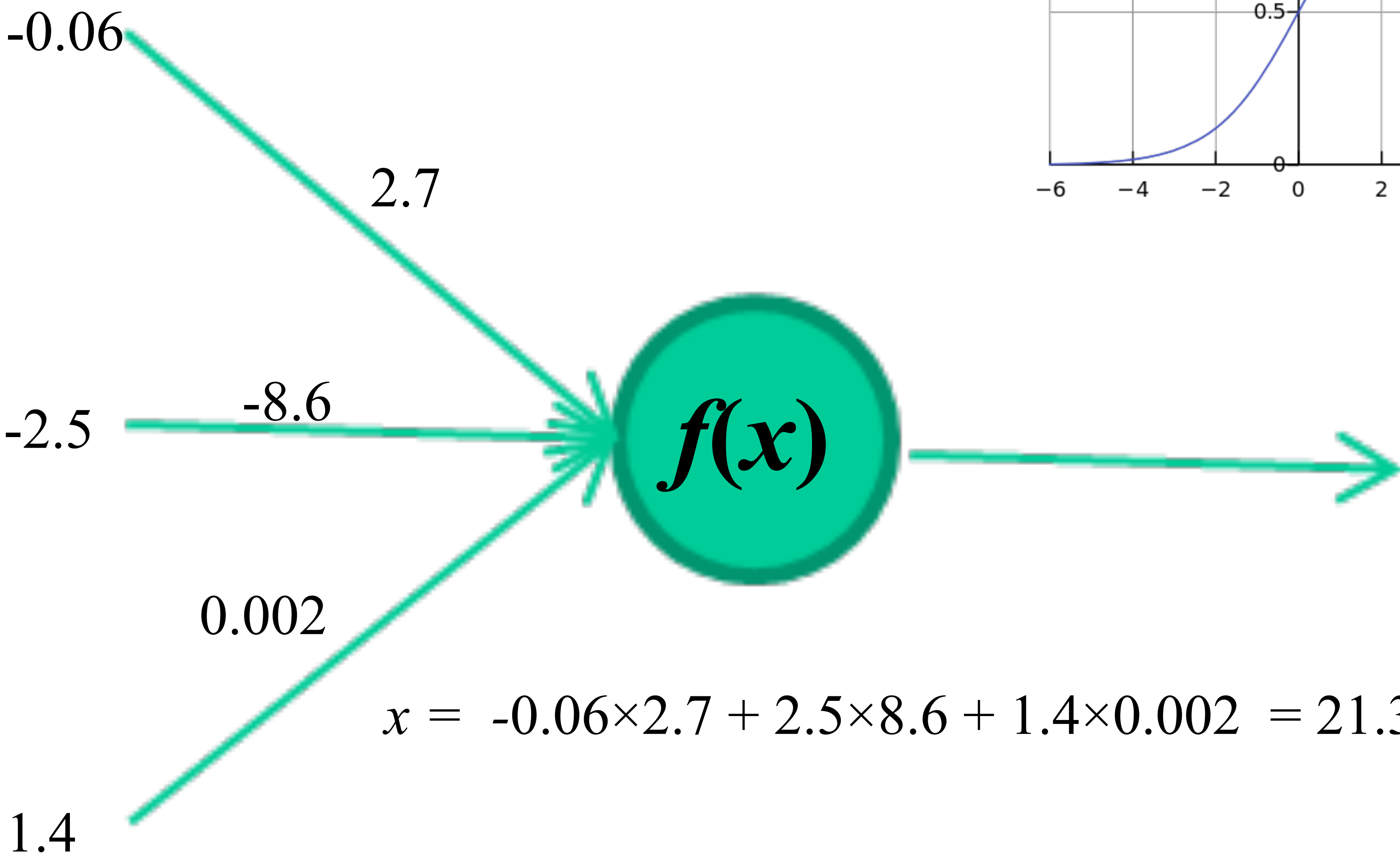
$$f(x) = \frac{1}{1 + e^{-x}}$$





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



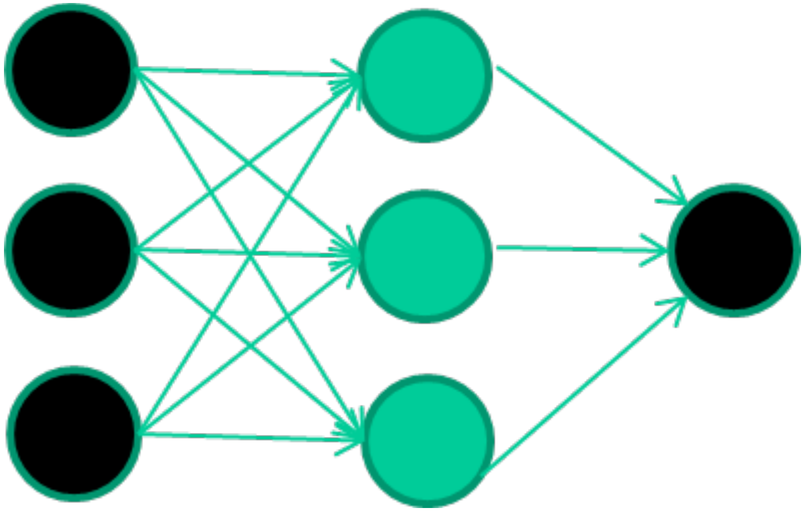


$$x = -0.06\times2.7 + 2.5\times8.6 + 1.4\times0.002 = 21.34$$

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

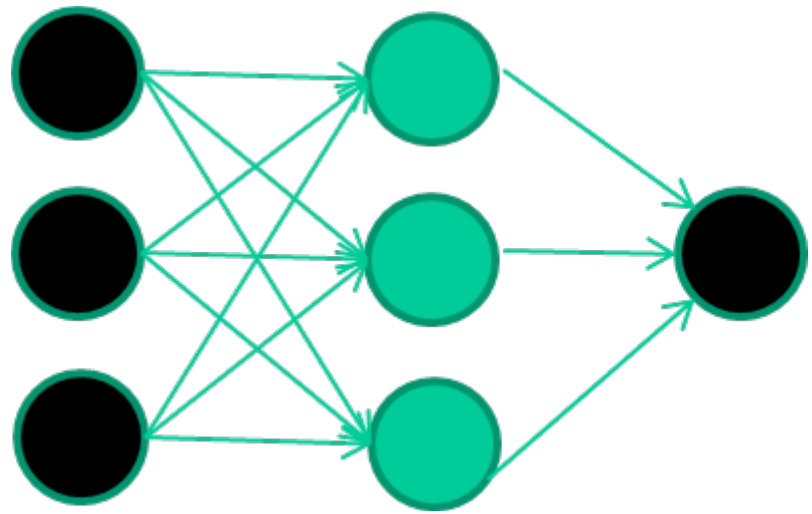
etc …



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

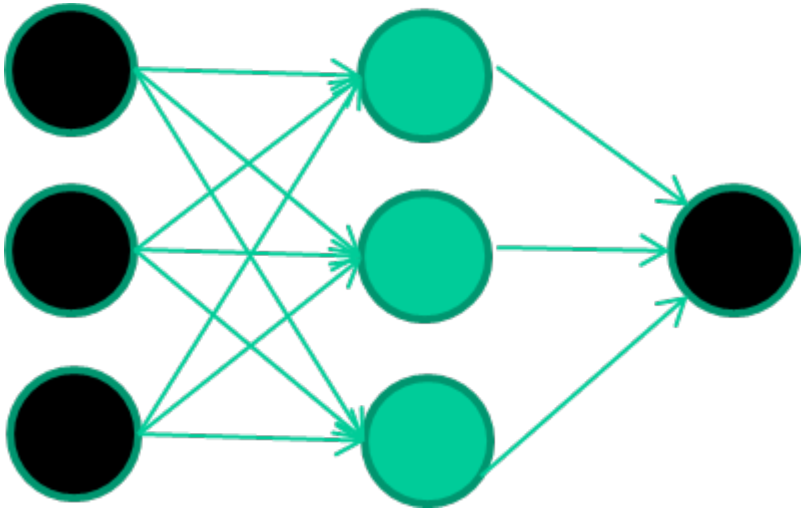
etc …



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 …			

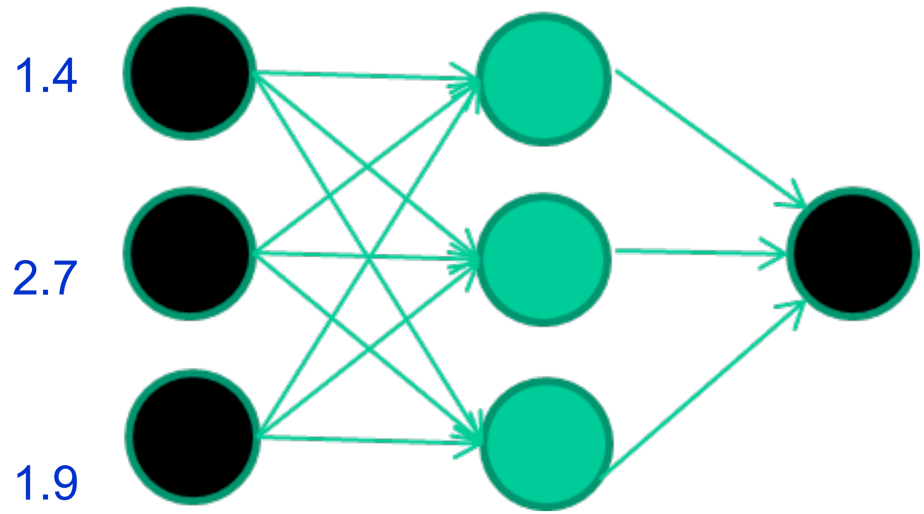
Initialise with random weights



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 …			

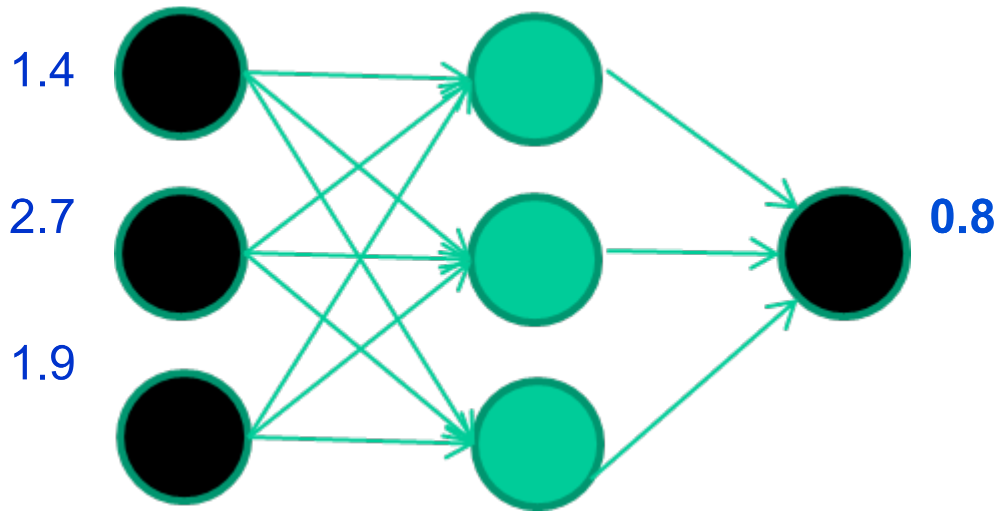
Present a training pattern



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 …			

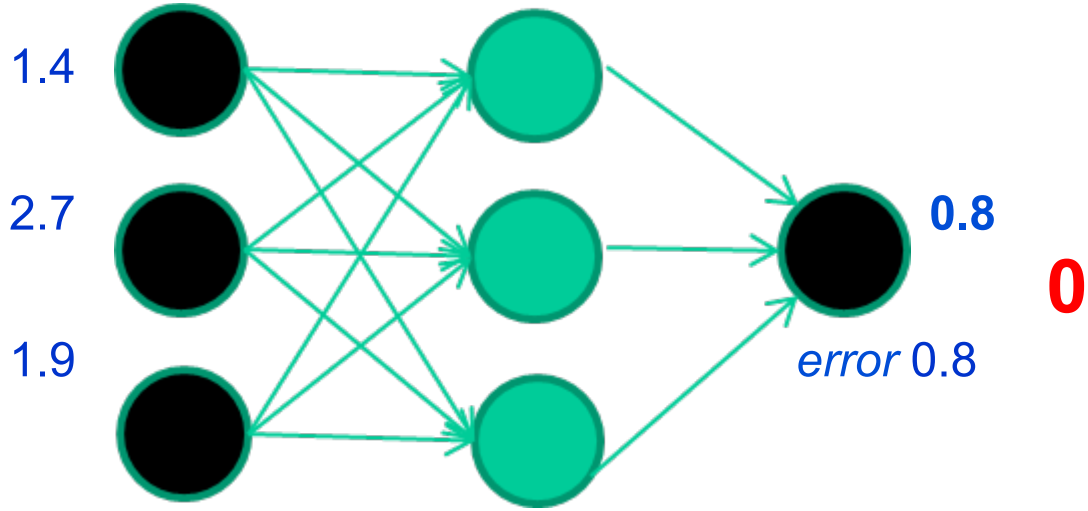
Feed it through to get output



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 …			

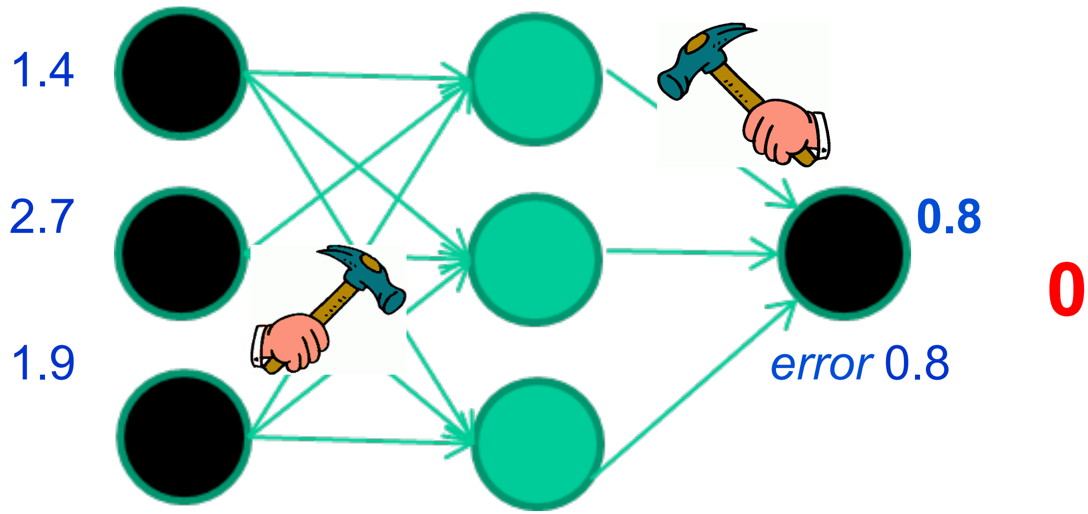
Compare with target output



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 …			

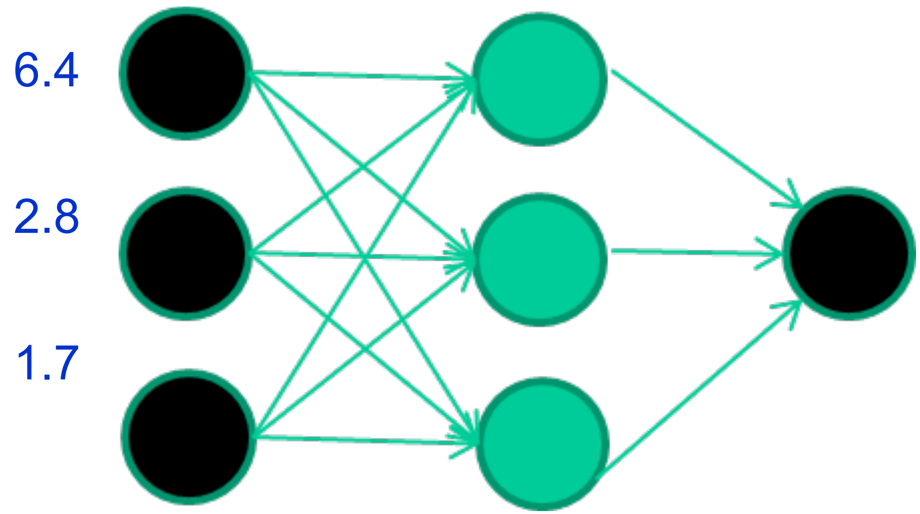
Adjust weights based on error



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 …			

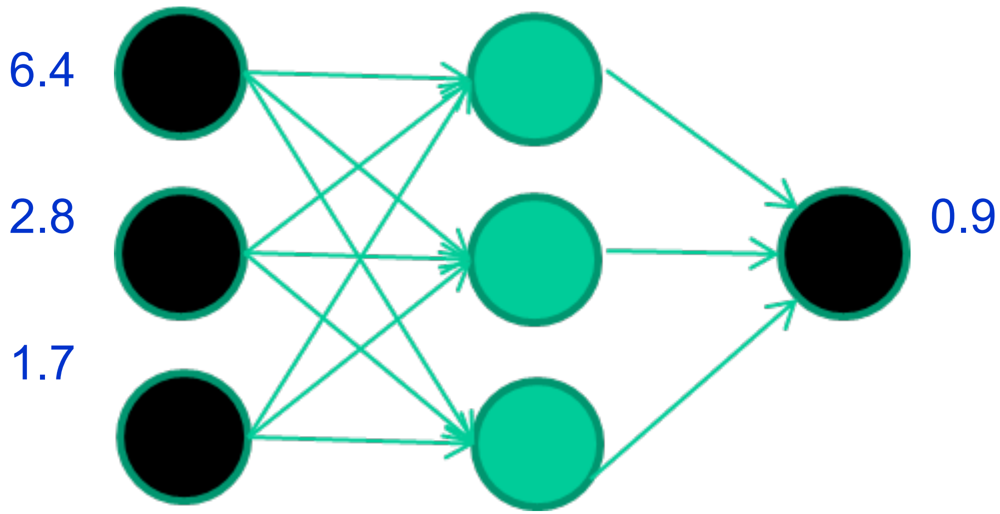
Present a training pattern



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 …			

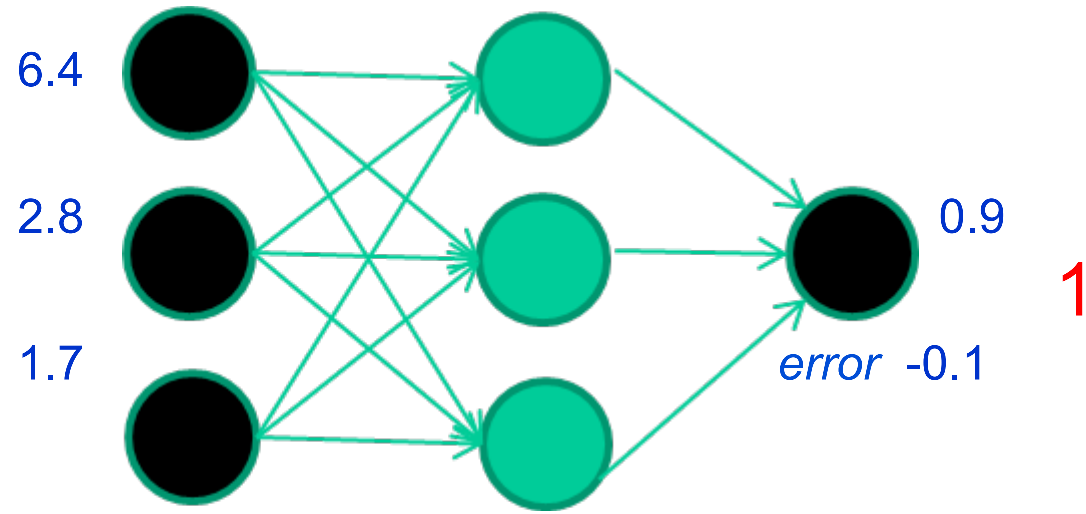
Feed it through to get output



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 …			

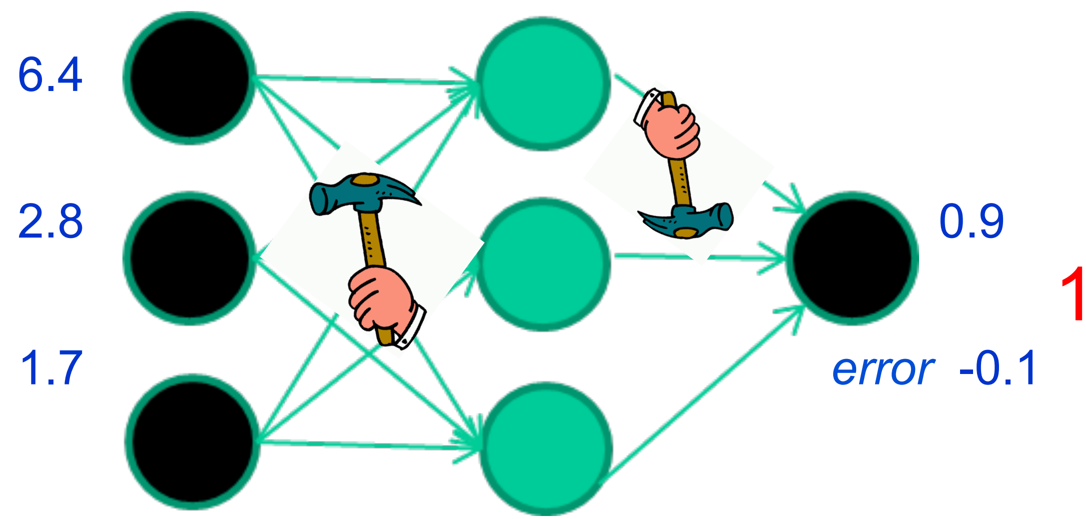
Compare with target output



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 …			

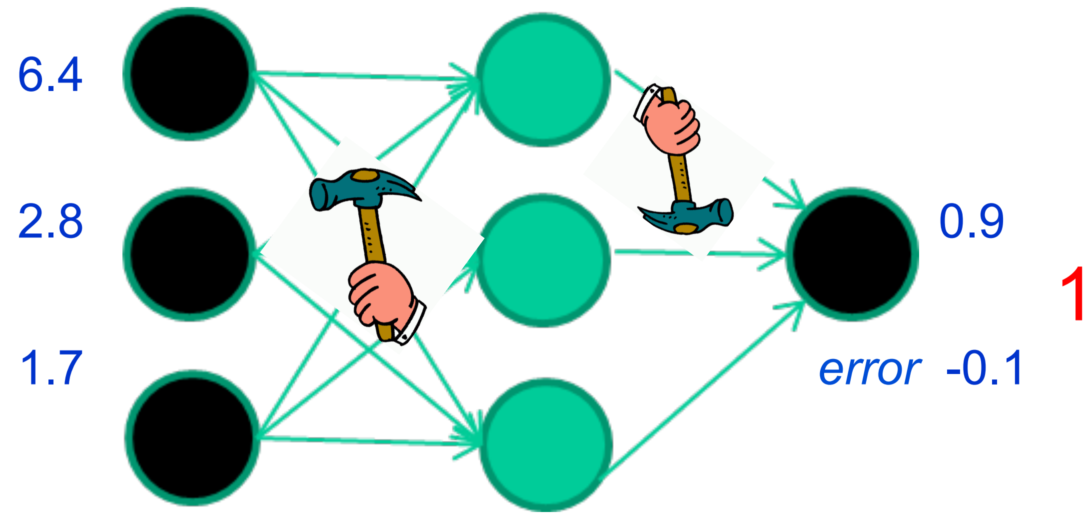
Adjust weights based on error



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 ….

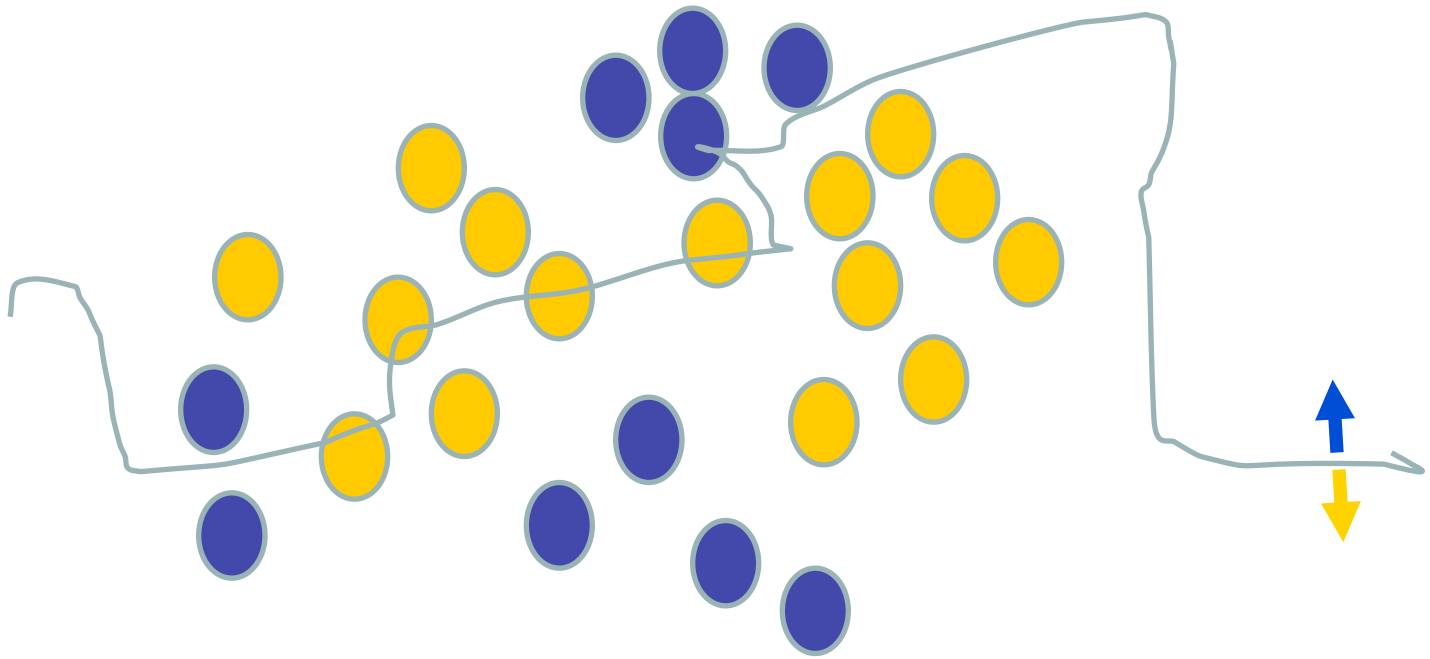


Repeat this thousands, maybe millions of times – each time taking a random training instance, and making slight weight adjustments

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

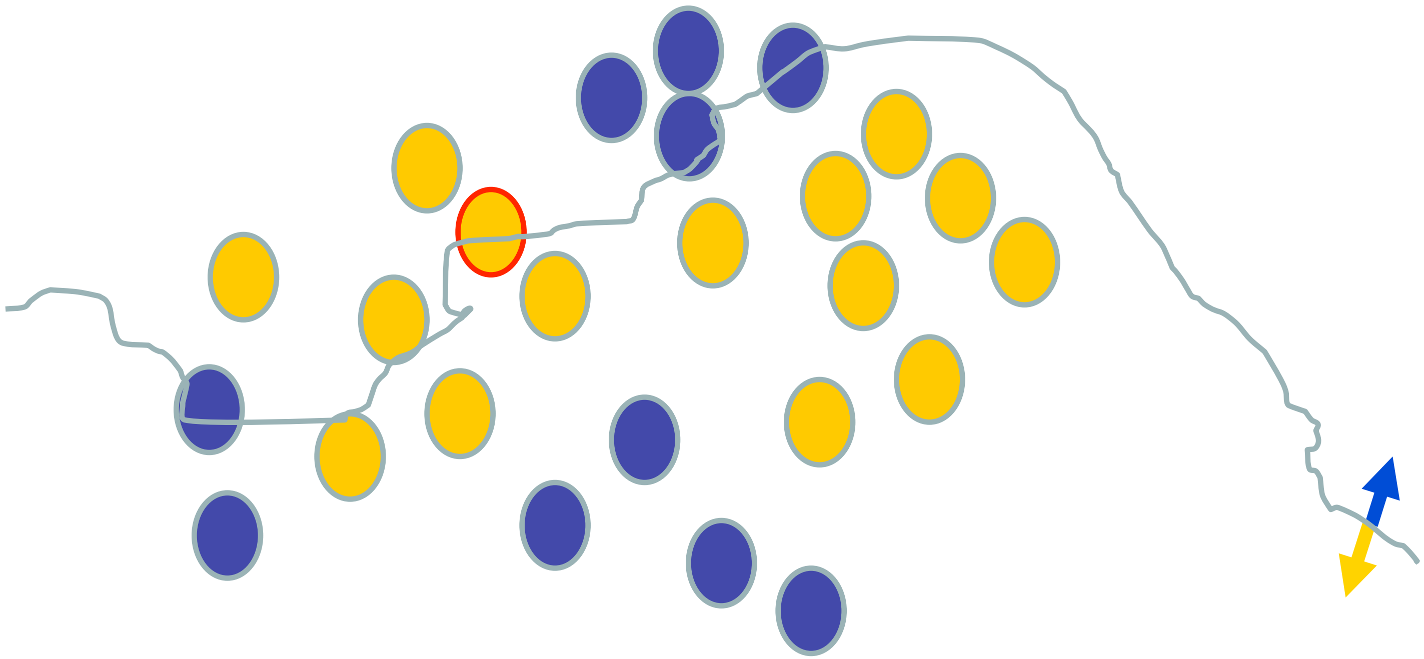
The decision boundary perspective…

Initial random weights



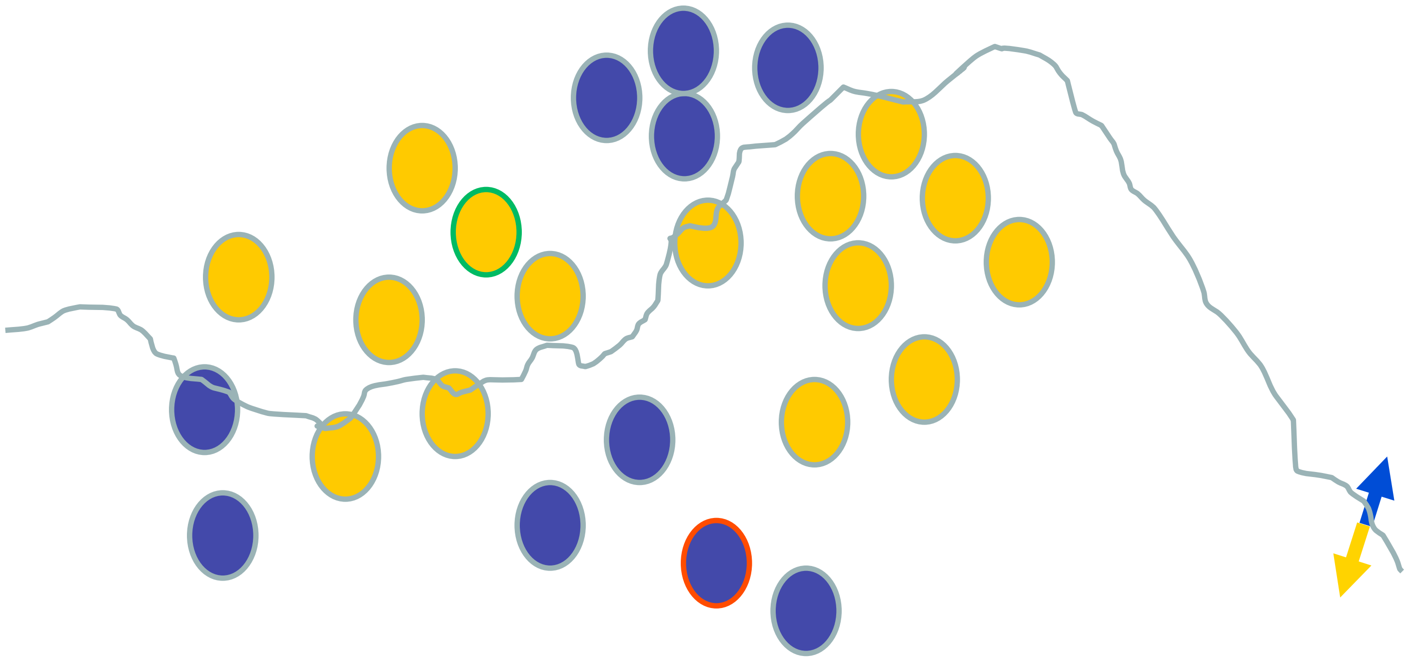
The decision boundary perspective…

Present a training instance / adjust the weights



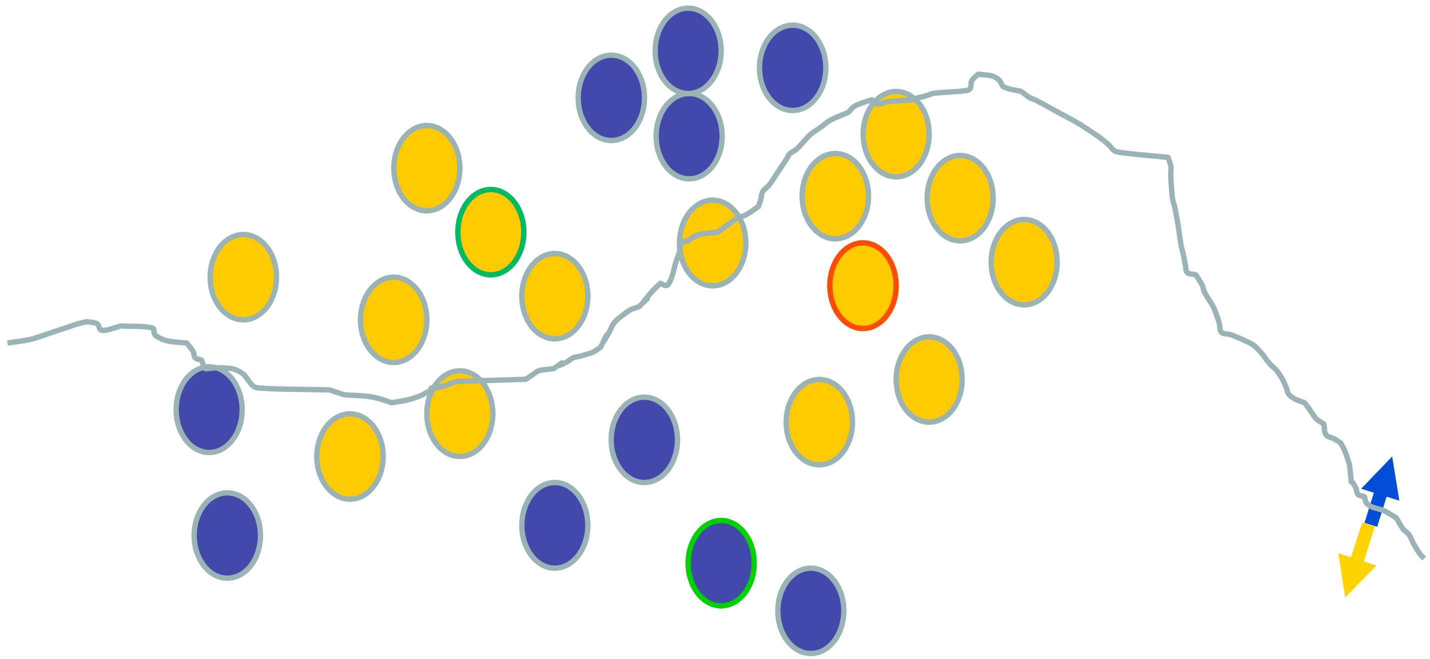
The decision boundary perspective…

Present a training instance / adjust the weights



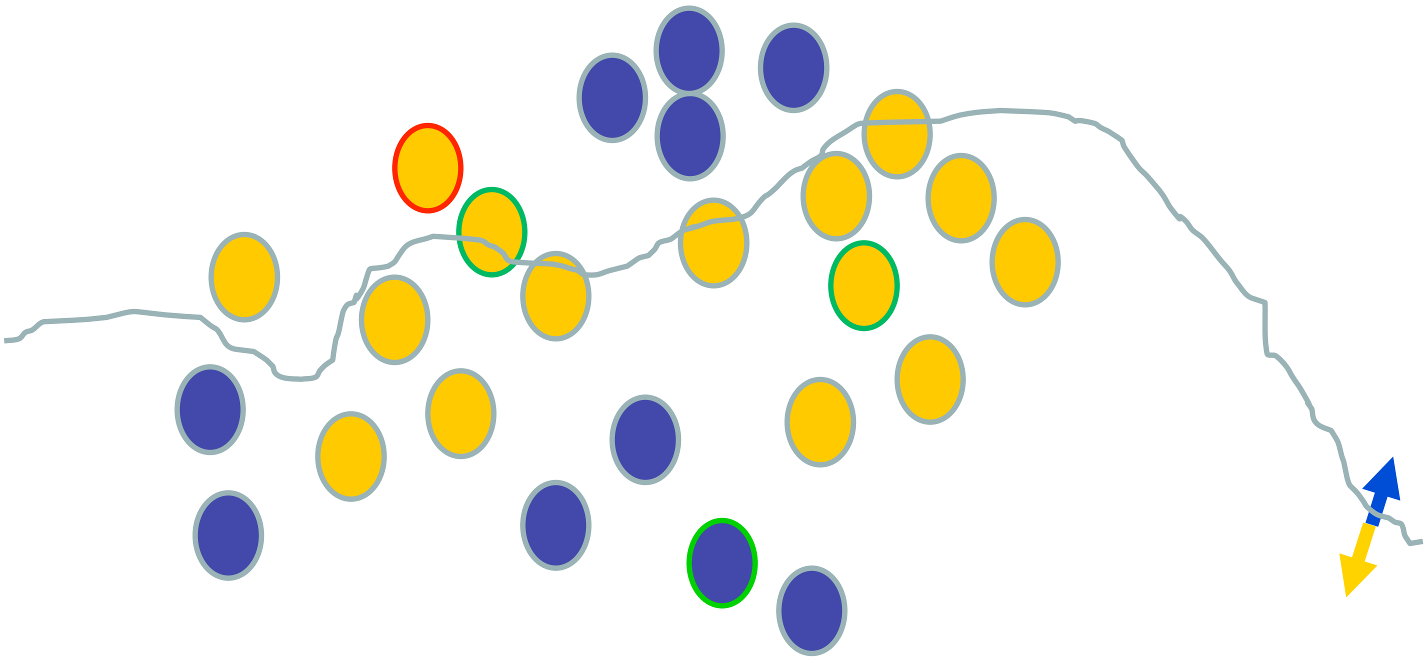
The decision boundary perspective…

Present a training instance / adjust the weights



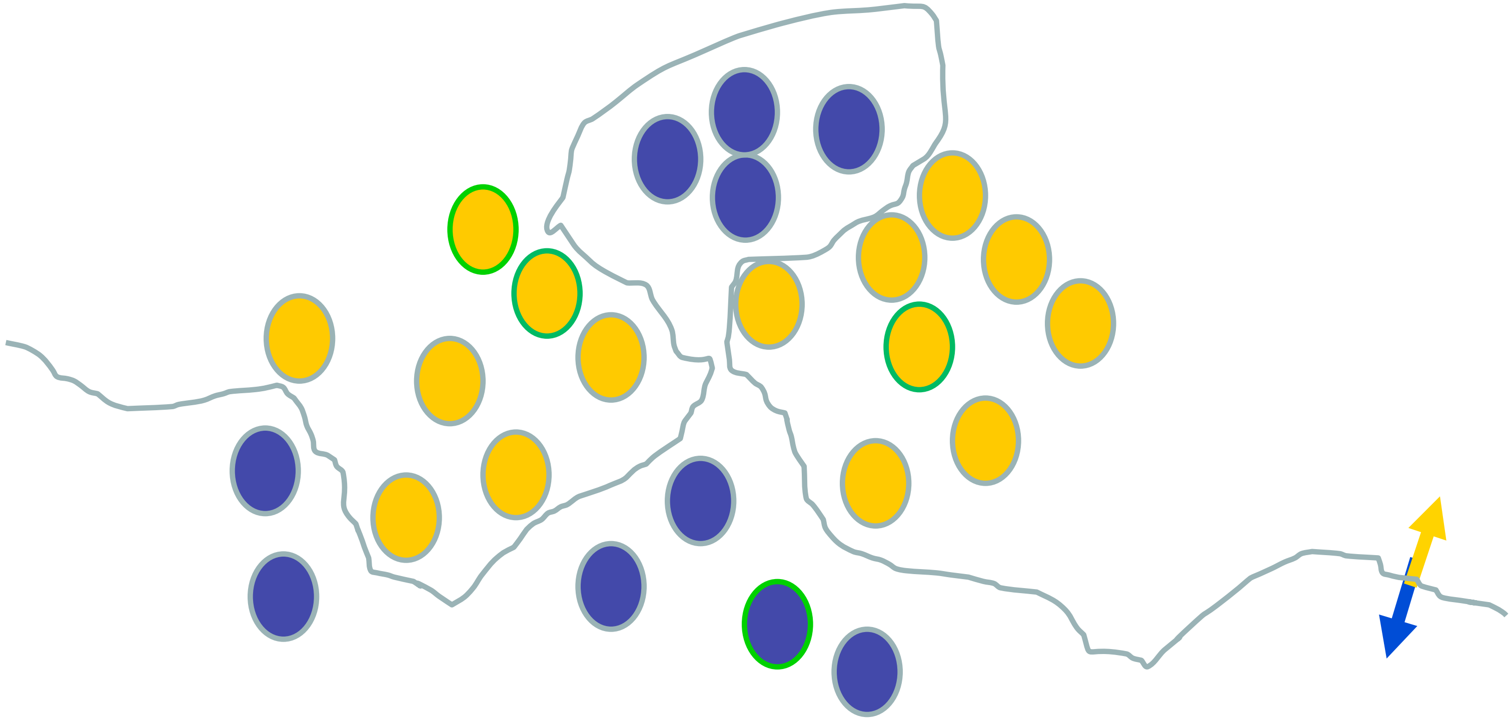
The decision boundary perspective…

Present a training instance / adjust the weights



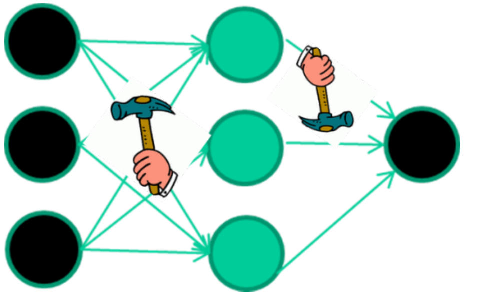
The decision boundary perspective…

Eventually ….



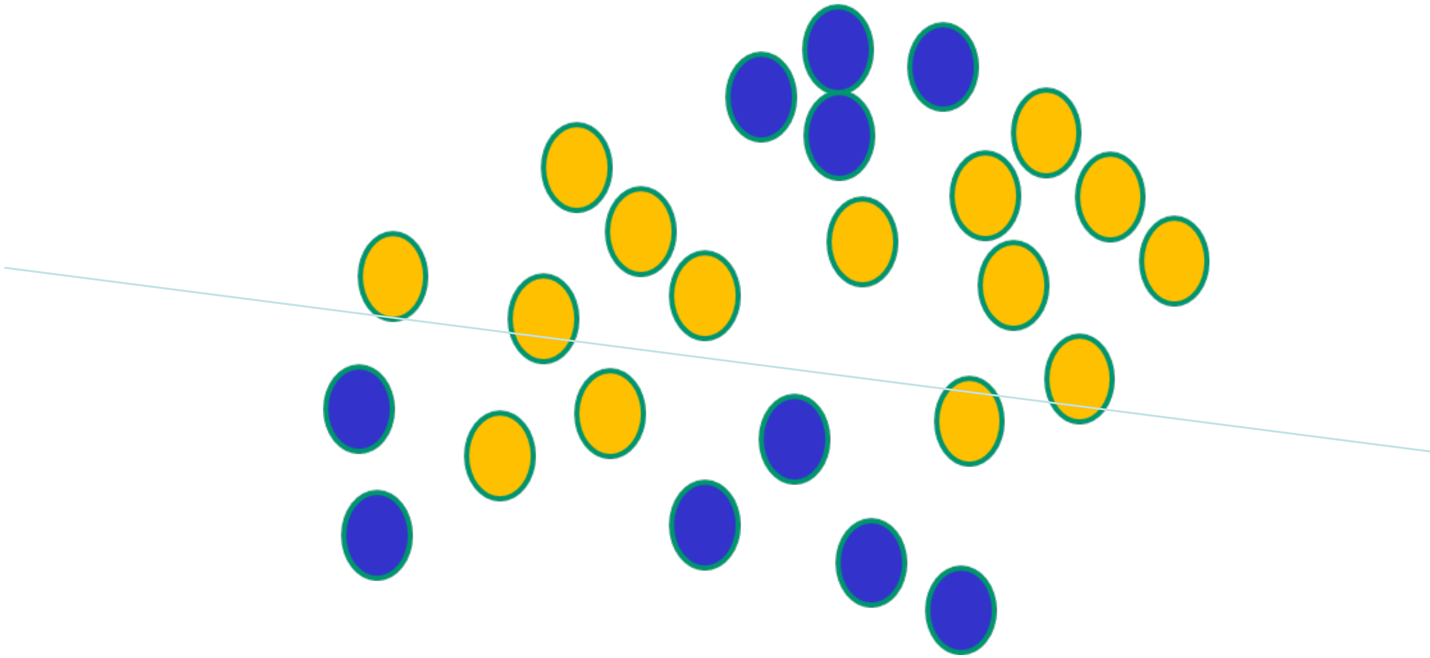
The point I am trying to make

- 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



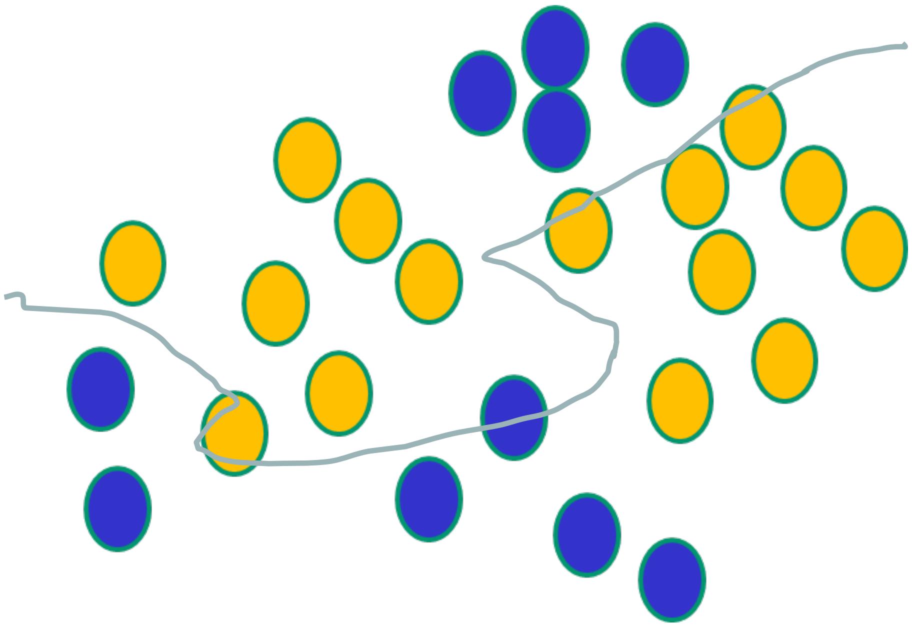
Some other 'by the way' points

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

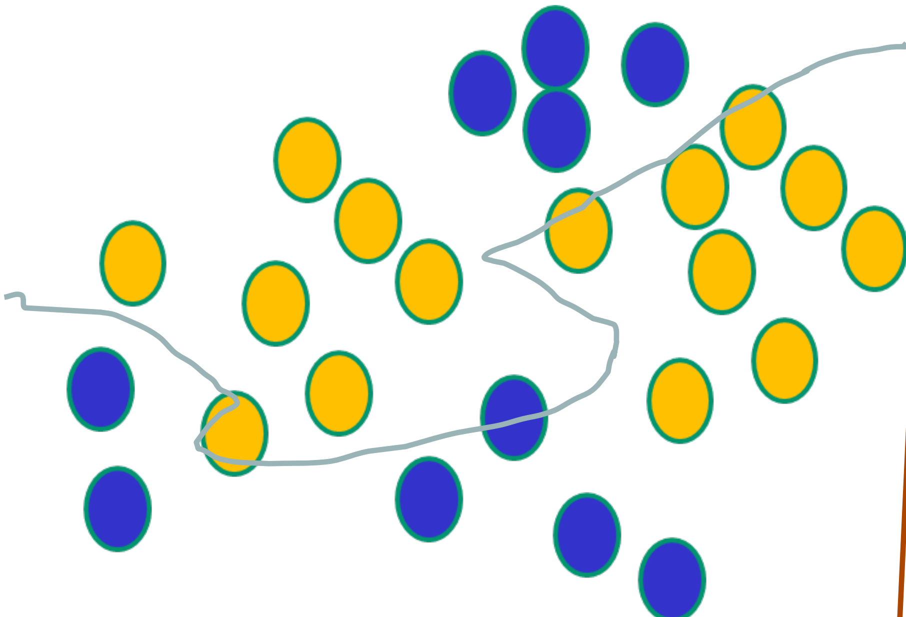
NNs use nonlinear *f*(x) so they can draw complex boundaries, but keep the data unchanged

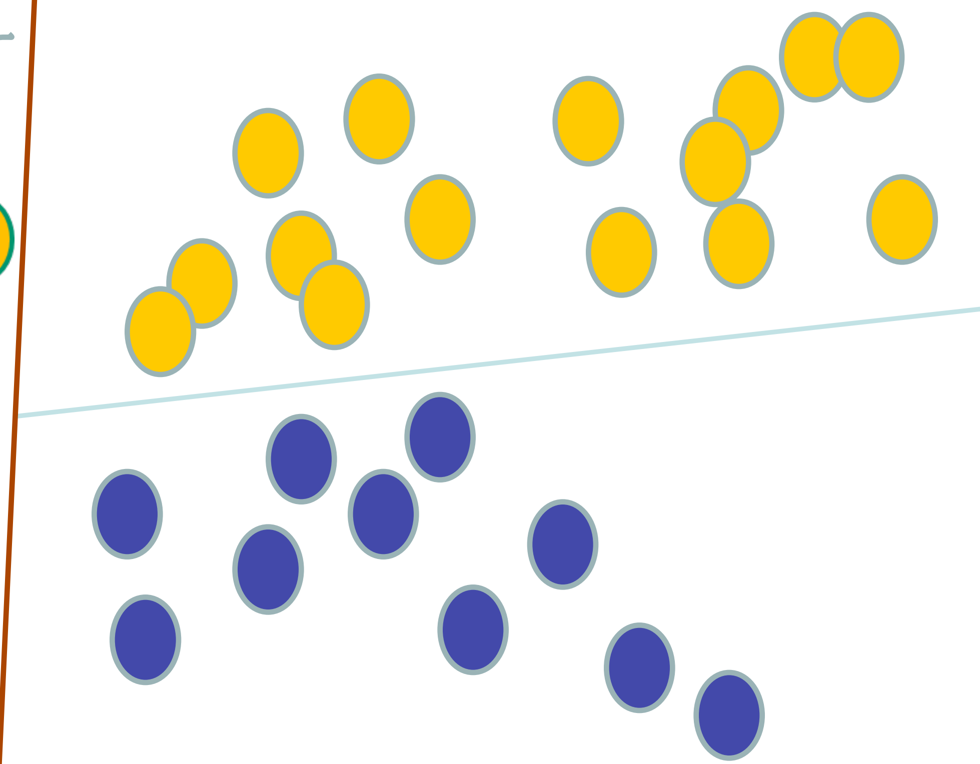


Some other 'by the way' points

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





Feature detectors

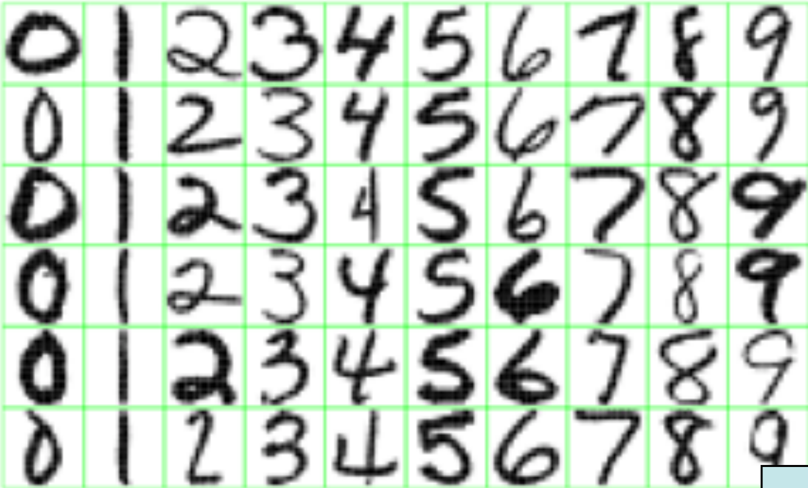
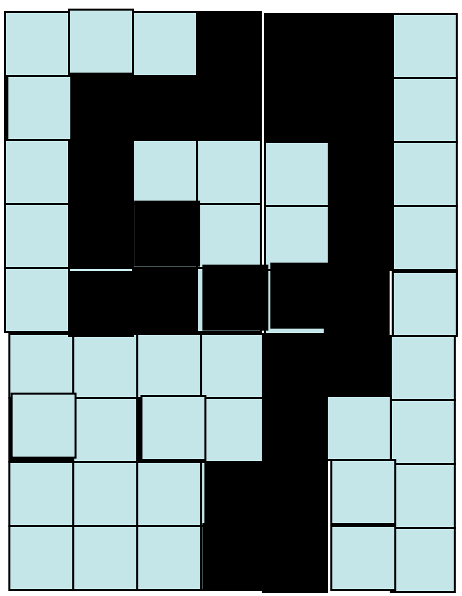
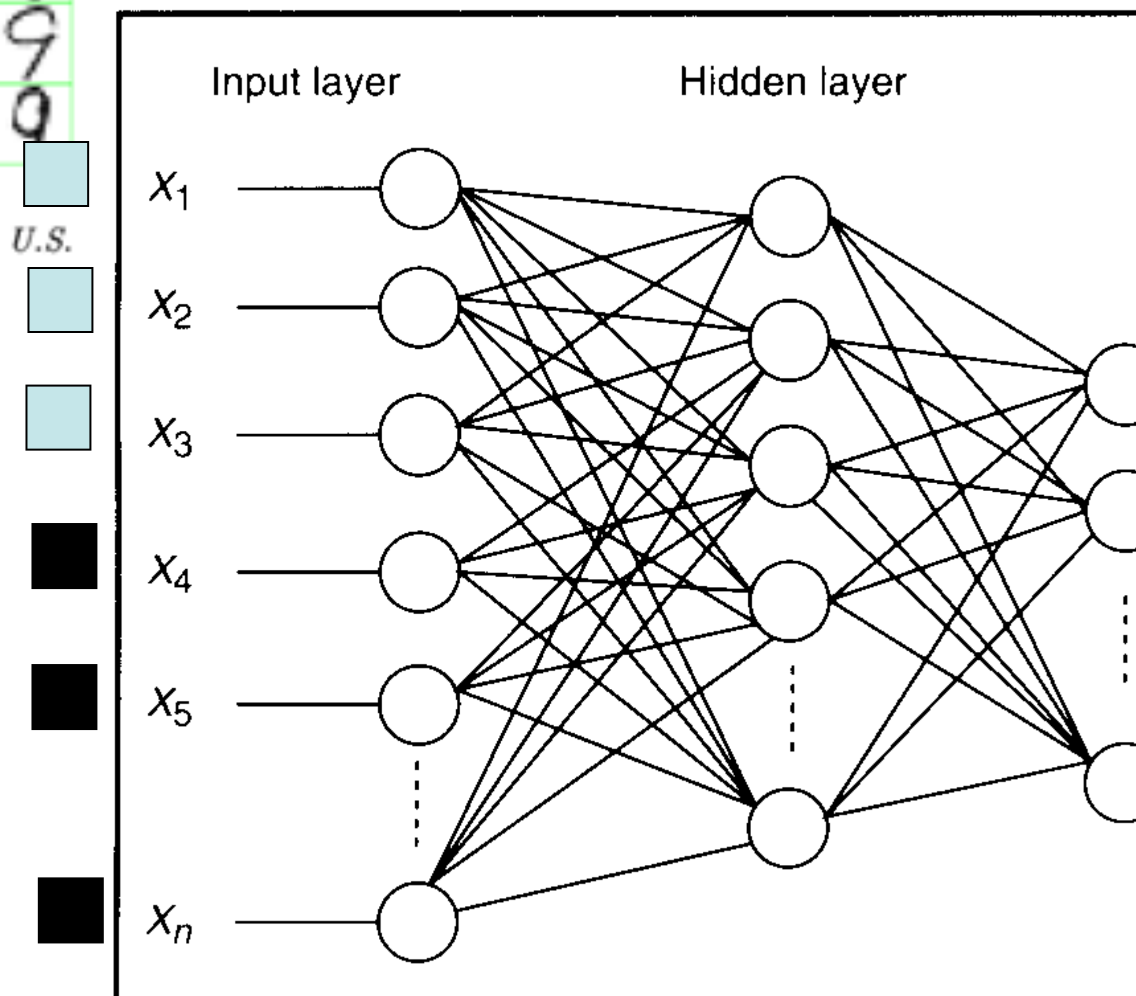


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*





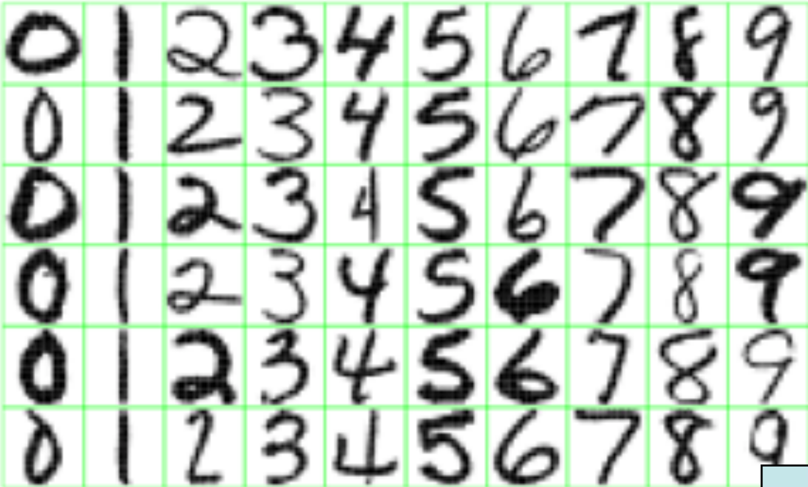
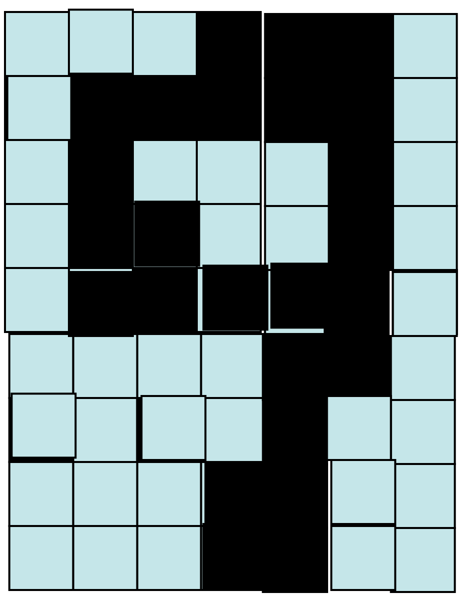
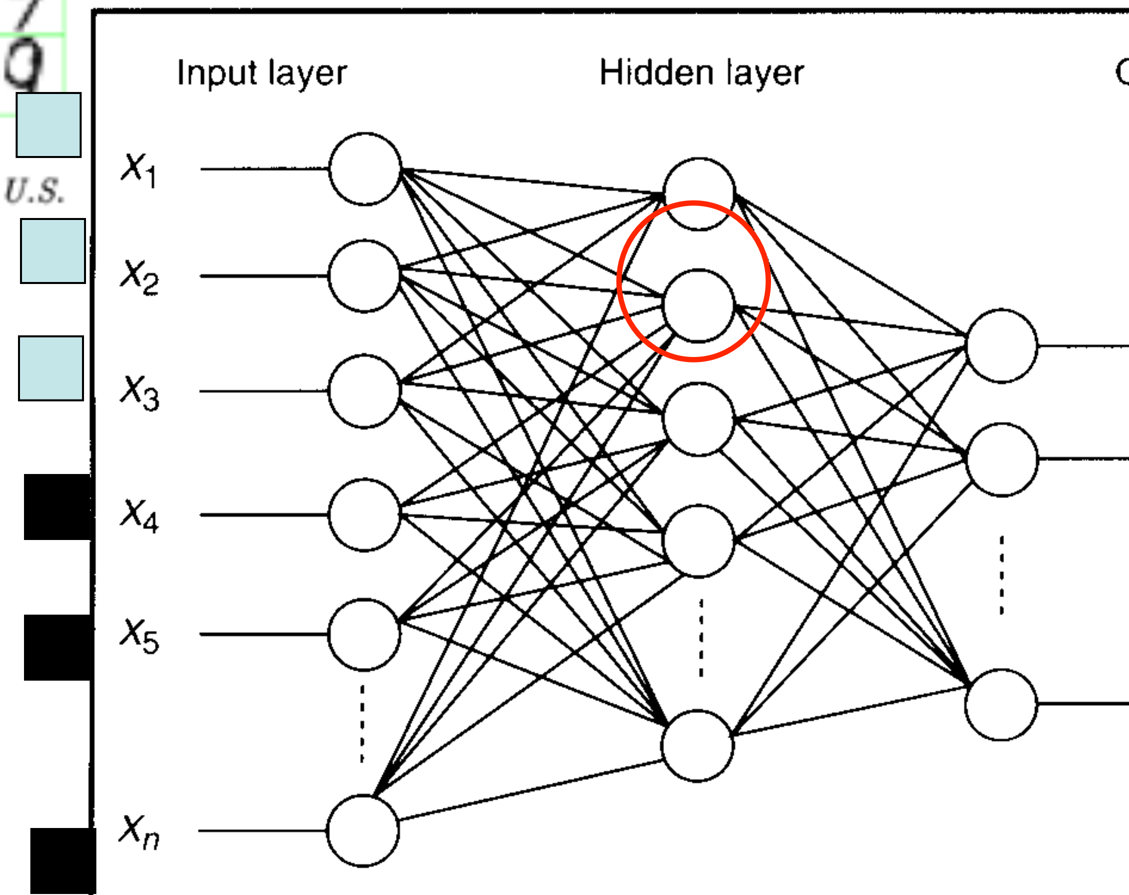


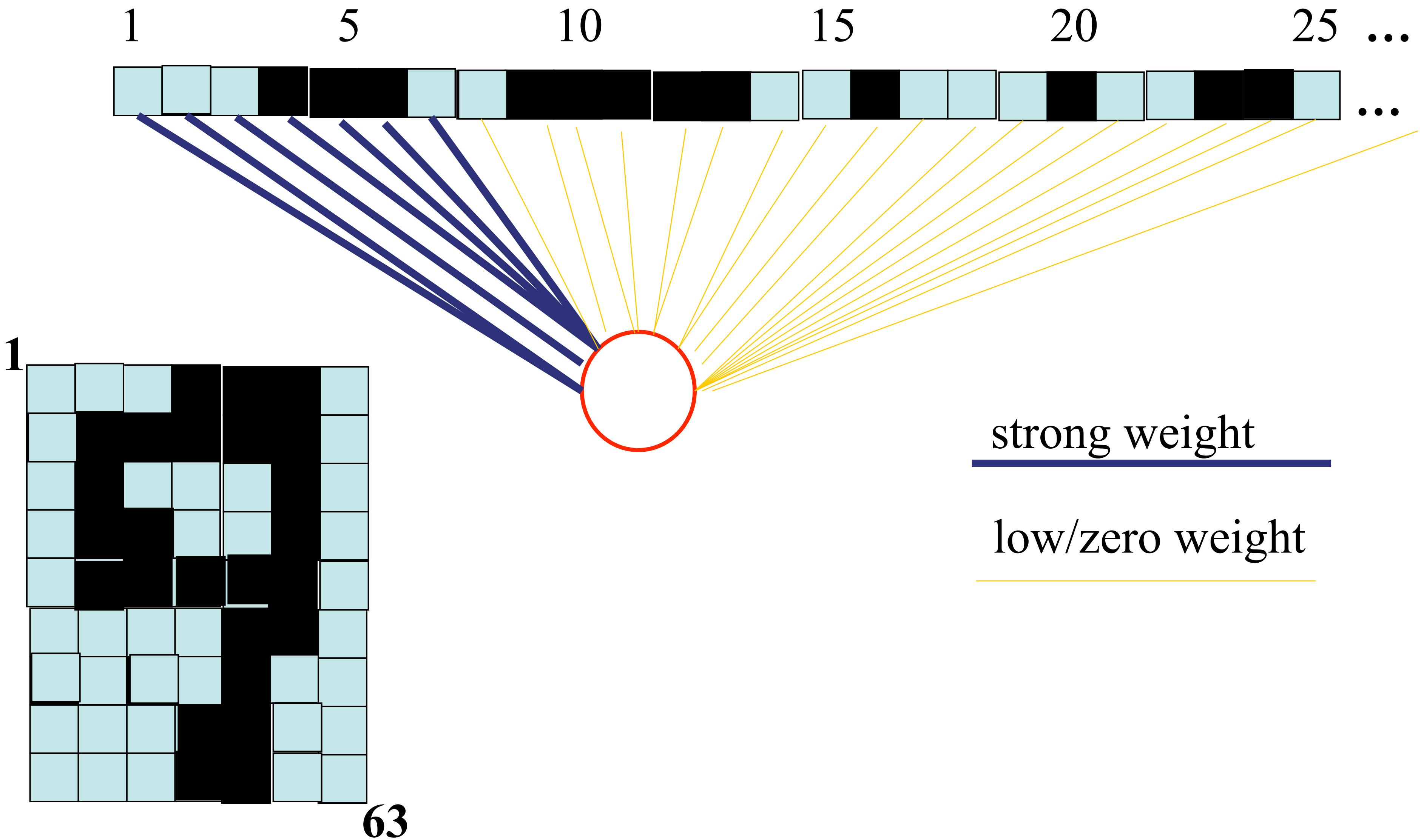
Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*



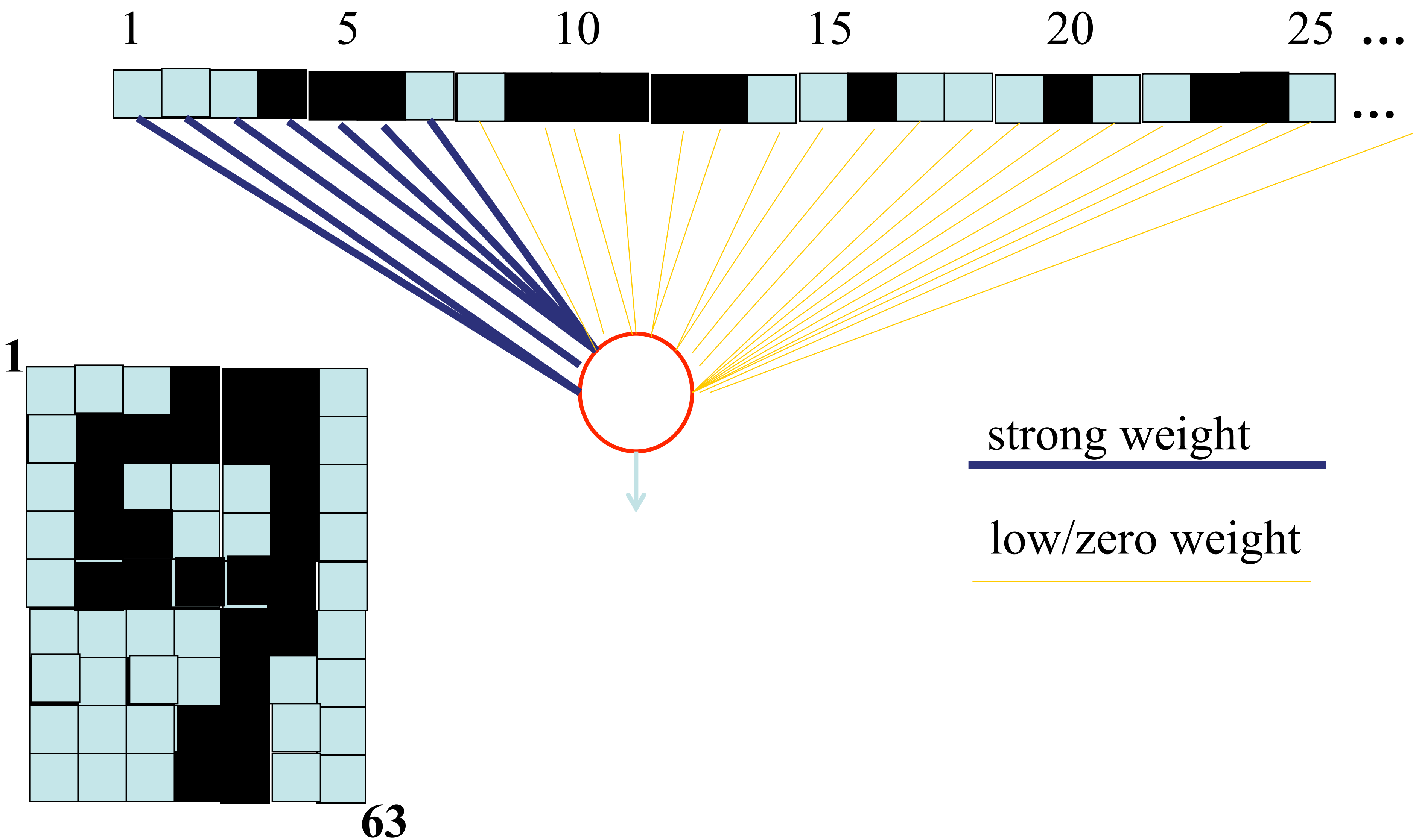
what is this unit doing?



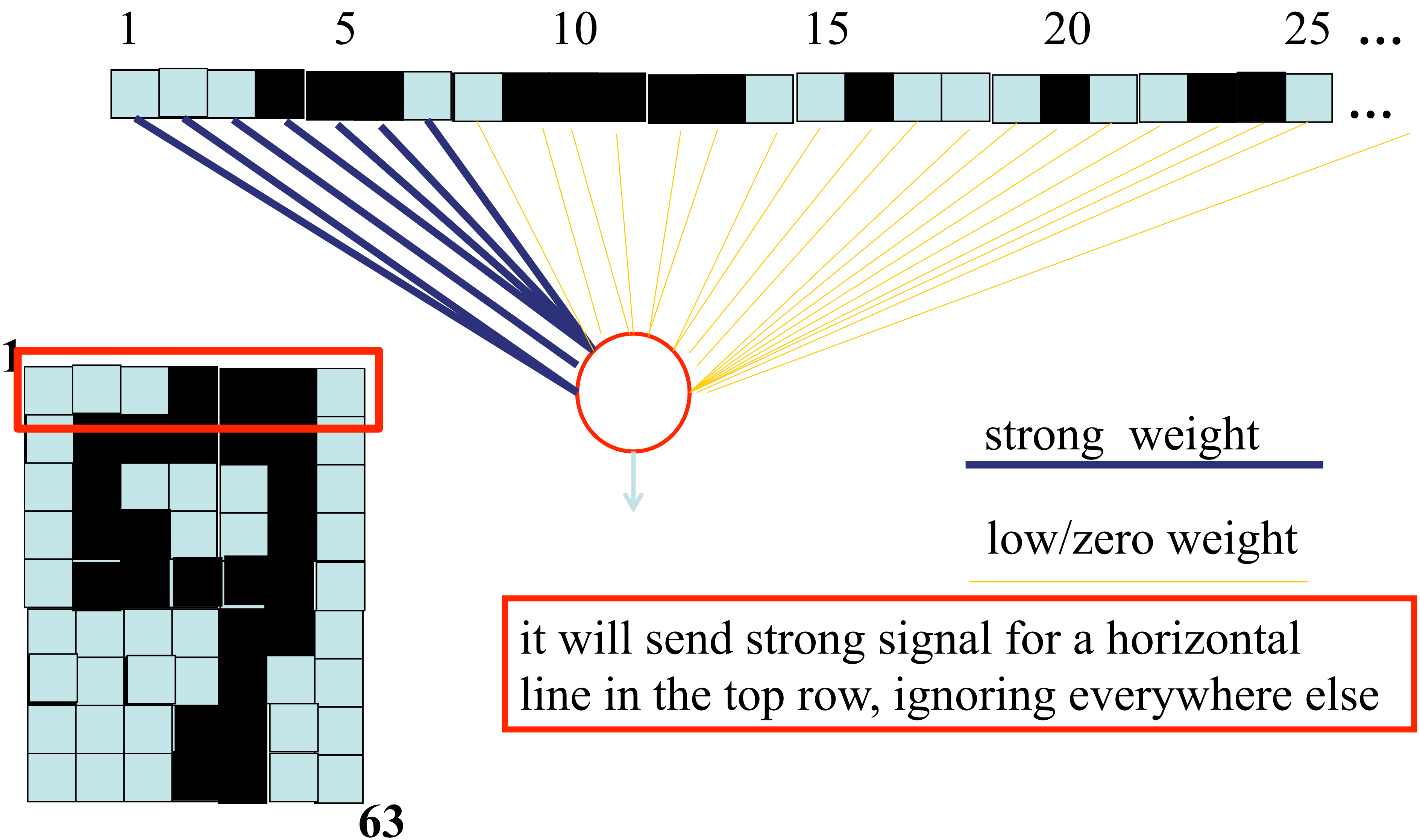
Hidden layer units become *self-organised feature detectors*



What does this unit detect?

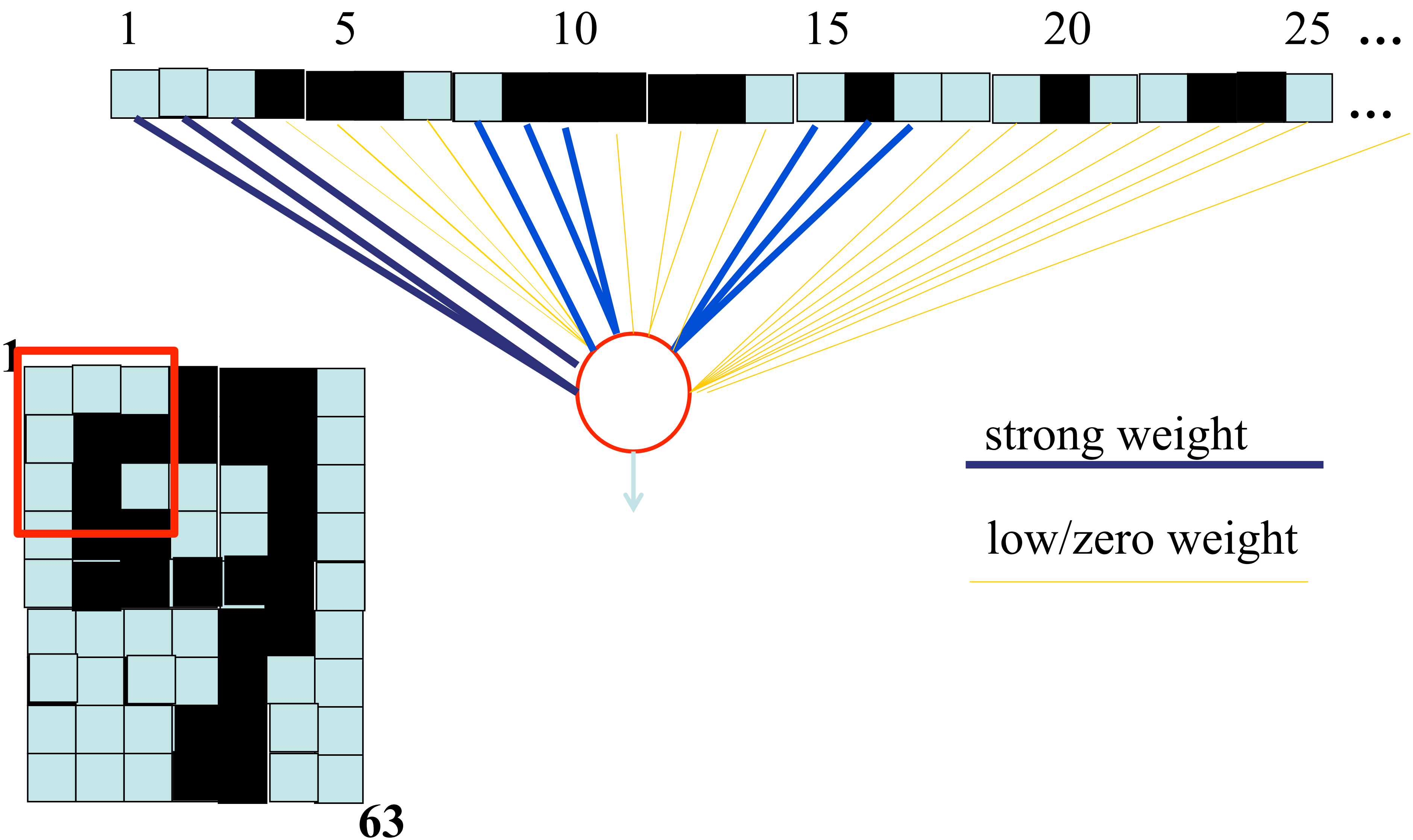


What does this unit detect?

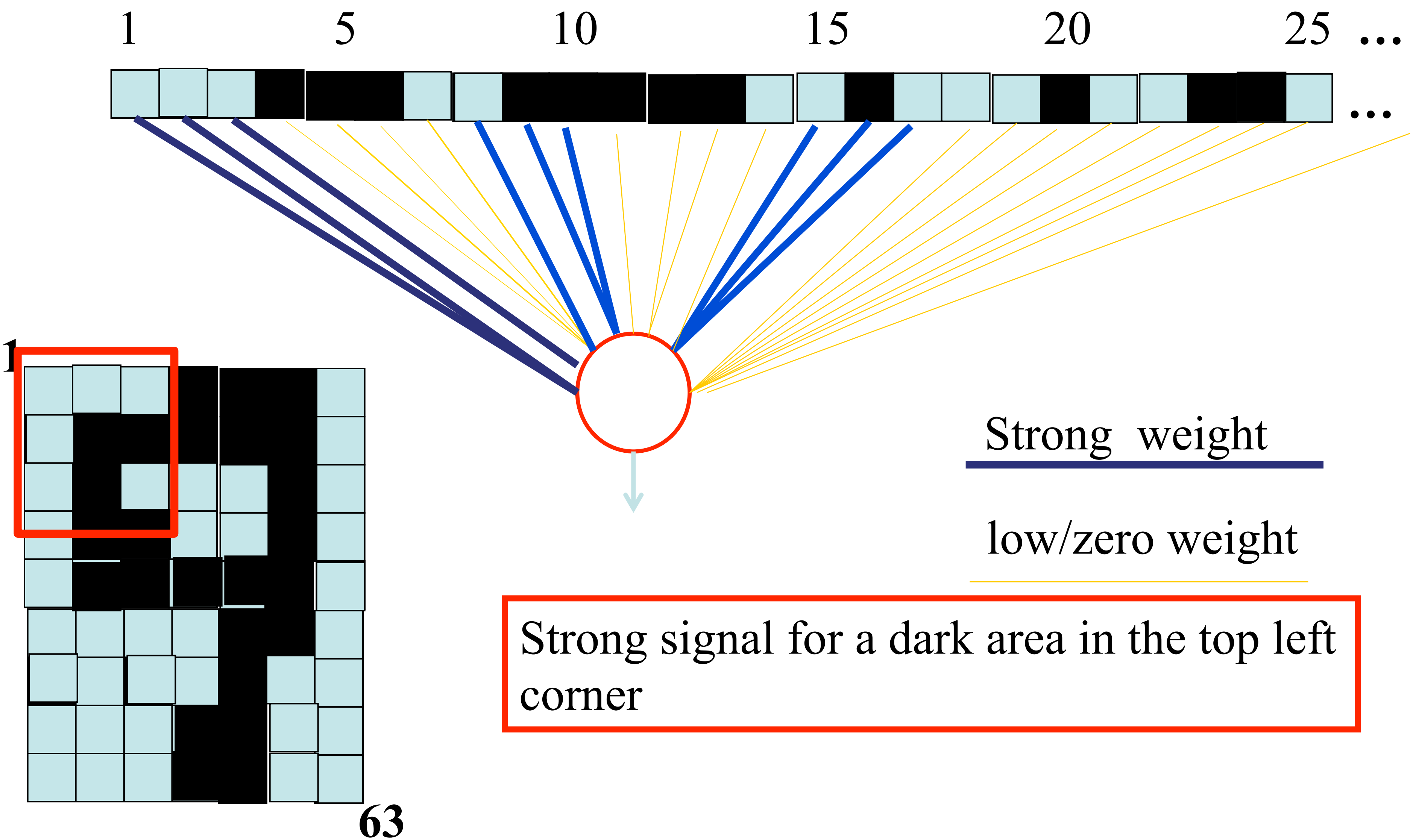


it will send strong signal for a horizontal line in the top row, ignoring everywhere else

What does this unit detect?



What does this unit detect?



Strong signal for a dark area in the top left corner

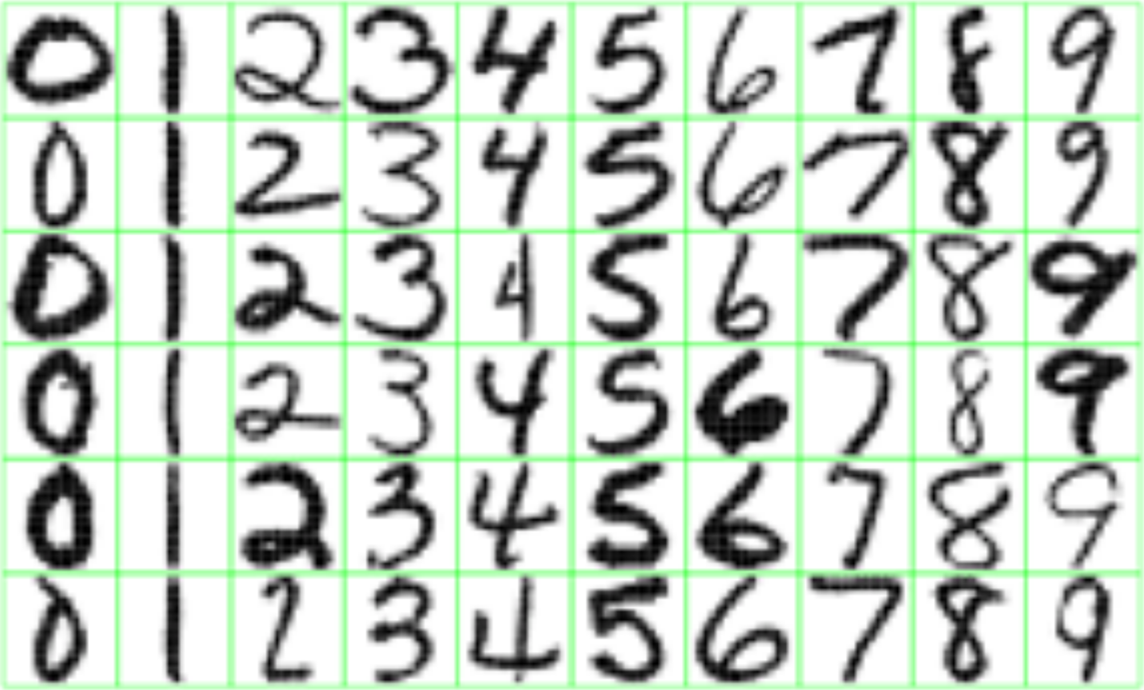


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

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

1

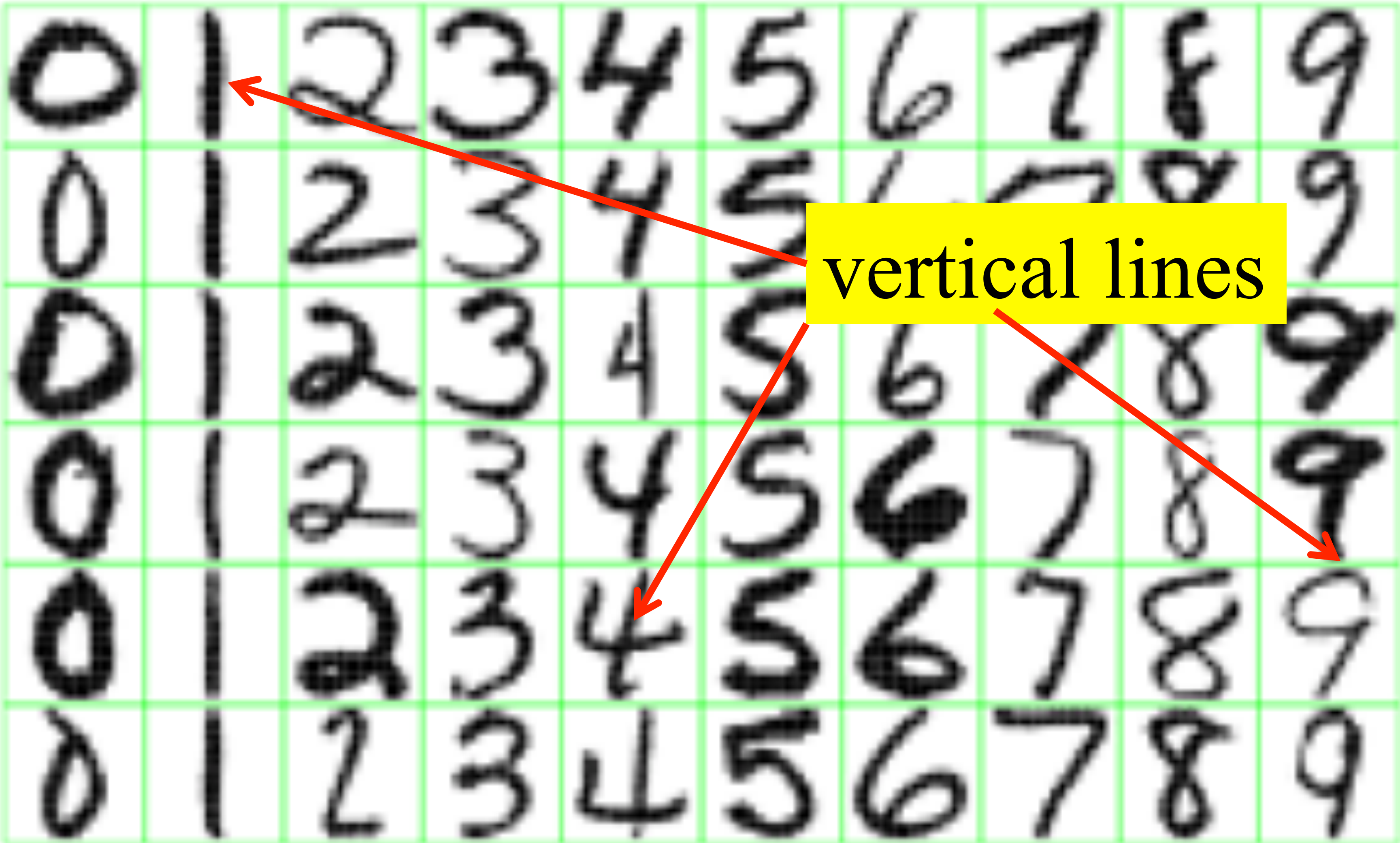


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

1

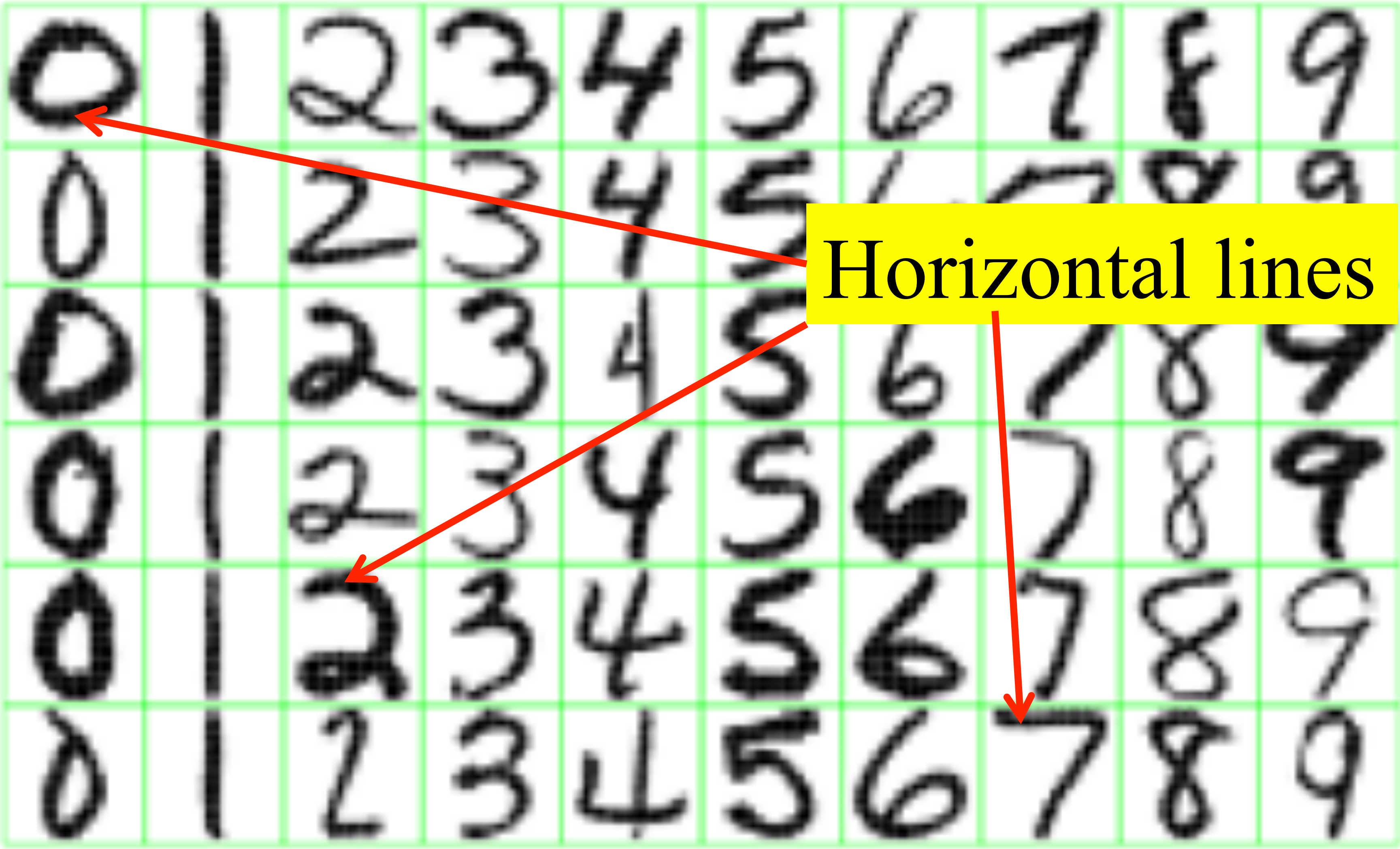


Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*

1

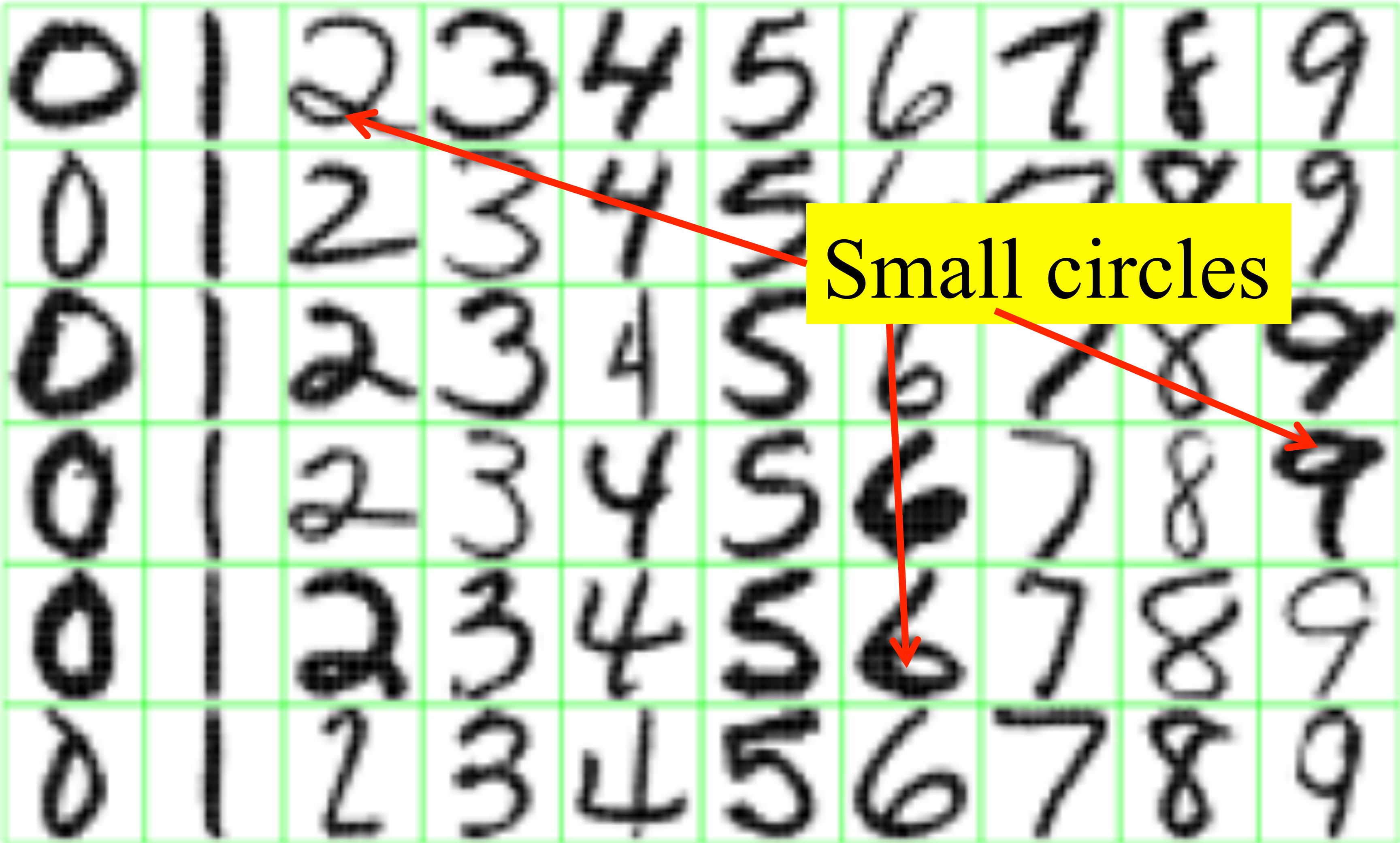
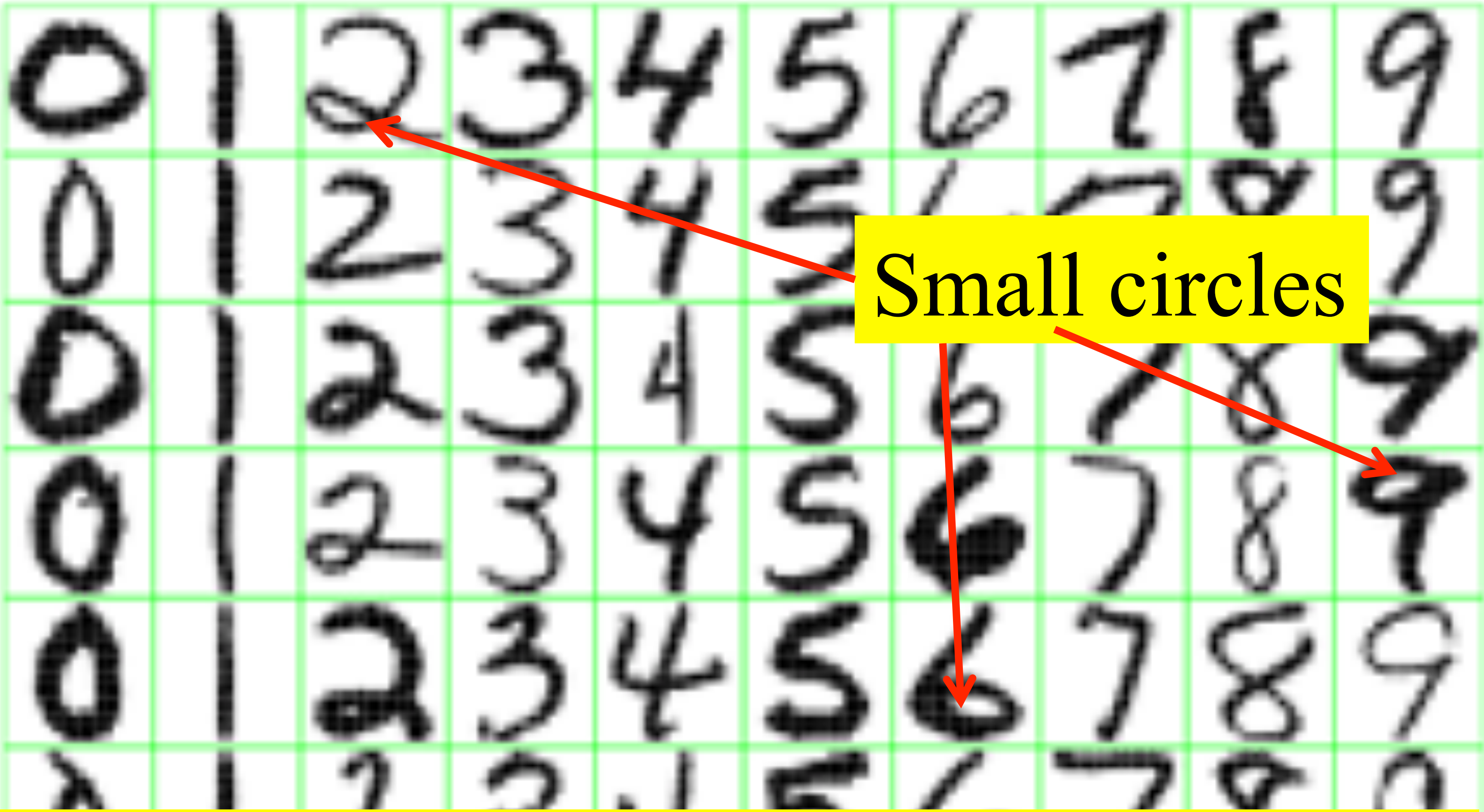


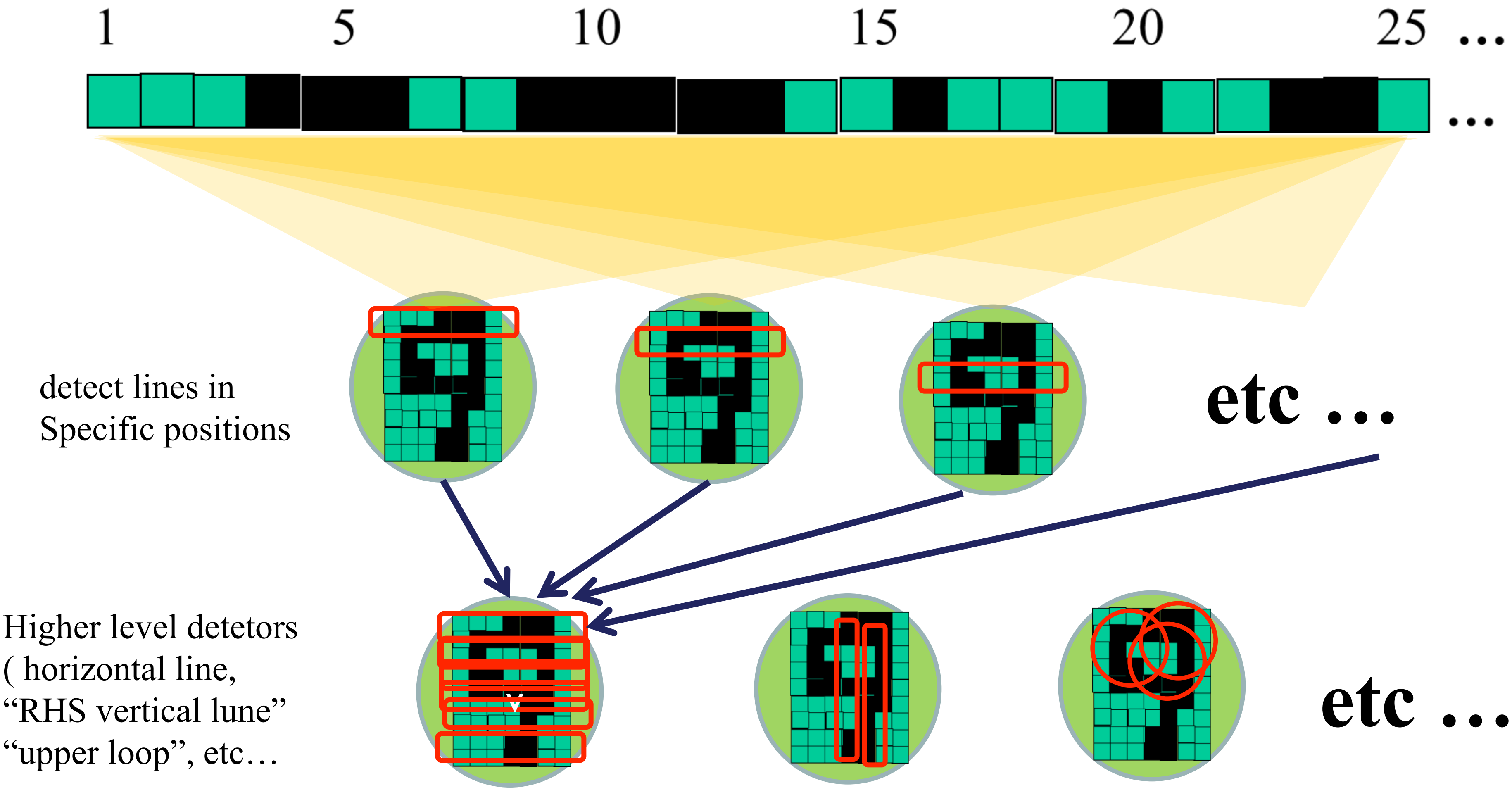
Figure 1.2: *Examples of handwritten digits from U.S. postal envelopes.*



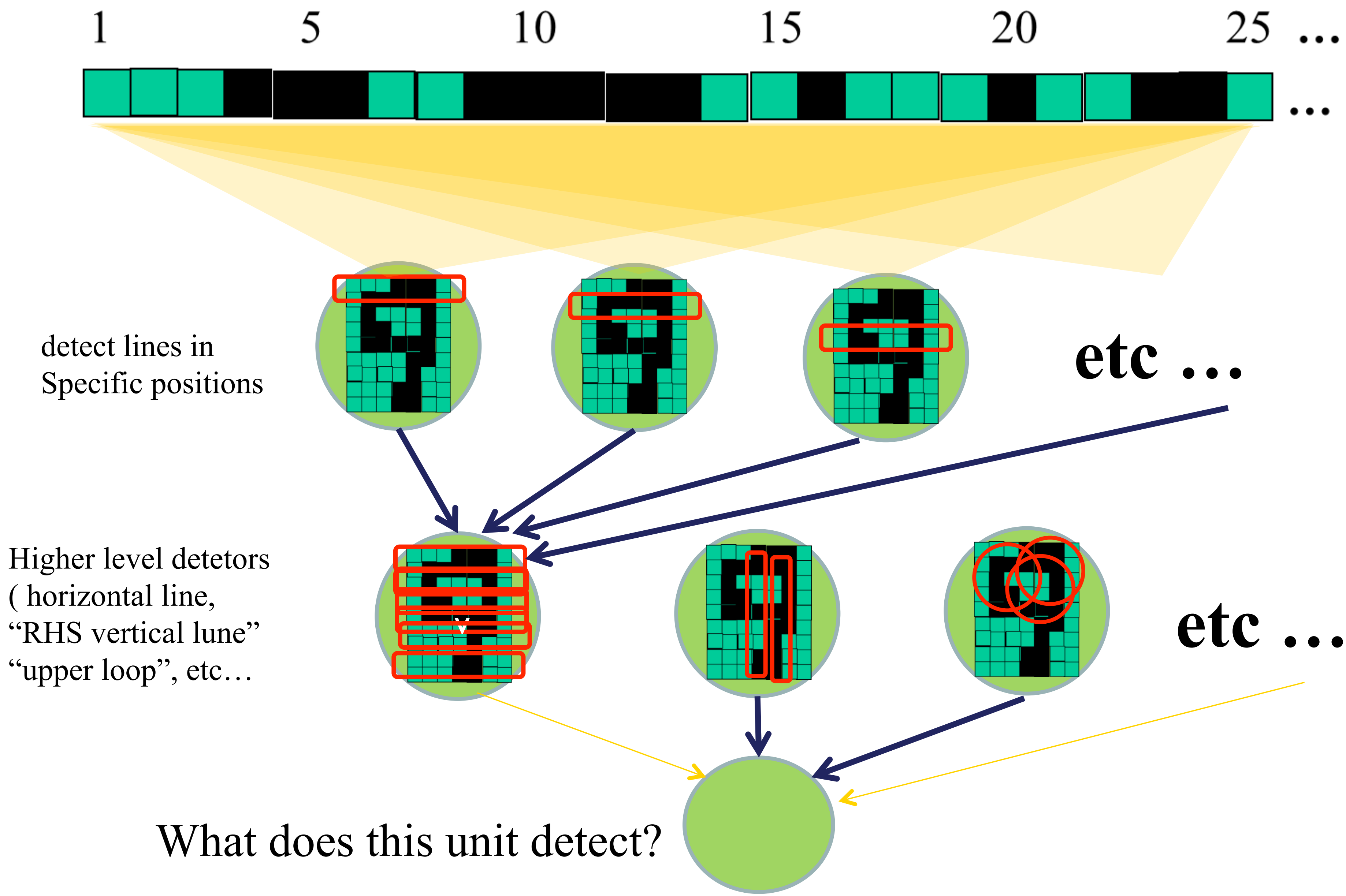
1

But what about position invariance ???
our example unit detectors were tied to
specific parts of the image

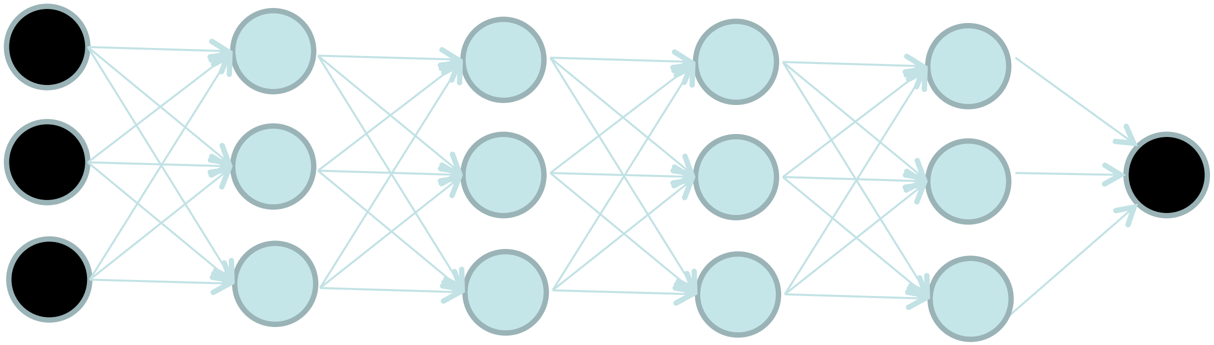
successive layers can learn higher-level features …



successive layers can learn higher-level features …

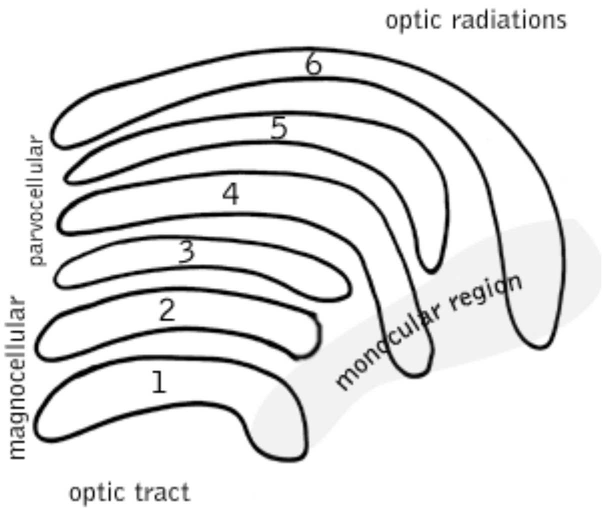


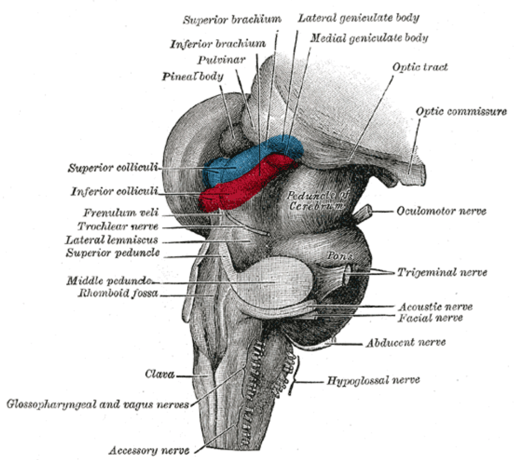
So: *multiple layers make sense*



So: *multiple layers make sense*

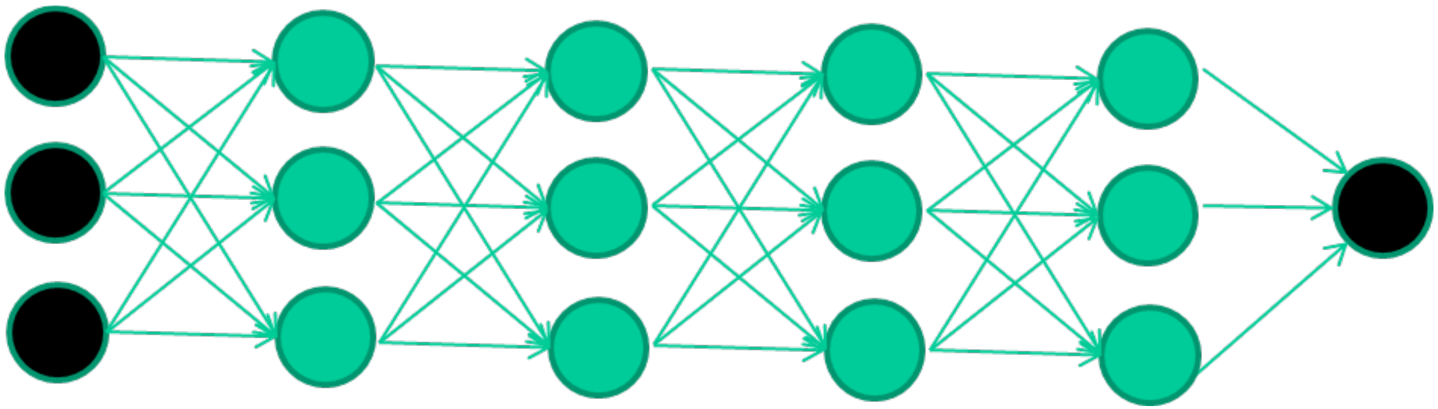
Your brain works that way



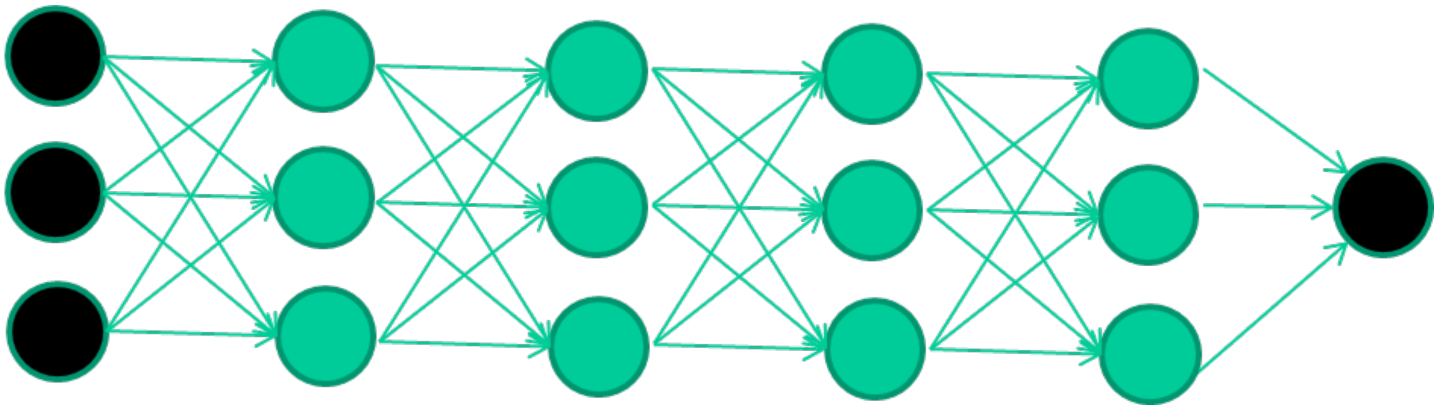


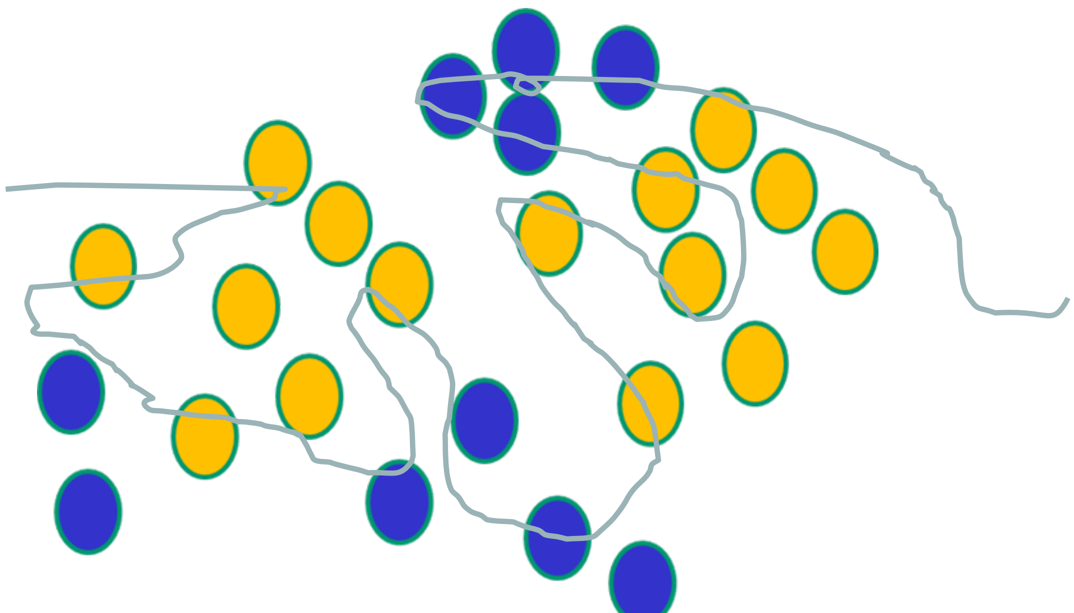
So: *multiple layers make sense*

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



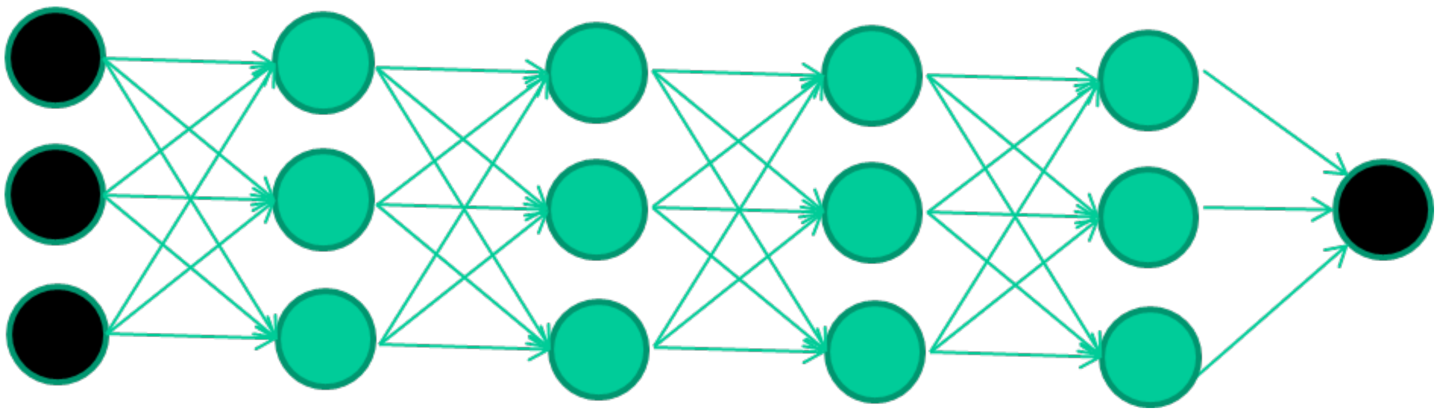
But, until very recently, our weight-learning algorithms simply did not work on multi-layer architectures



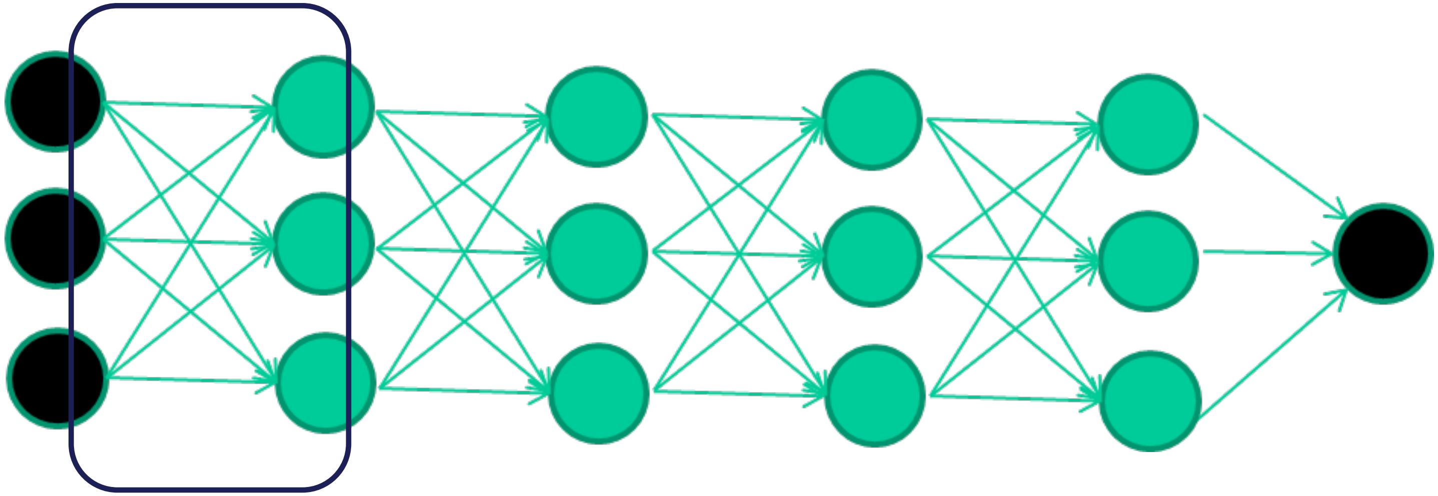


Along came deep learning …

The new way to train multi-layer NNs…

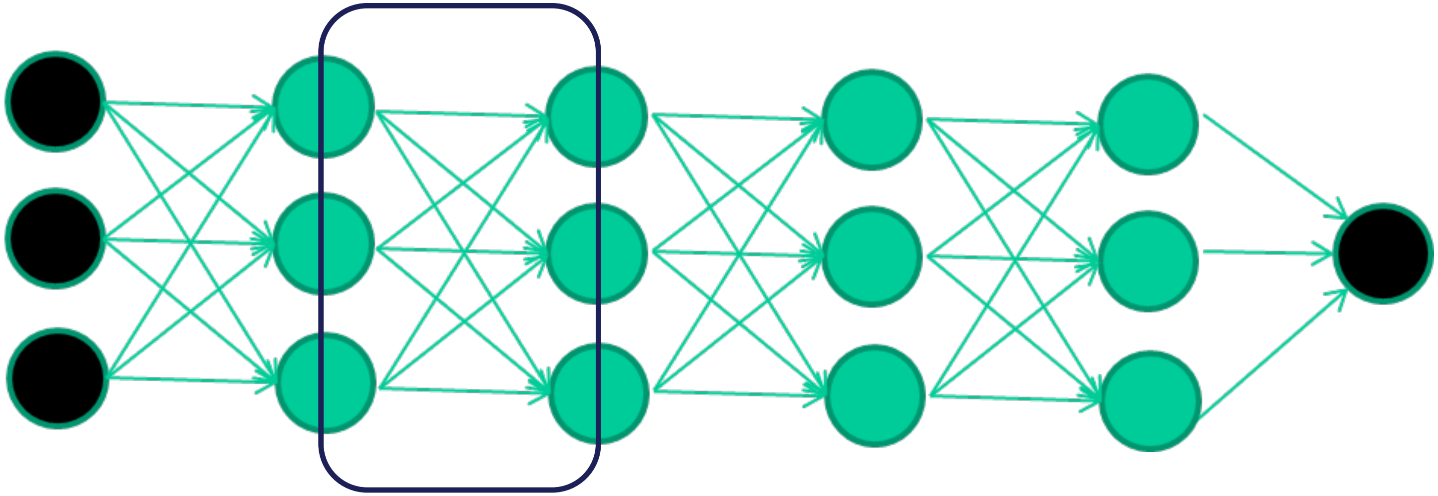


The new way to train multi-layer NNs…



Train **this** layer first

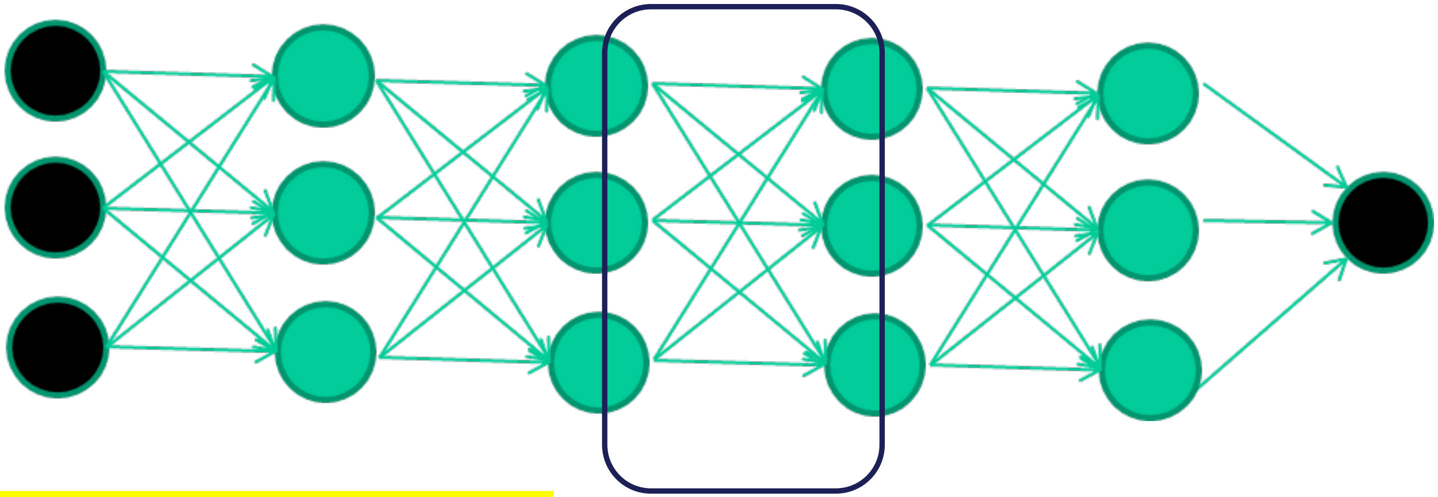
The new way to train multi-layer NNs…



Train **this** layer first

then **this** layer

The new way to train multi-layer NNs…

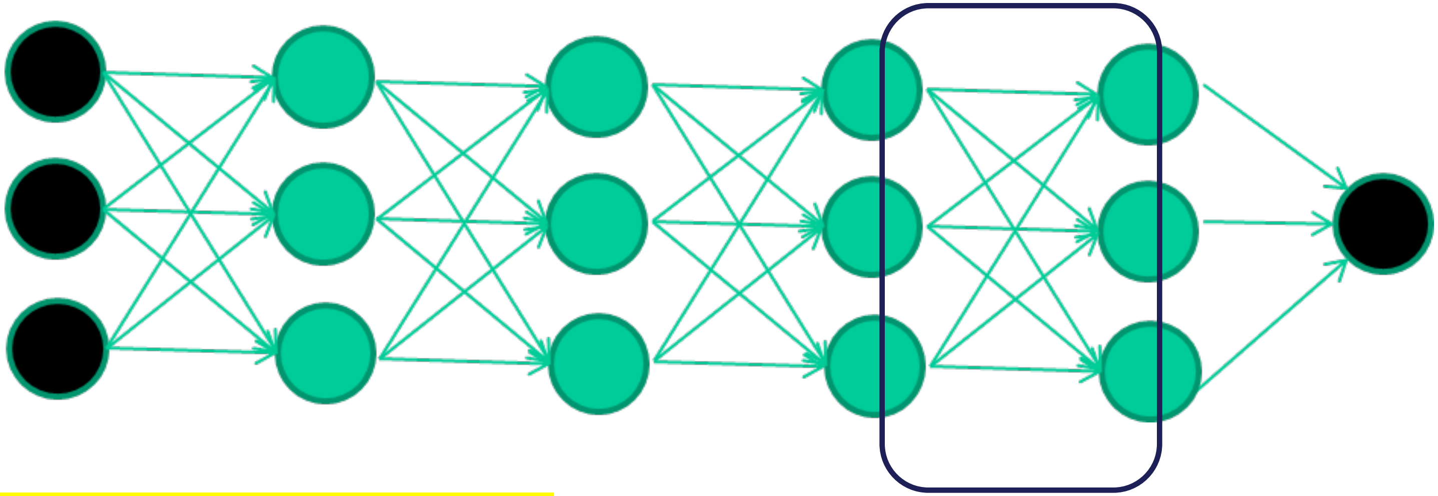


Train **this** layer first

then **this** layer

then **this** layer

The new way to train multi-layer NNs…



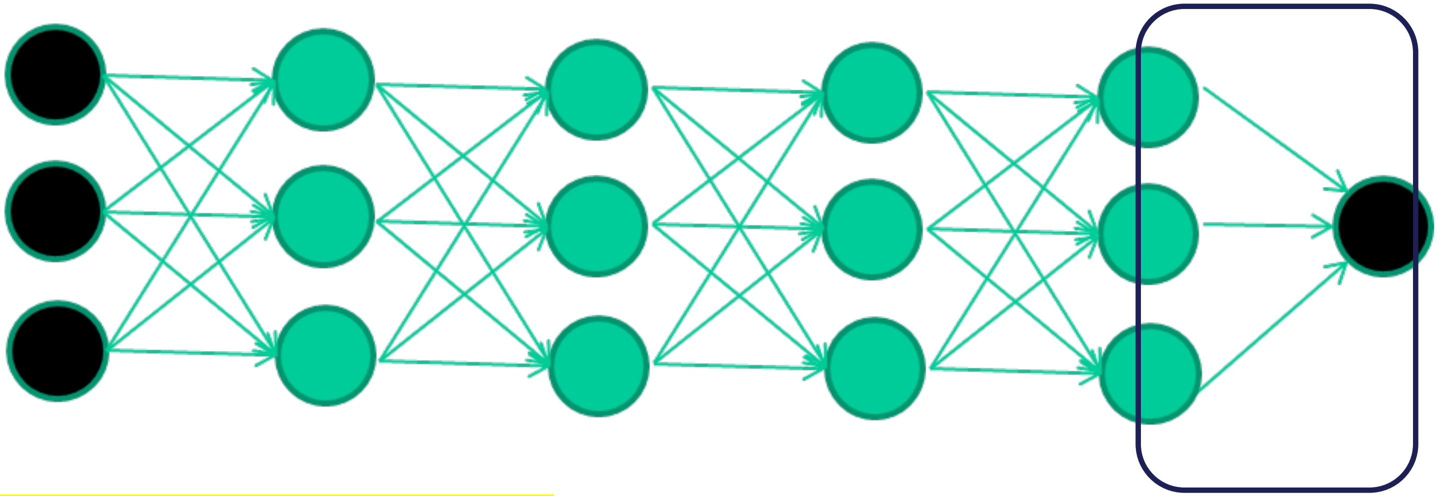
Train **this** layer first

then **this** layer

then **this** layer

then **this** layer

The new way to train multi-layer NNs…



Train **this** layer first

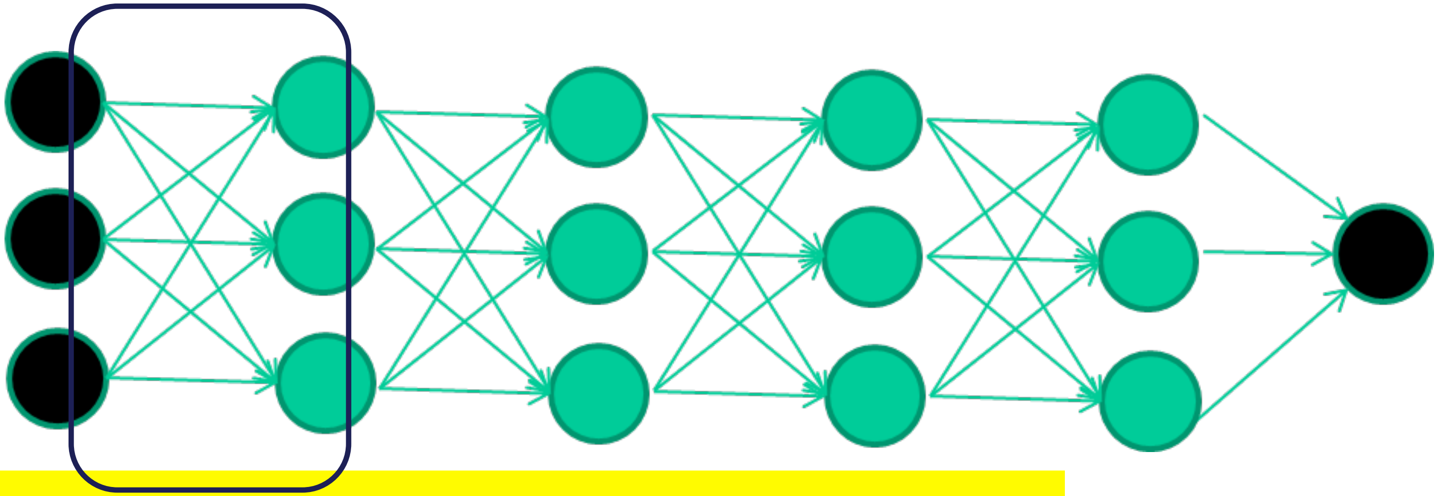
then **this** layer

then **this** layer

then **this** layer

finally **this** layer

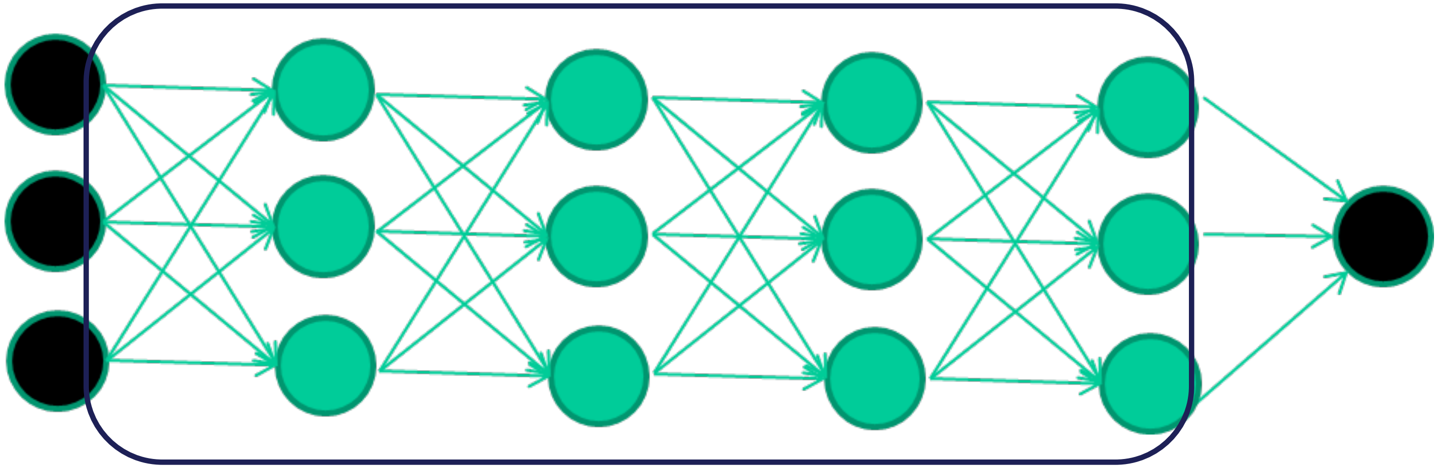
The new way to train multi-layer NNs…



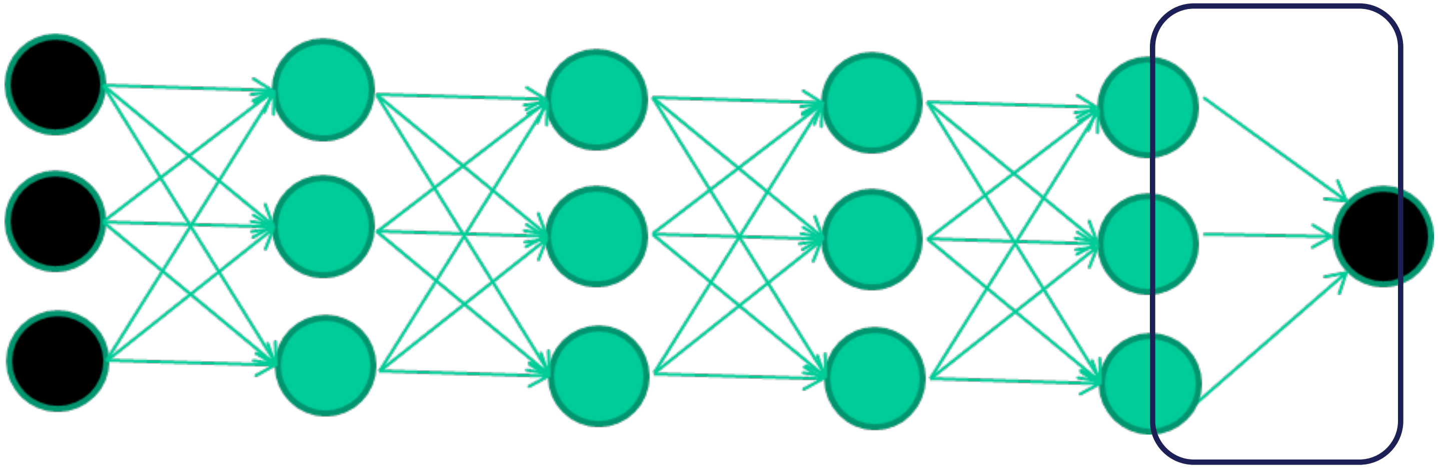
EACH of the (non-output) layers is trained to be an ***auto-encoder***

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

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

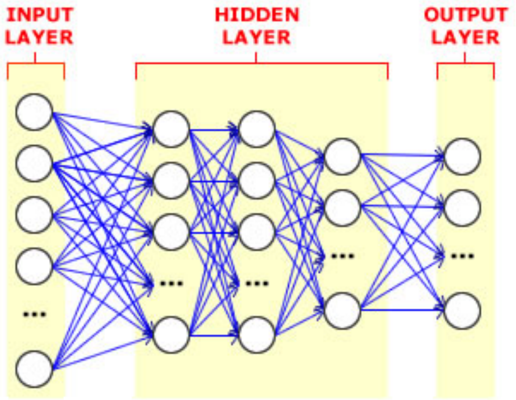


Final layer trained to predict class based on outputs from previous layers

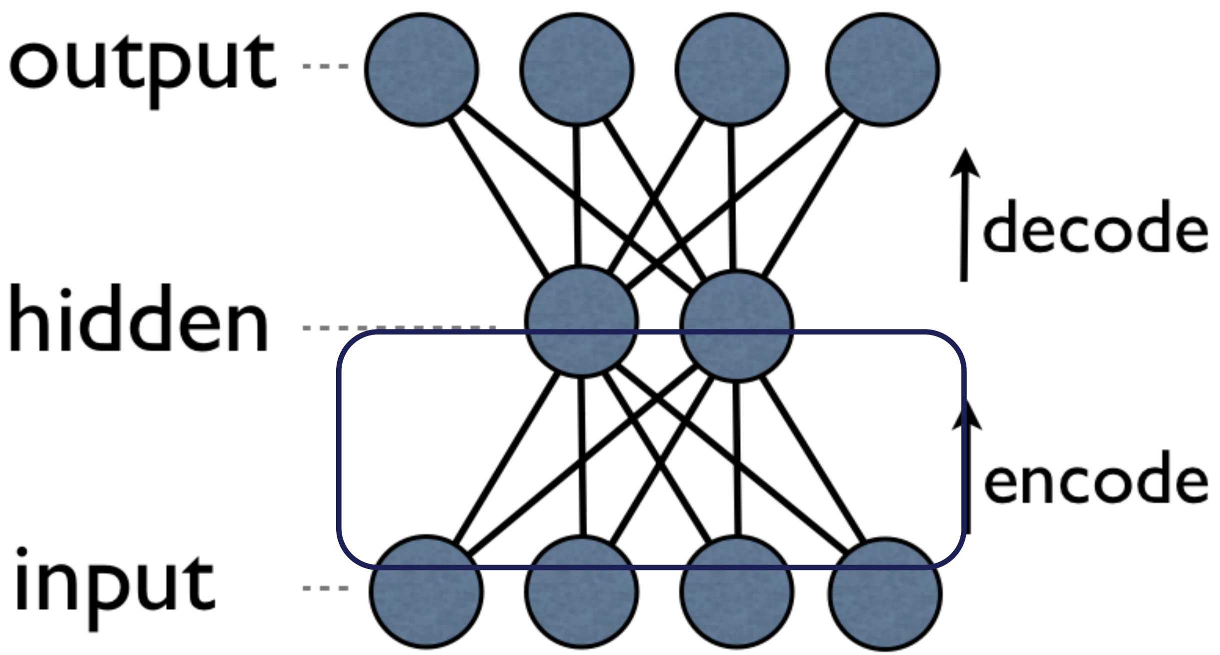


And that's that

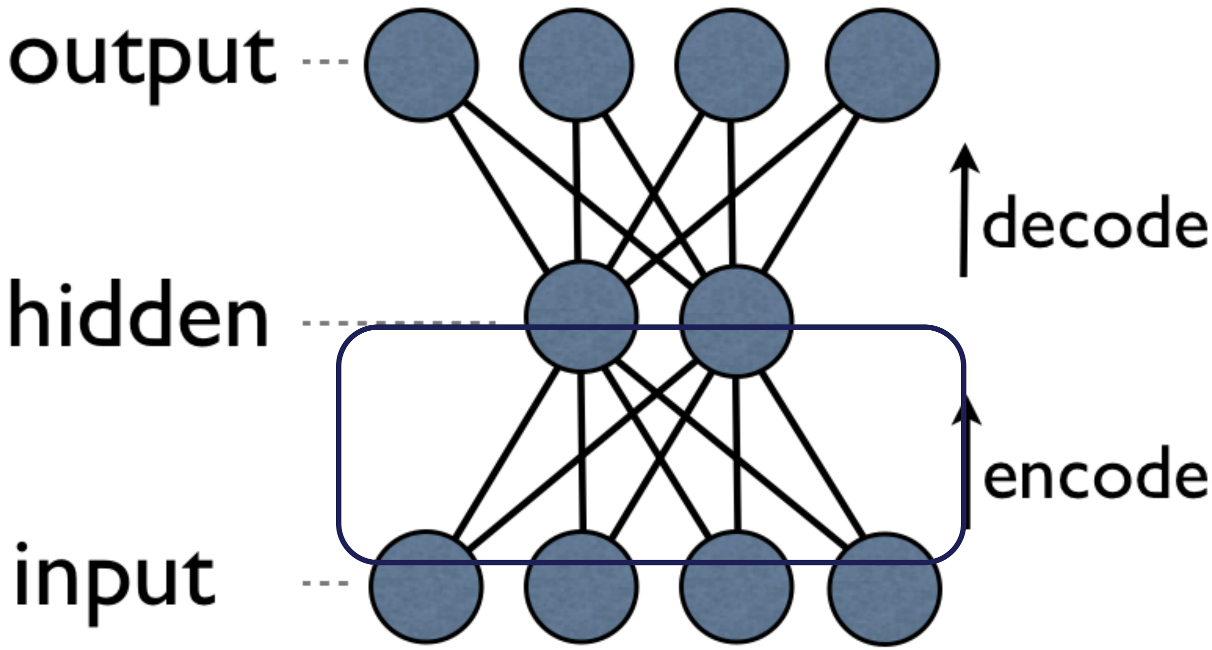
- That's the basic idea
- There are many many types of deep learning,
- different kinds of autoencoder, variations on architectures and training algorithms, etc…
- Very fast growing area …



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



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