DEPARTMENT OF
INFORMATION SYSTEMS

CIS 508 – Data Mining I
Week 5 – Lecture 5.2



ASU W. P. CAREY SCHOOL of BUSINESS
ARIZONA STATE UNIVERSITY

Deep Learning

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.

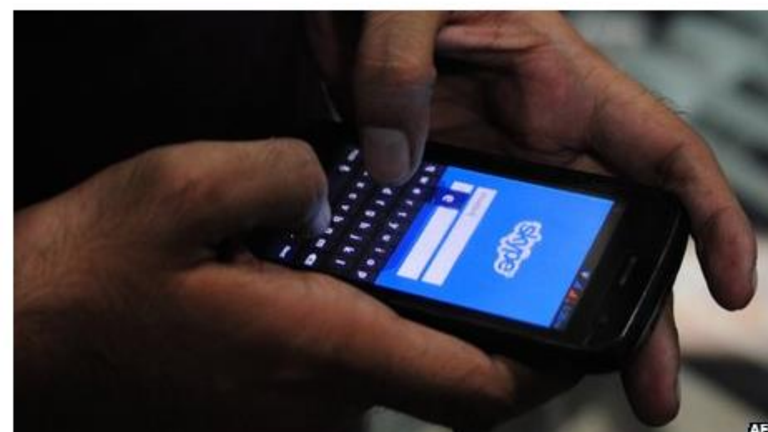
/ MOBILE

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 gadget that let Kirk and Spock talk

Skype to get 'real-time' translator



Analysts say the translation feature could have wide ranging applications

Inside Microsoft Research

Advancing the state of the art in computing

Microsoft Research | News center | Research areas | About | Contact

DNN Research Improves Bing Voice Search

Inside Microsoft Research 17 Jun 2013 10:00 AM 4

Like 22

Tweet 17

Posted by Rob Knies

We live in a society obsessed with speed. Whether it's download times on a mobile phone or Usain Bolt's time in the 100 meters, the faster the better. We also live during an era when accuracy has become not just preferable but essential. The technological marvels of the 21st century demand it.

Speed=good. Accuracy=good. Put them together, and you've got a leap forward, such as **recent advancements in Bing Voice Search** for **Windows Phone** that enable customers to get faster, more accurate results

Microsoft

The Fire Hose

Covering the news of the day at Microsoft

Bing helps you find better images in your searches with 'deep learning'

22 Nov 2013 9:21 AM

MIT Technology Review

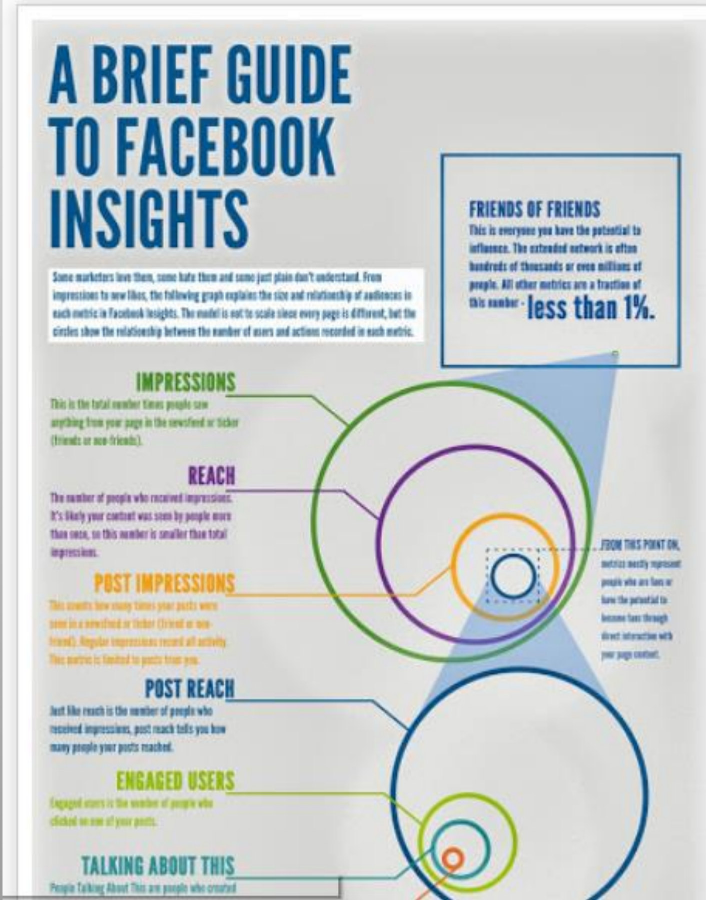
Facebook Launches Advanced AI Effort to Find Meaning in Your Posts

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

By Tom Simonite on September 20, 2013

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



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."

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*

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?

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

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.**

W. P. CAREY SCHOOL of BUSINESS
ARIZONA STATE UNIVERSITY

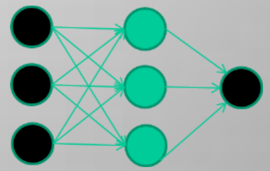
but:

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

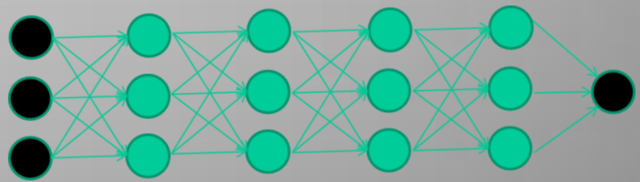
but:

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-later networks</u>**

longer answers

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

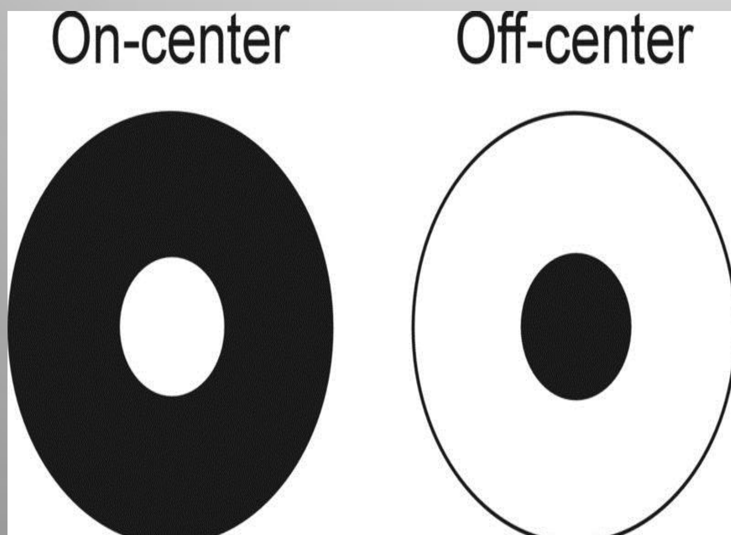
ASU W. P. CAREY SCHOOL of BUSINESS ARIZONA STATE UNIVERSITY

How the brain works

- The cortical visual system is composed of multiple visual areas with different functions.
- V1 neurons describe object features.
- The principle of columnar organization.
- Two visual streams – ‘what’ and ‘how’ (or ‘where’).
- MT neurons describe motion and depth (dorsal stream).
- IT neurons describe objects (ventral stream).

Classical Receptive field in vision:

a two-dimensional region in visual space whose size can range from a few minutes of arc to tens of degrees



On-center and Off-center receptive fields. The receptive fields of **retinal ganglion cells and thalamic neurons** are organized as two concentric circles with different contrast polarities. On-center neurons respond to the presentation of a light spot on a dark background and off-center neurons to the presentation of a dark spot on a light background.

Convergence

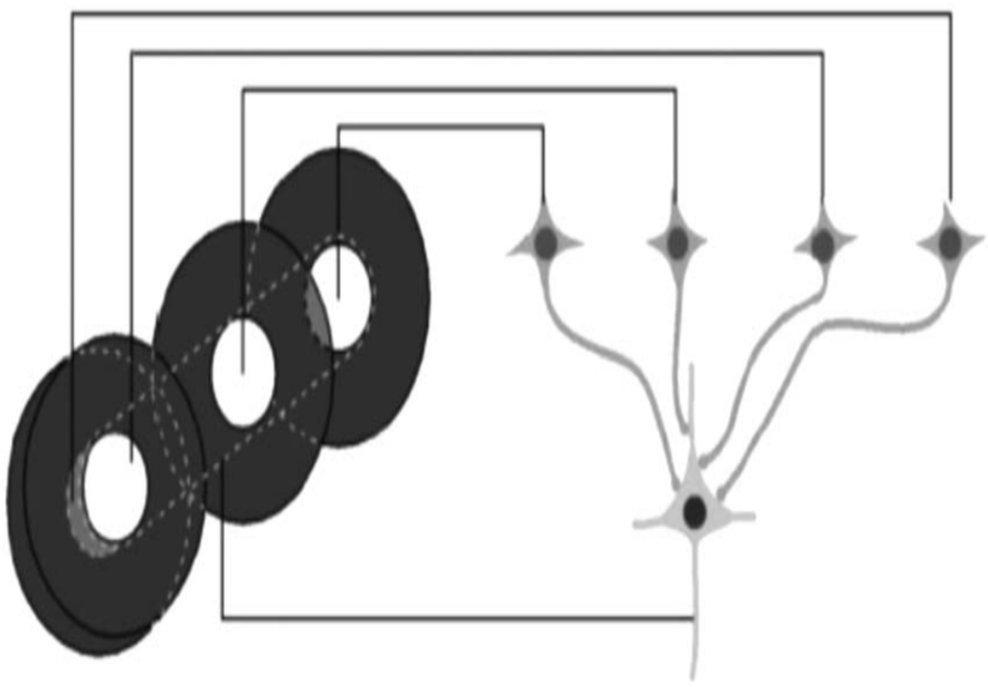
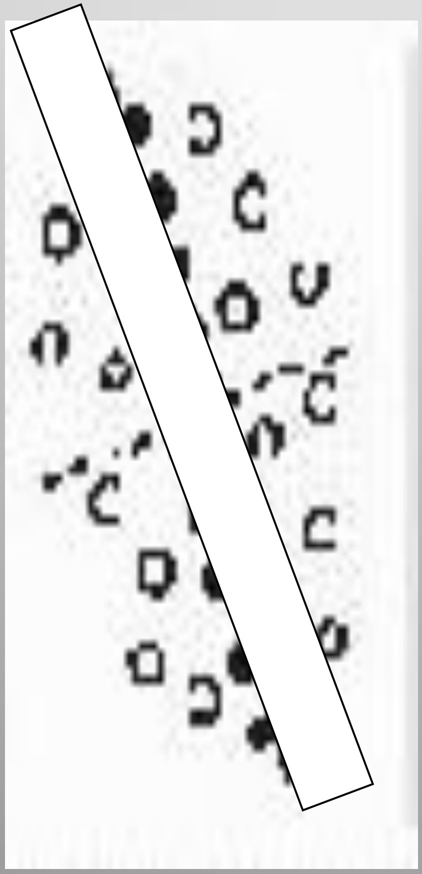


Fig. 1. Classic feedforward model from LGN to simple cells in V1 cortex. Adapted with permission from Hubel and Wiesel (1962). Four LGN cells are drawn as converging onto a single V1 cell. The circular LGN receptive fields aligned in a row on the left side of the diagram make the receptive field of the cortical cell elongated.

Shapley et al. 2007

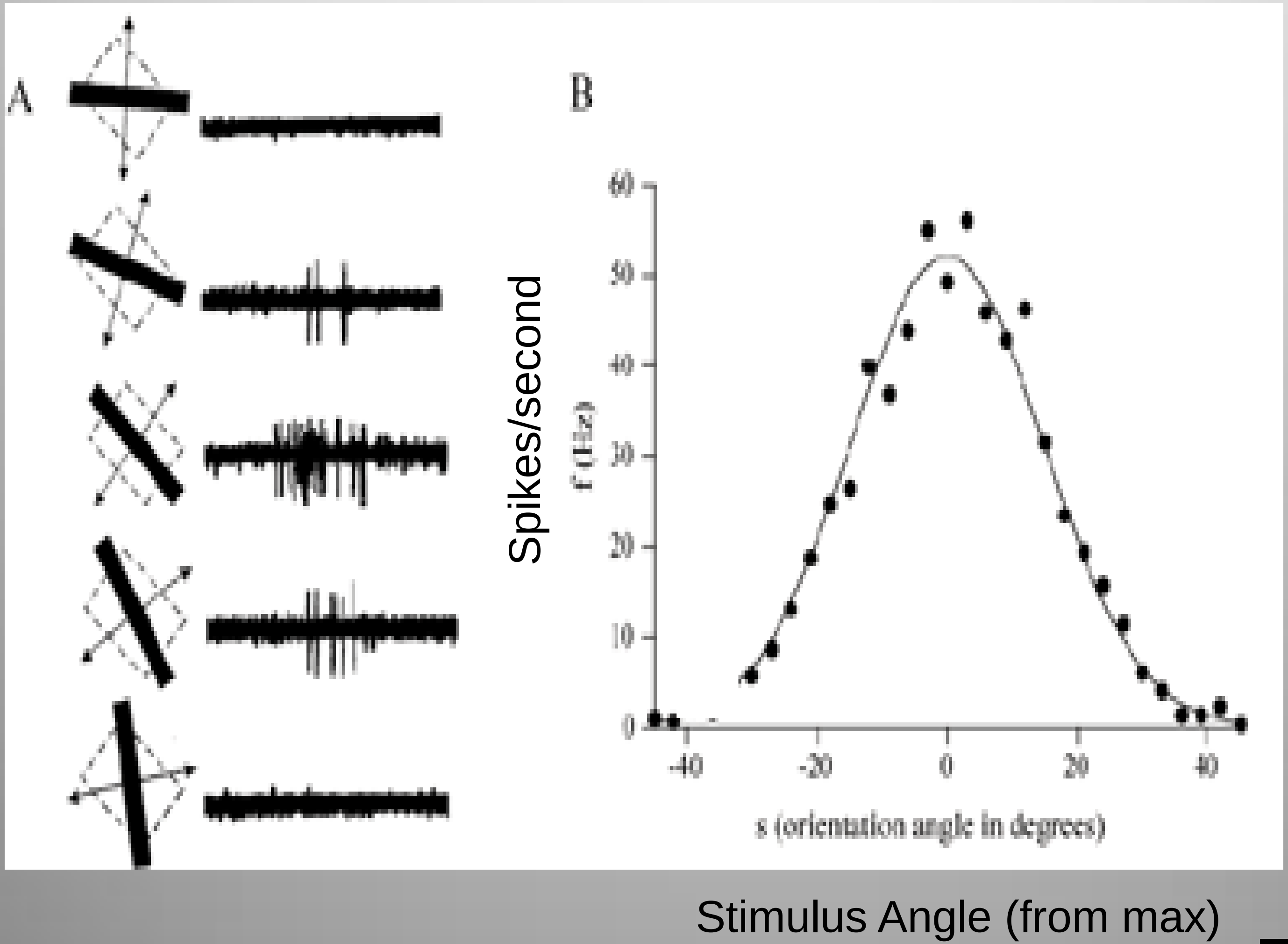
V1 simple cell is most responsive to an oriented line



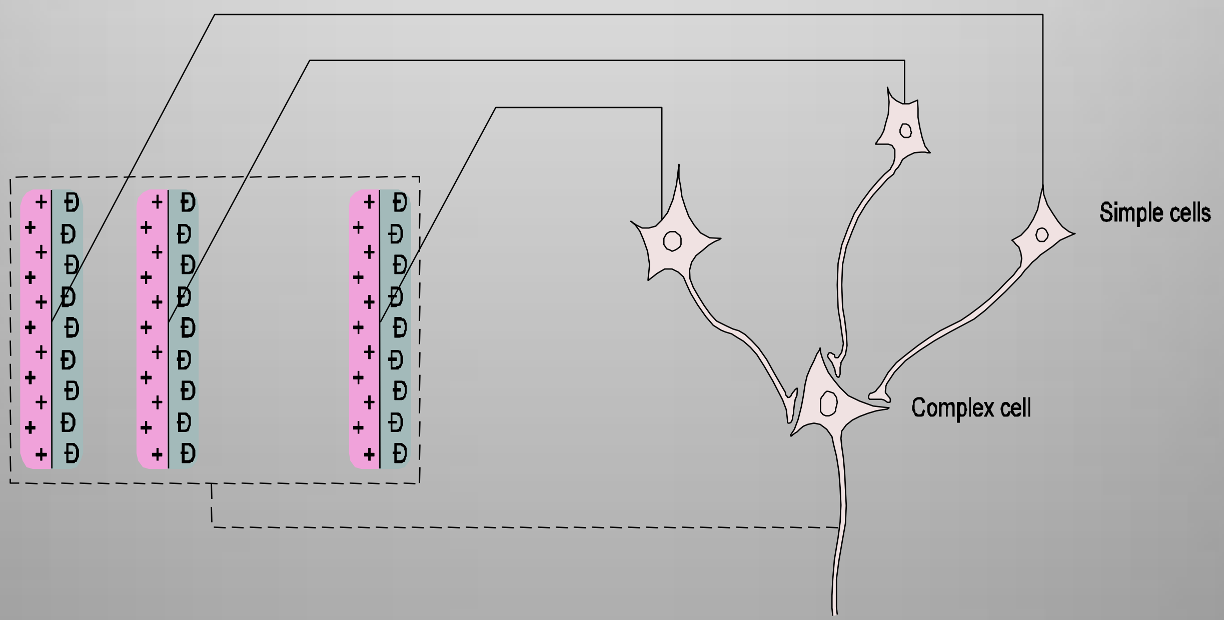
Off-response

On-response

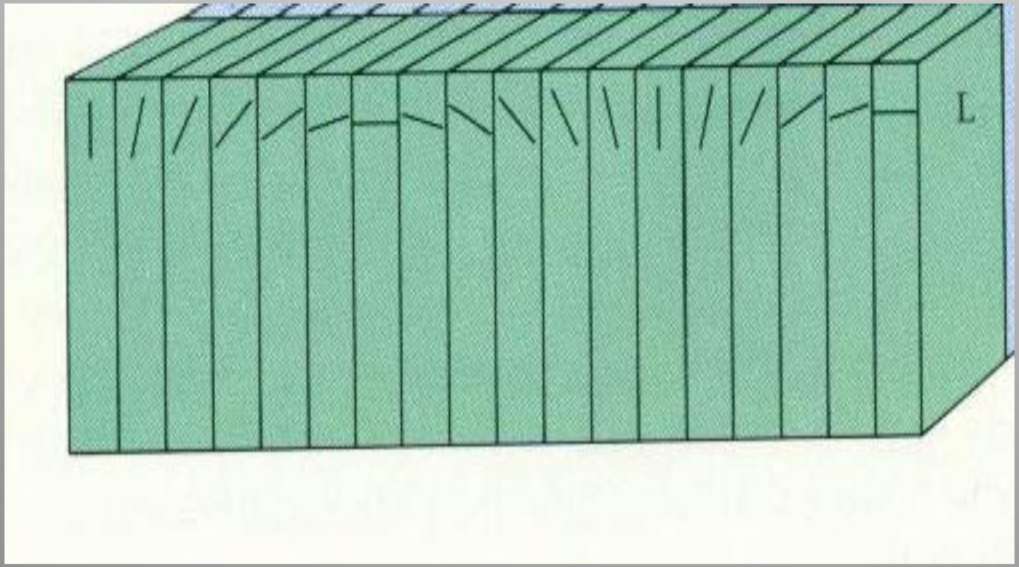
Orientation tuning in a V1 simple cell



Complex cells can be constructed from an array of similarly oriented simple cells

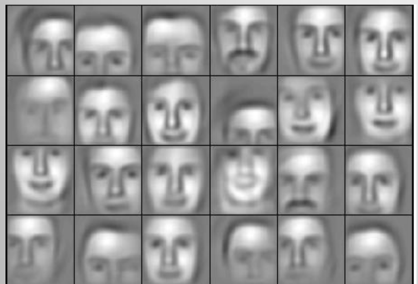


Cells with similar orientation preferences lie in the same column

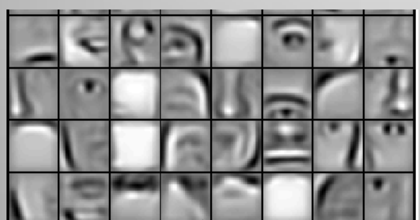


Feature Hierarchies

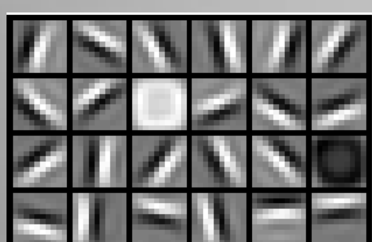
Why feature hierarchies



object models



object parts
(combination
of edges)

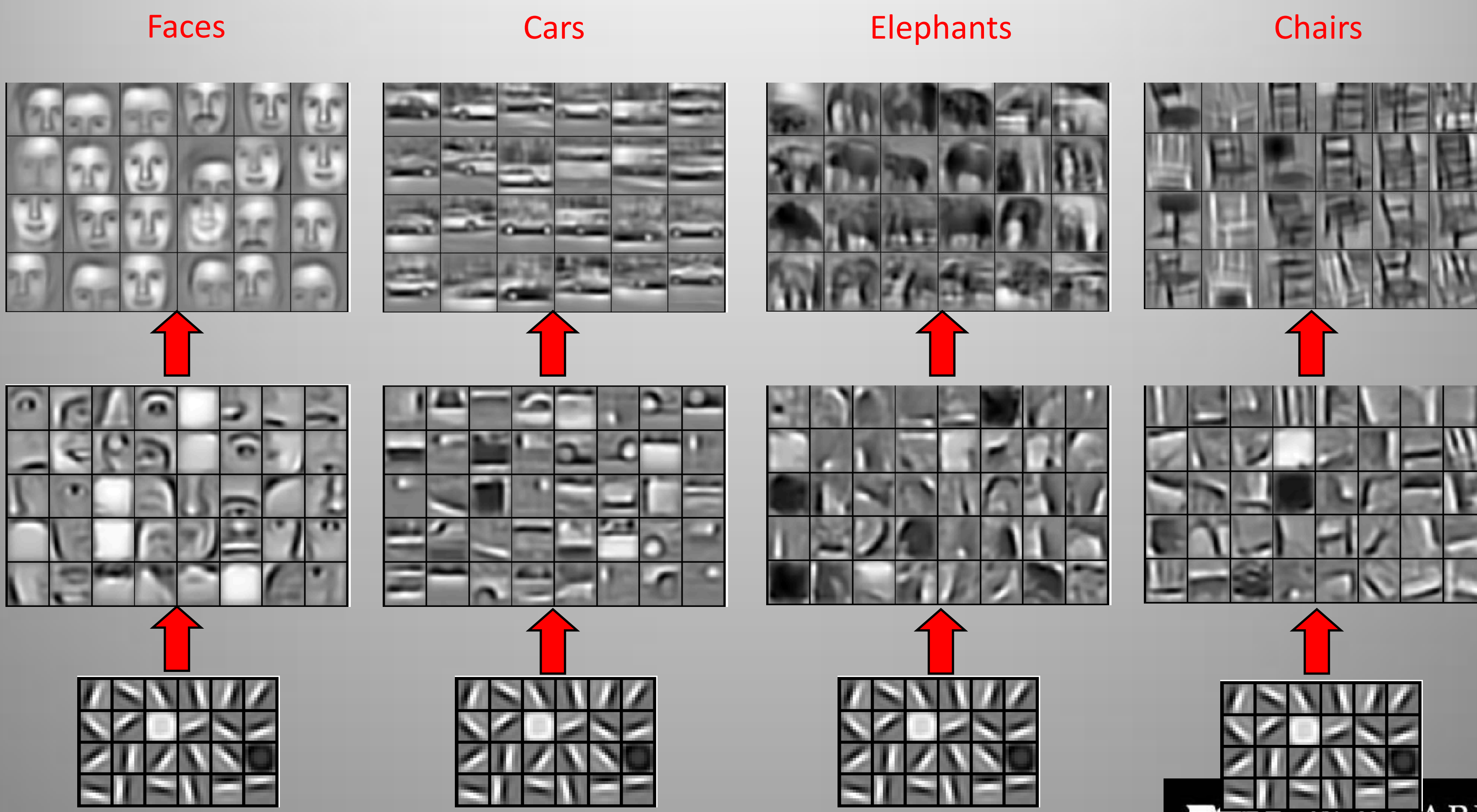


edges

pixels

Learning of object parts

Examples of learned object parts from object categories

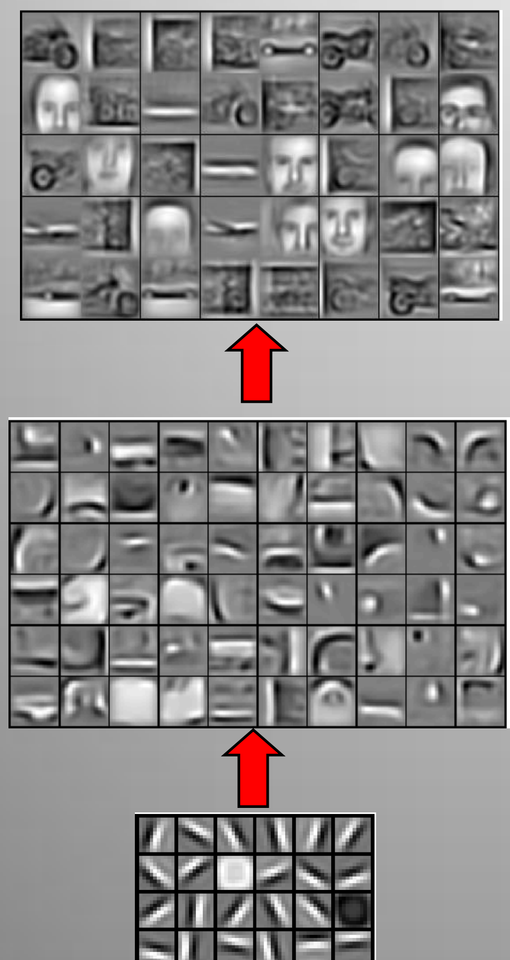


Training on multiple objects

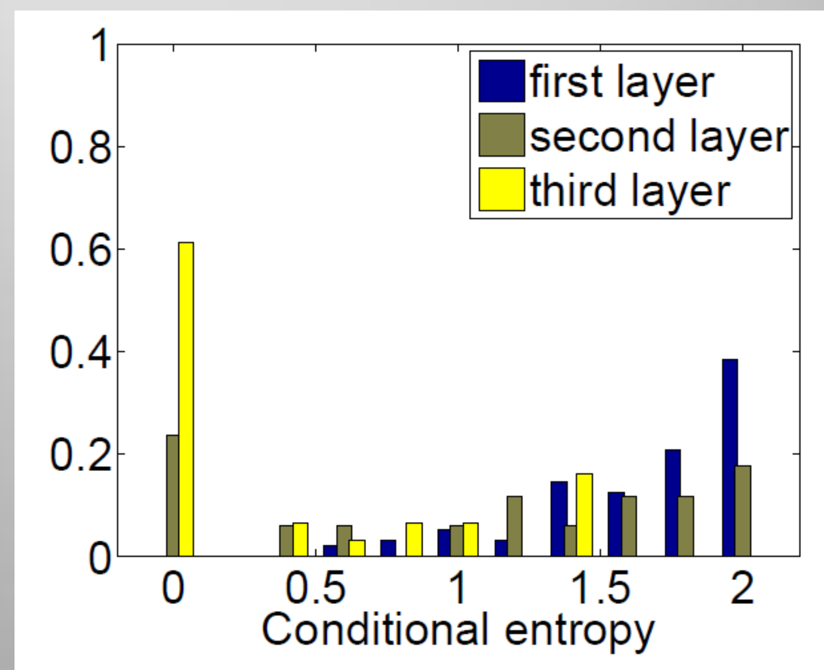
Trained on 4 classes (cars, faces, motorbikes, airplanes).

Second layer: Shared-features and object-specific features.

Third layer: More specific features.

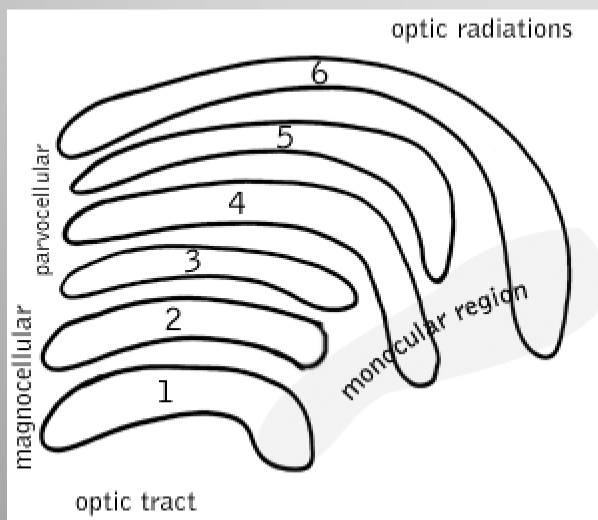


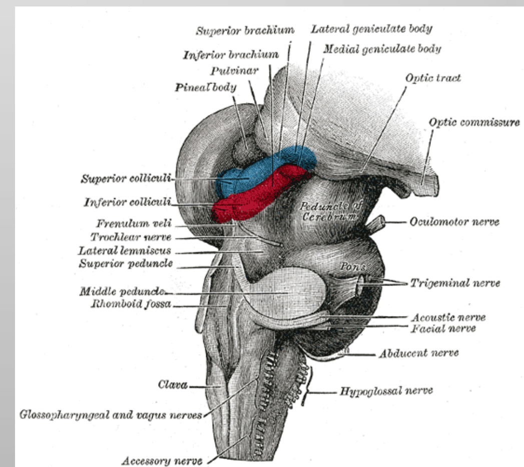
Plot of $H(\text{class}|\text{neuron active})$



So: *multiple layers make sense*

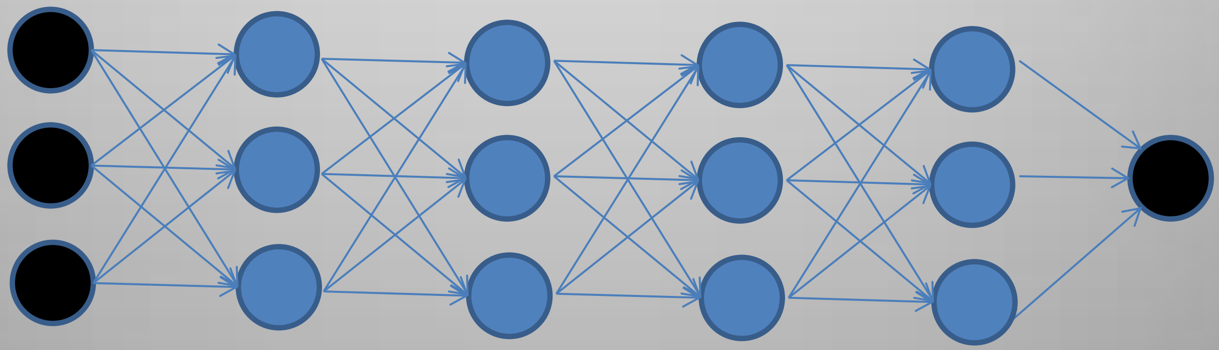
Your brain works that way





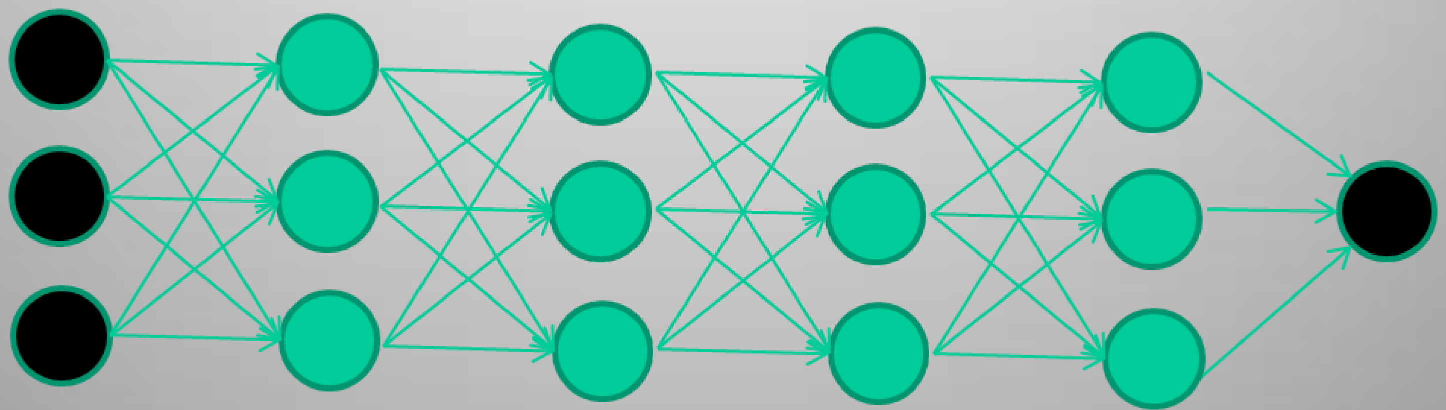
ASU W. P. CAREY SCHOOL of BUSINESS
ARIZONA STATE UNIVERSITY

So: *multiple layers make sense*

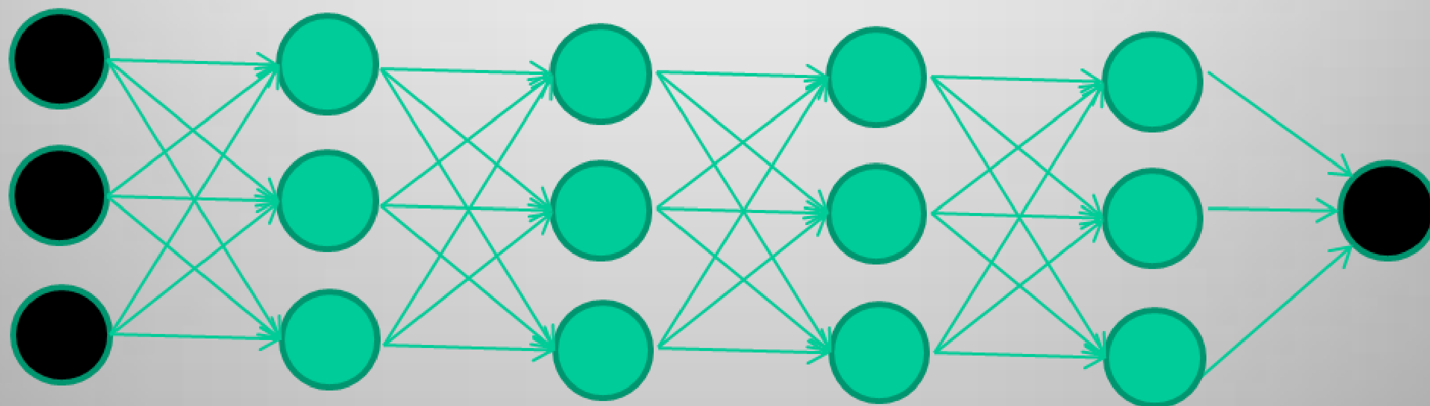


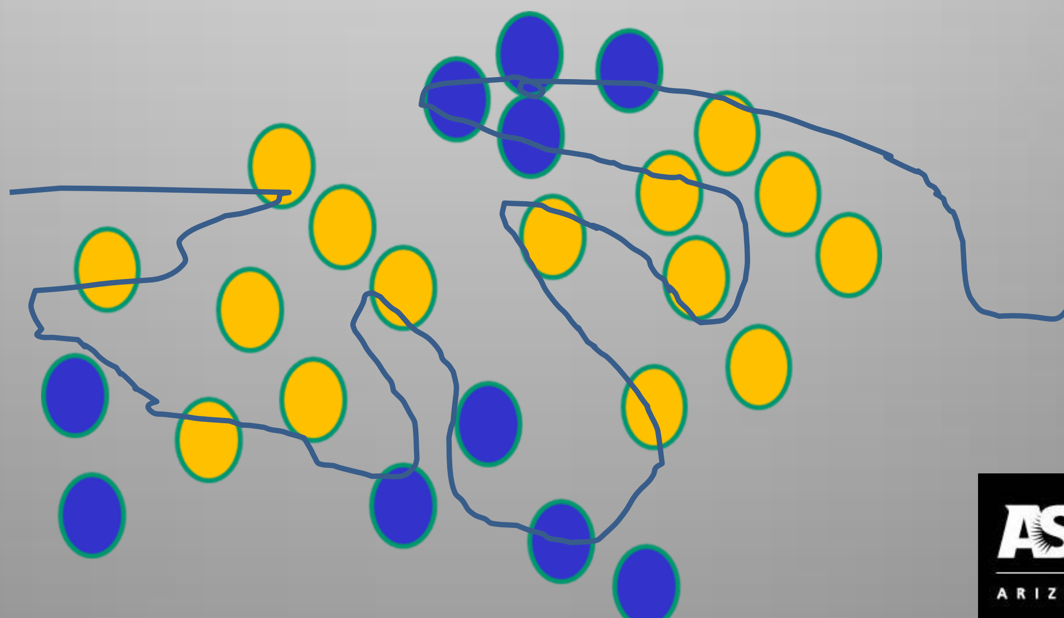
So: *multiple layers make sense*

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



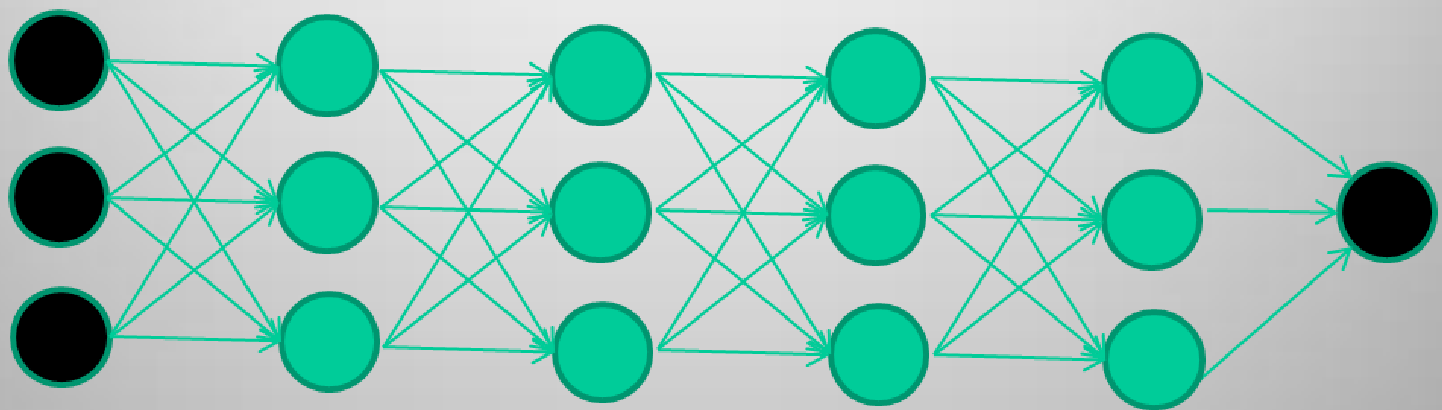
But, until very recently, our weight-learning algorithms simply did not work on multi-layer architectures



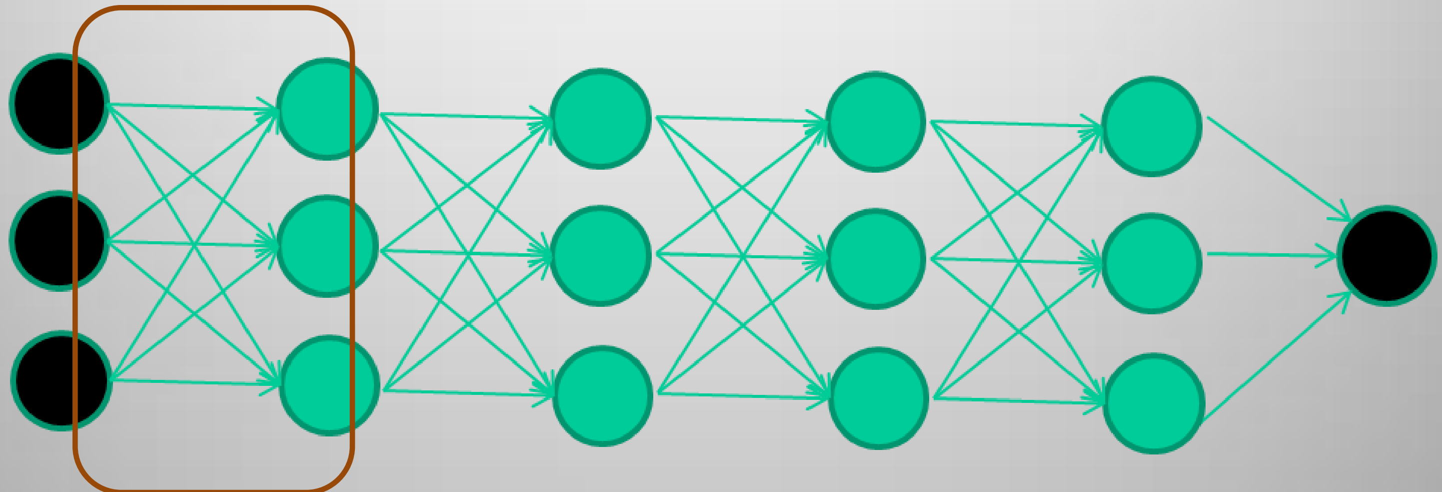


Along came deep learning …

The new way to train multi-layer NNs…

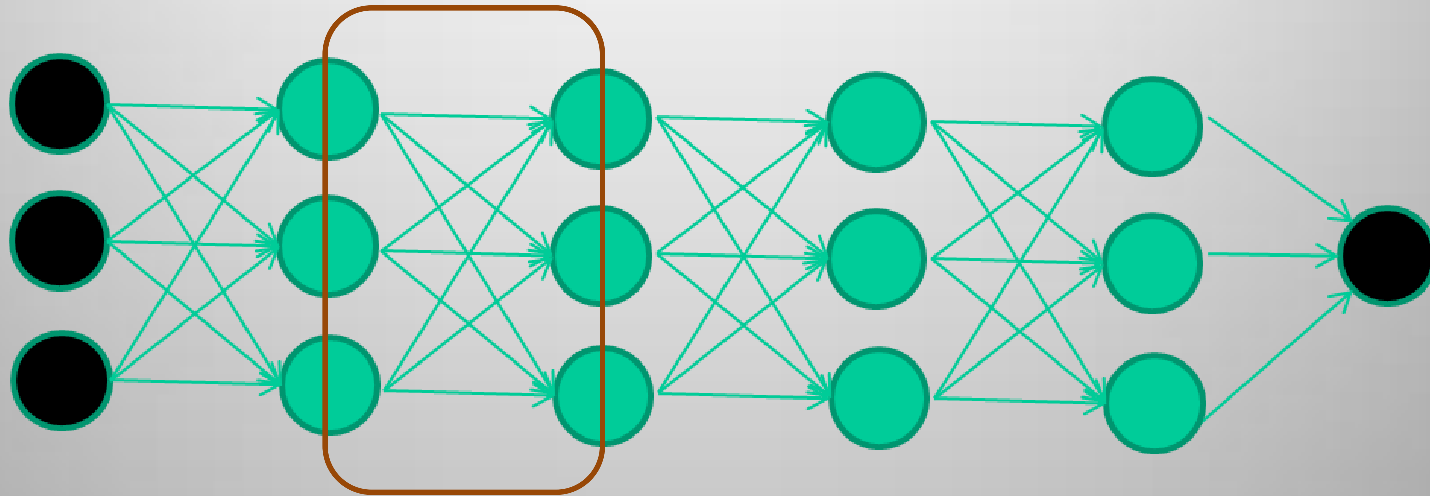


The new way to train multi-layer NNs…



Train **this** layer first

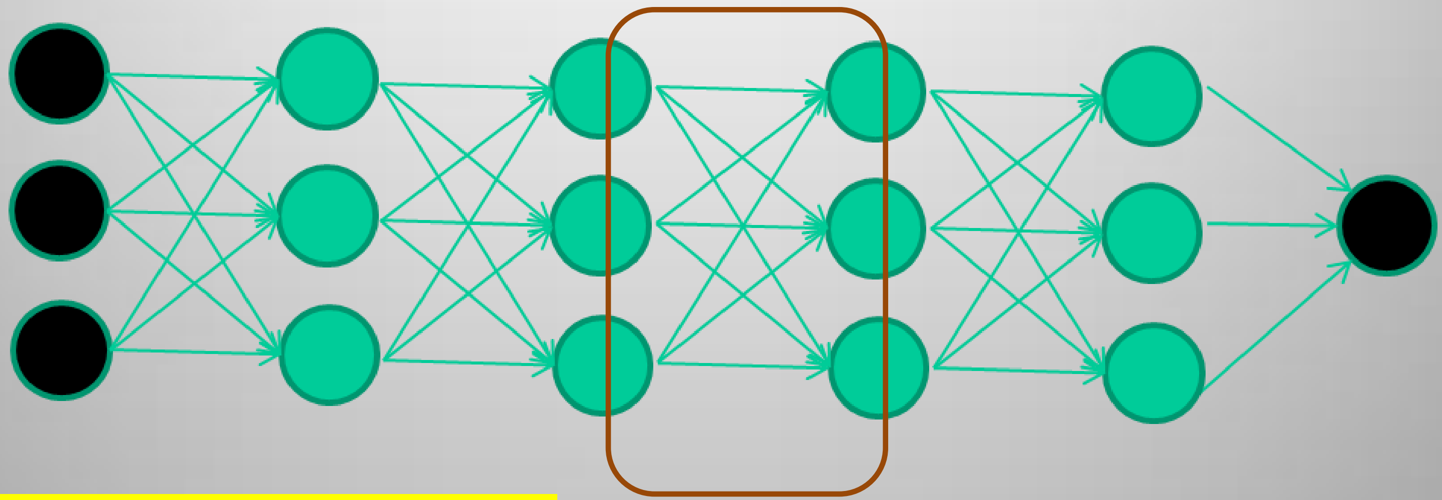
The new way to train multi-layer NNs…



Train **this** layer first

then **this** layer

The new way to train multi-layer NNs…



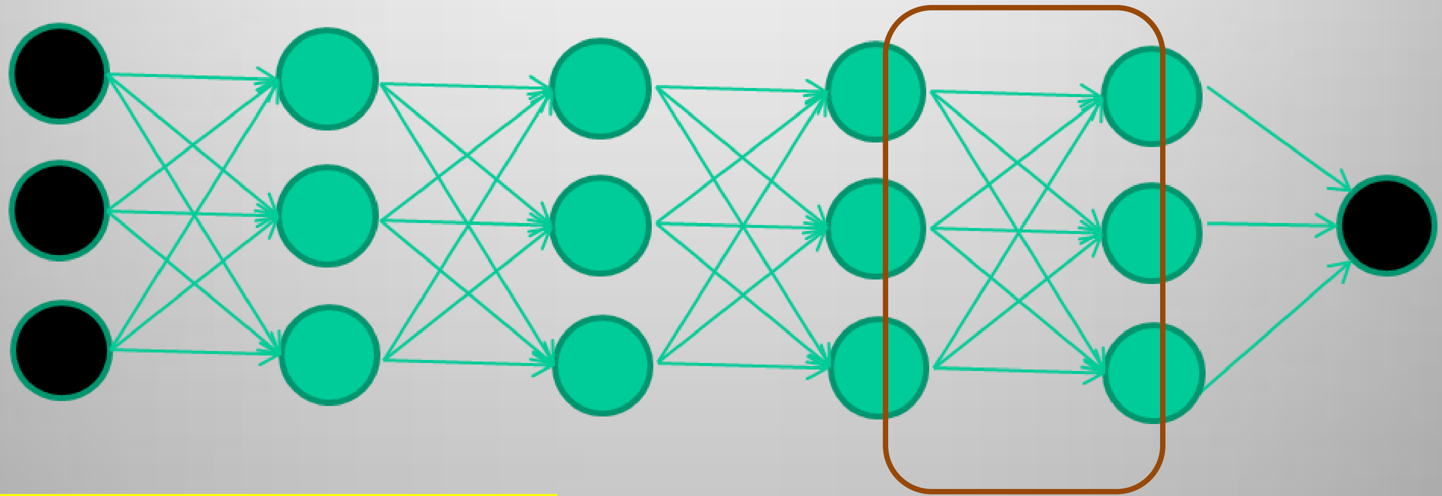
Train **this** layer first

then **this** layer

then **this** layer

ASU W. P. CAREY SCHOOL of BUSINESS ARIZONA STATE UNIVERSITY

The new way to train multi-layer NNs…



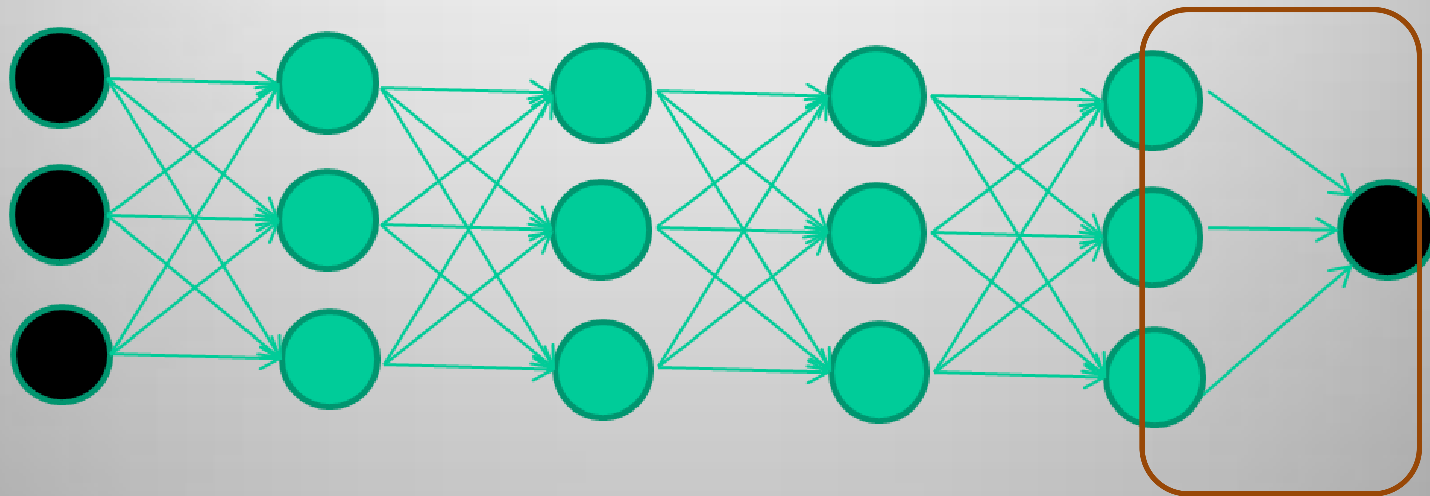
Train **this** layer first

then **this** layer

then **this** layer

then **this** layer

The new way to train multi-layer NNs…



Train **this** layer first

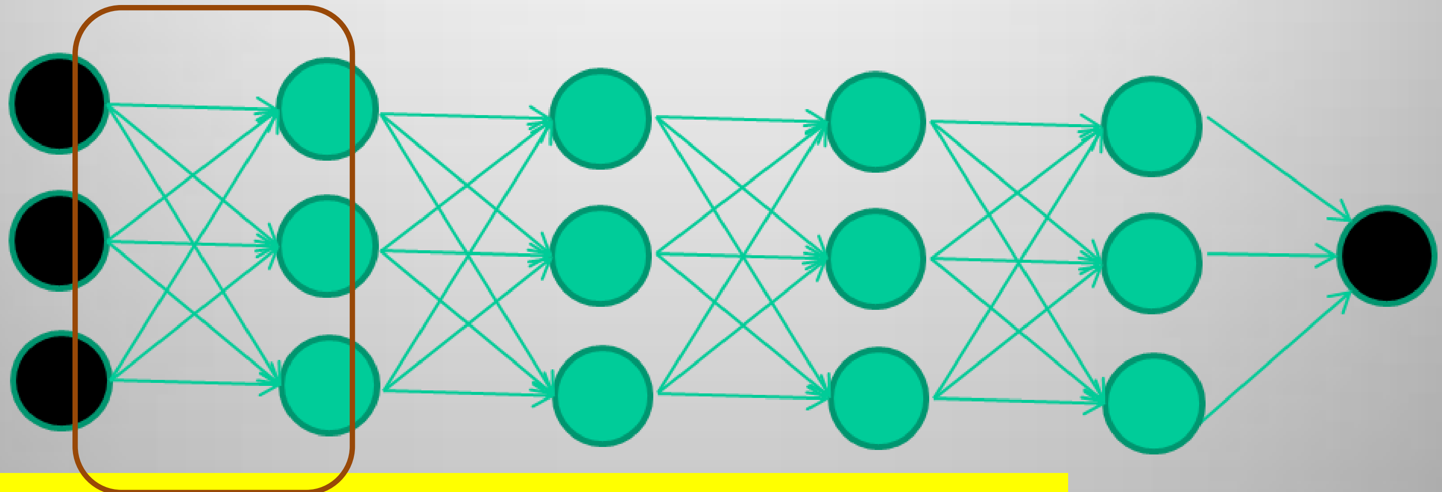
then **this** layer

then **this** layer

then **this** layer

finally **this** layer

The new way to train multi-layer NNs…

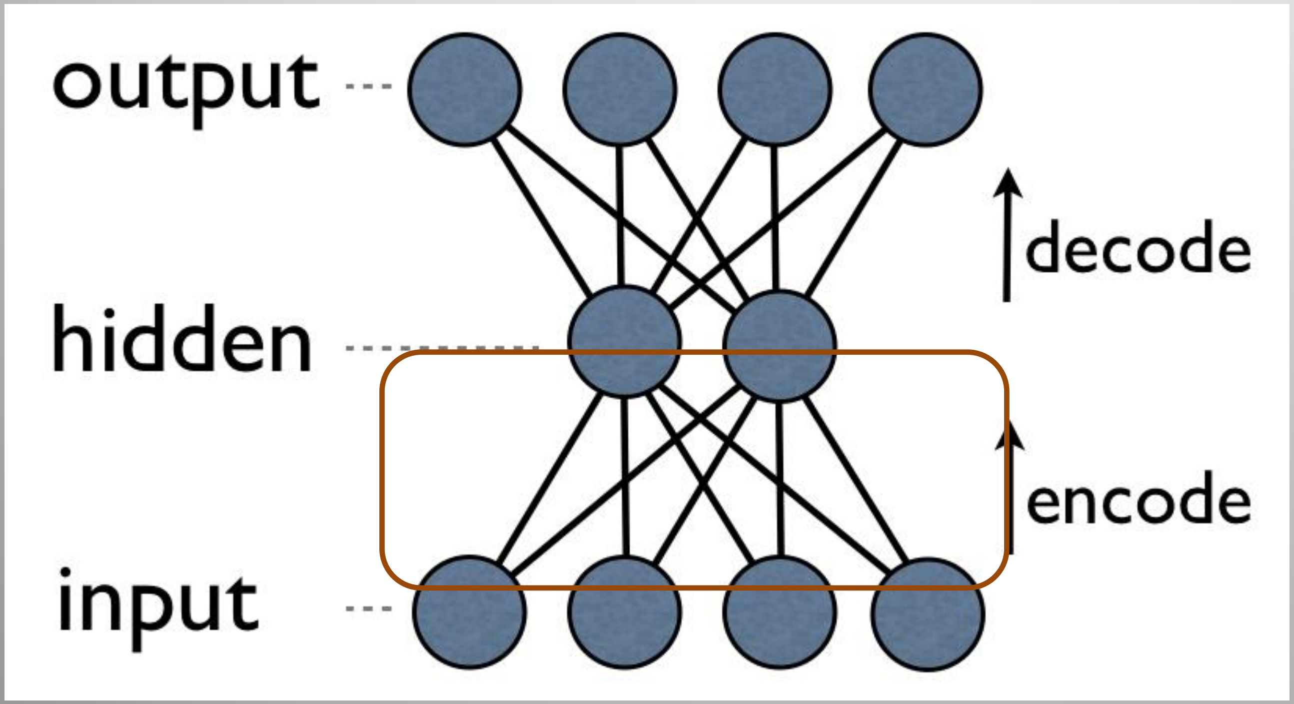


EACH of the (non-output) layers is trained to be an ***auto-encoder***

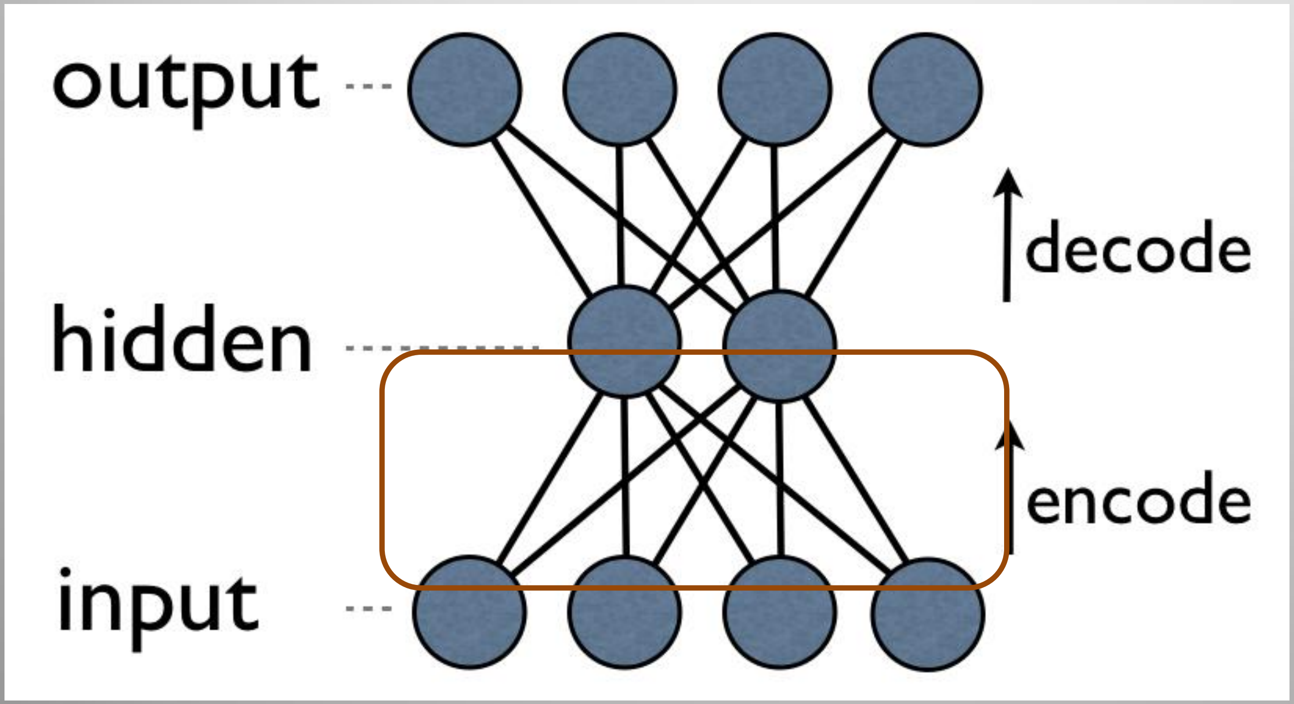
Basically, it is forced to learn good features that describe what comes from the previous layer

CAREY
BUSINESS
IVERSITY

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

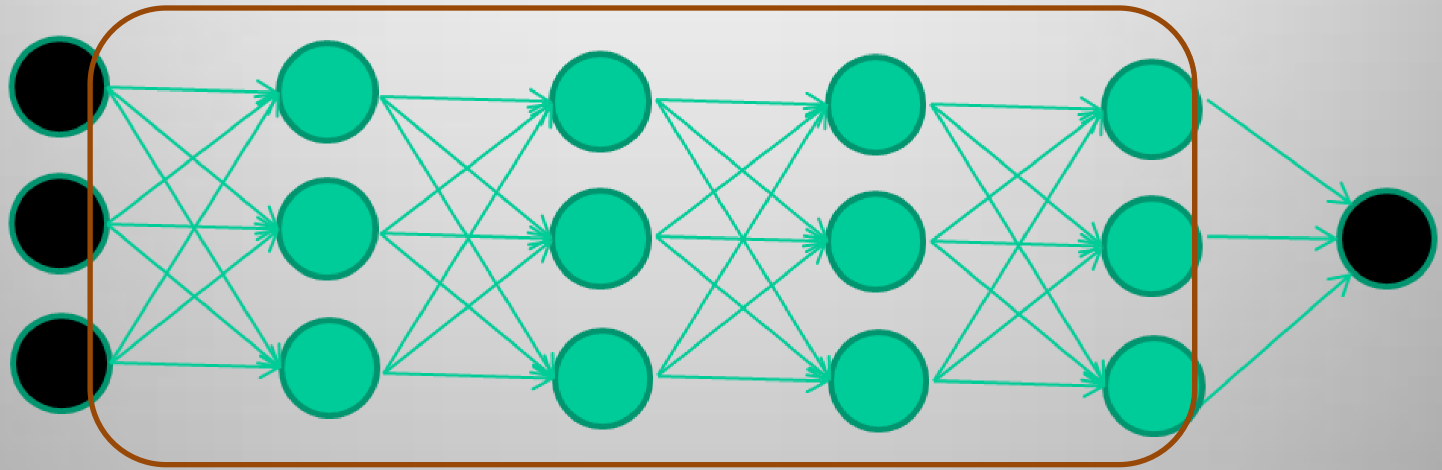


an auto-encoder is trained, with an absolutely standard weight-adjustment algorithm to reproduce the input

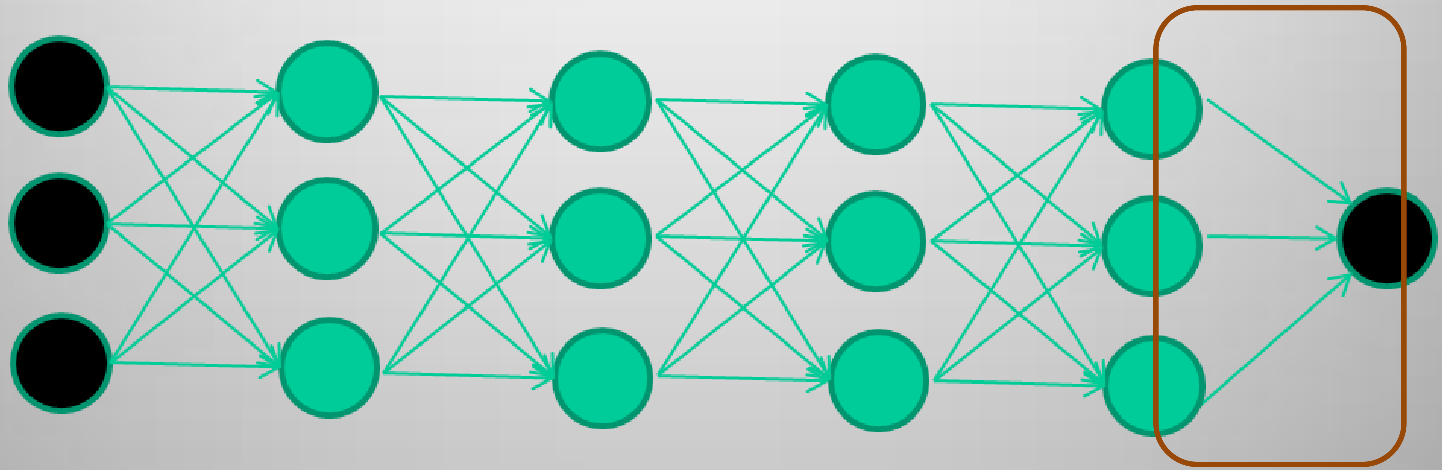


By making this happen with (many) fewer units than the inputs, this forces the 'hidden layer' units to become good feature detectors

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

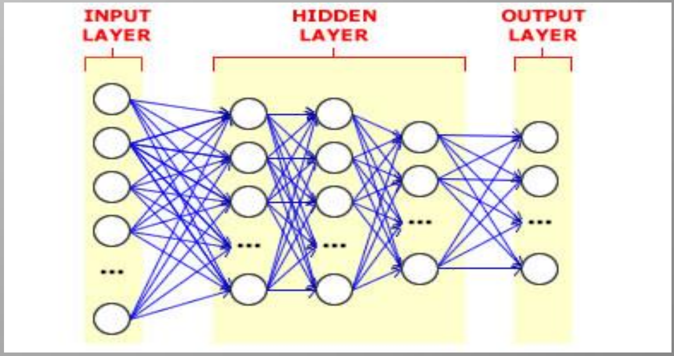


Final layer trained to predict class based on outputs from previous layers



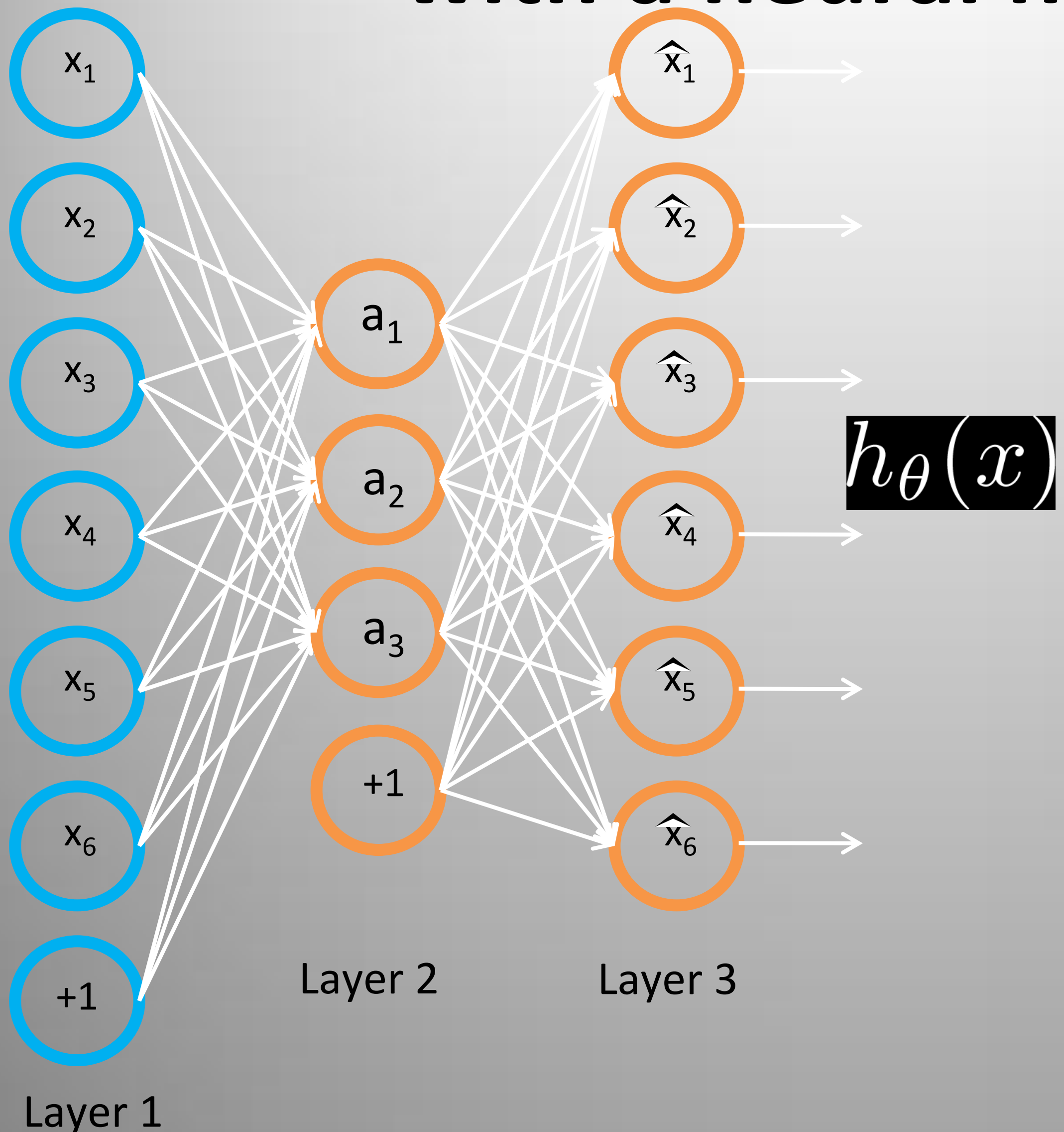
And that's that

- 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 …



Feature Hierarchies and Auto-Encoders

Unsupervised feature learning with a neural network



Autoencoder.

Network is trained to output the input (learn identify function).

$$h_\theta(x) \approx x$$

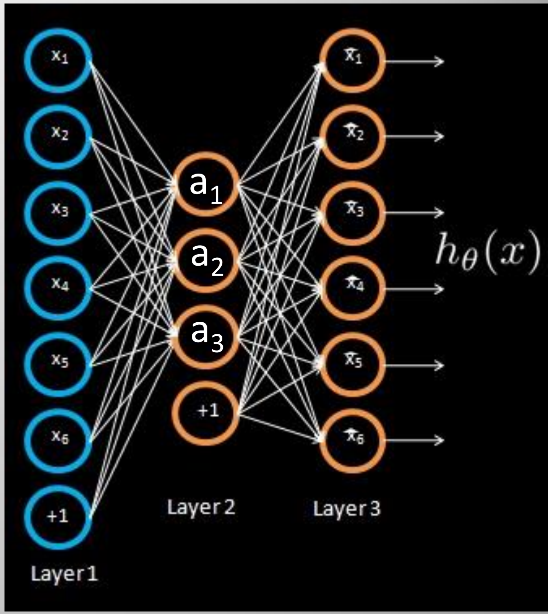
Trivial solution unless:
- Constrain number of units in Layer 2 (learn compressed representation), or
- Constrain Layer 2 to be **sparse**.

Unsupervised feature learning with a neural network

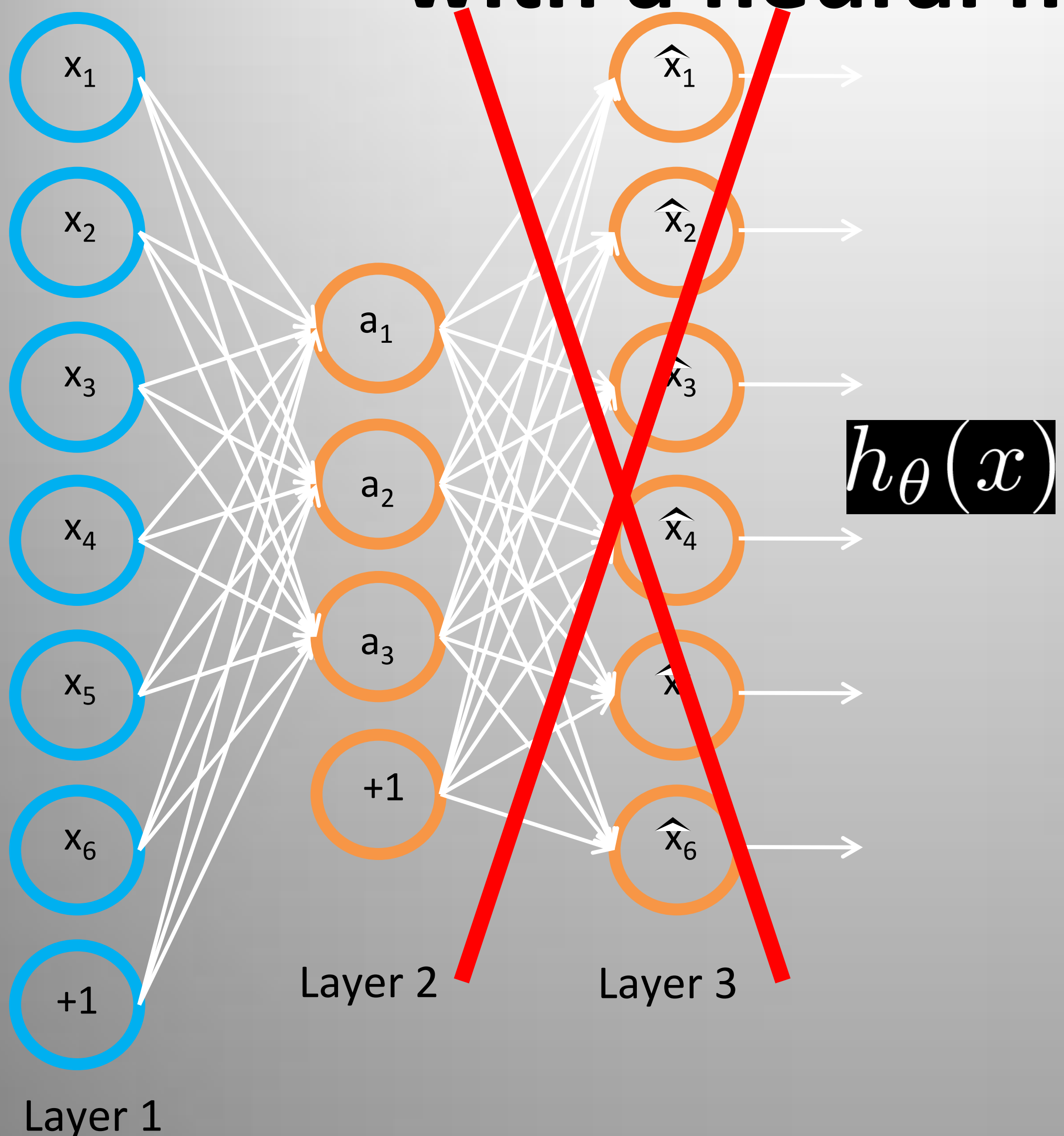
Training a sparse autoencoder.

Given unlabeled training set $x_1, x_2, \ldots$

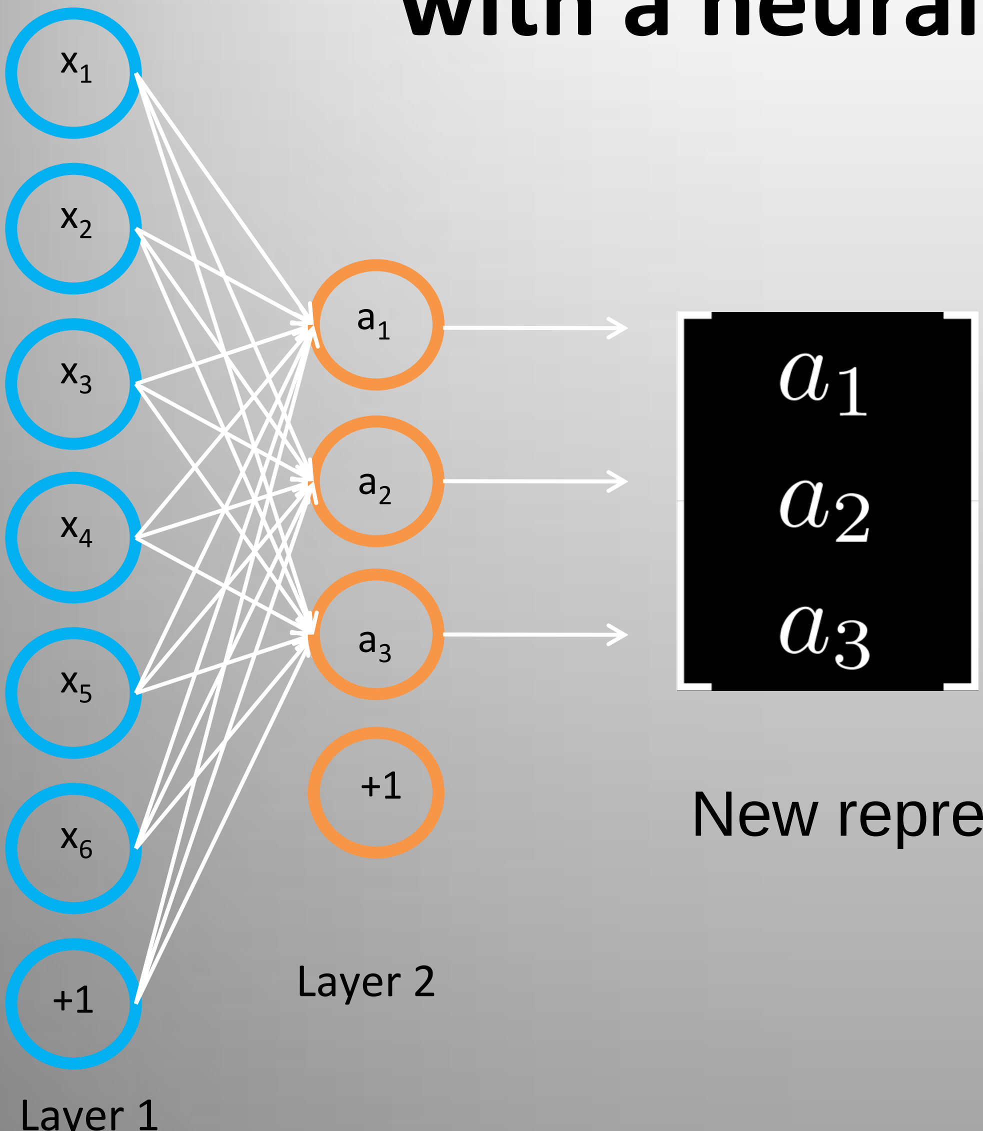
$$\min_{\theta} \underbrace{\|h_\theta(x) - x\|^2}_{\text{Reconstruction error term}} + \underbrace{\lambda \sum_i |a_i|}_{L_1 \text{ sparsity term}}$$



Unsupervised feature learning with a neural network

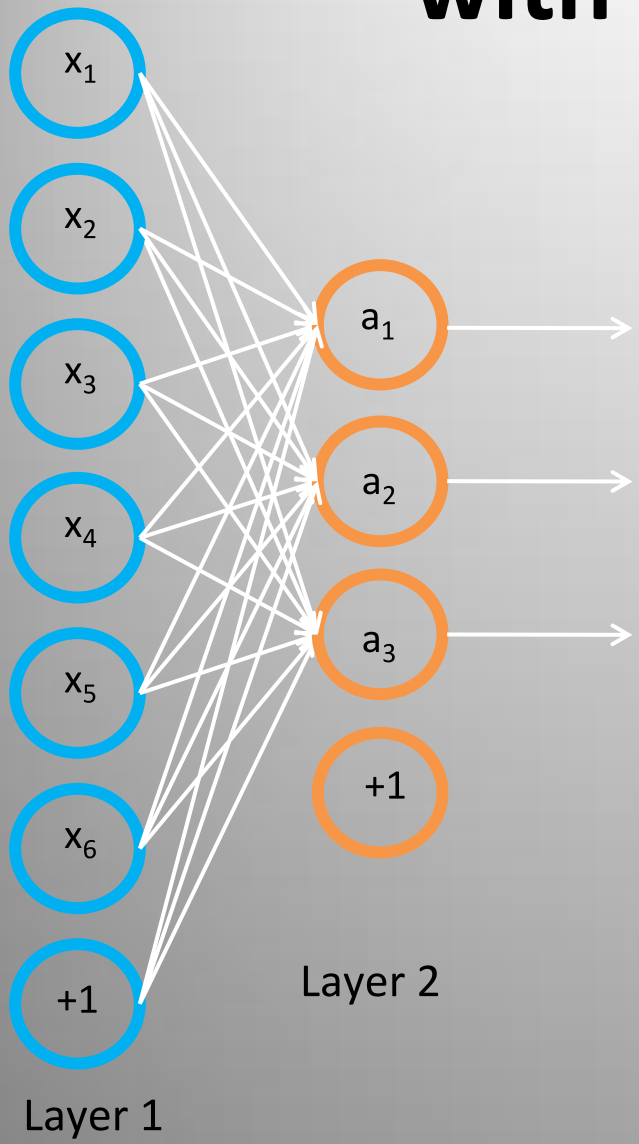


Unsupervised feature learning with a neural network

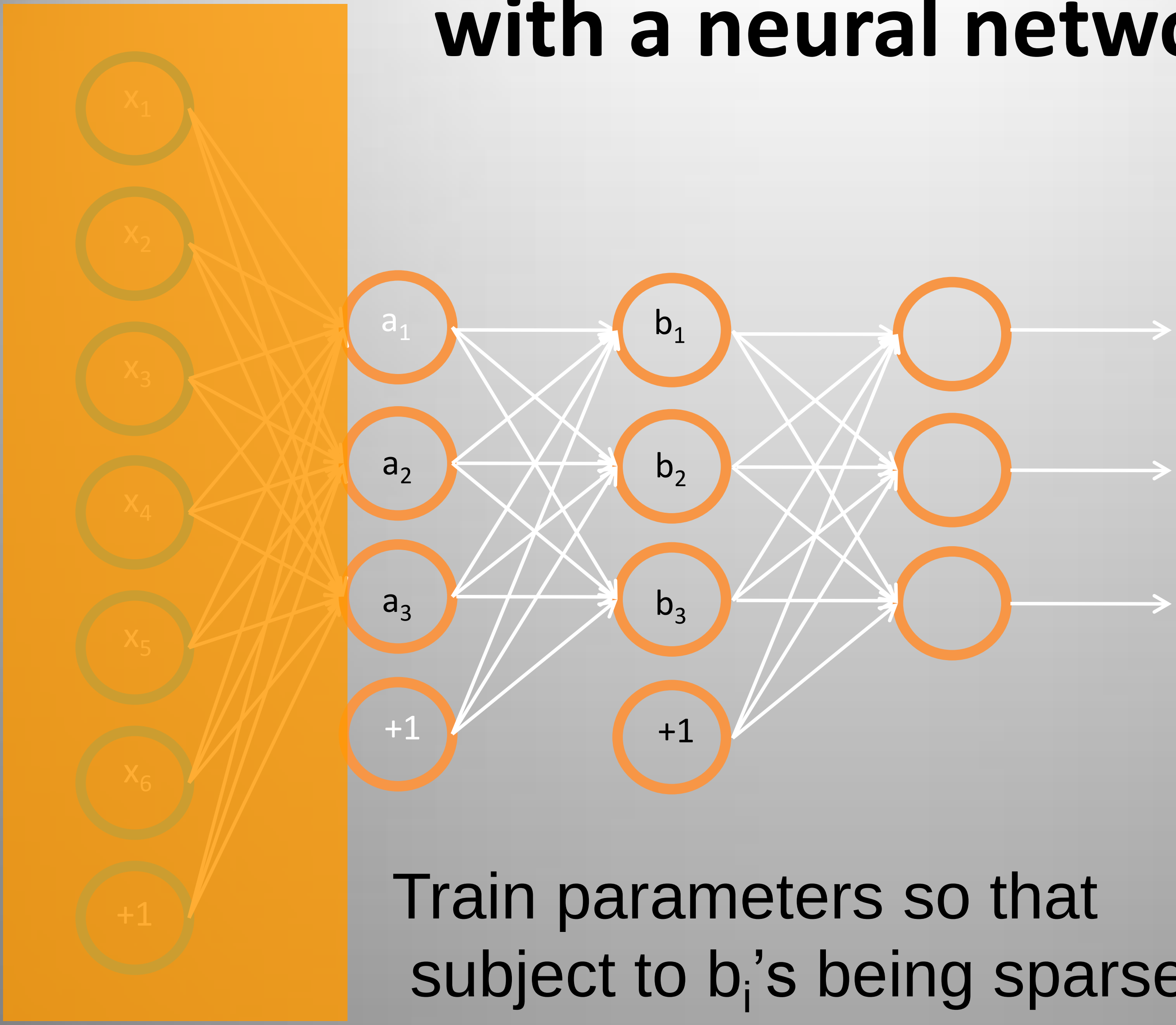


New representation for input.

Unsupervised feature learning with a neural network



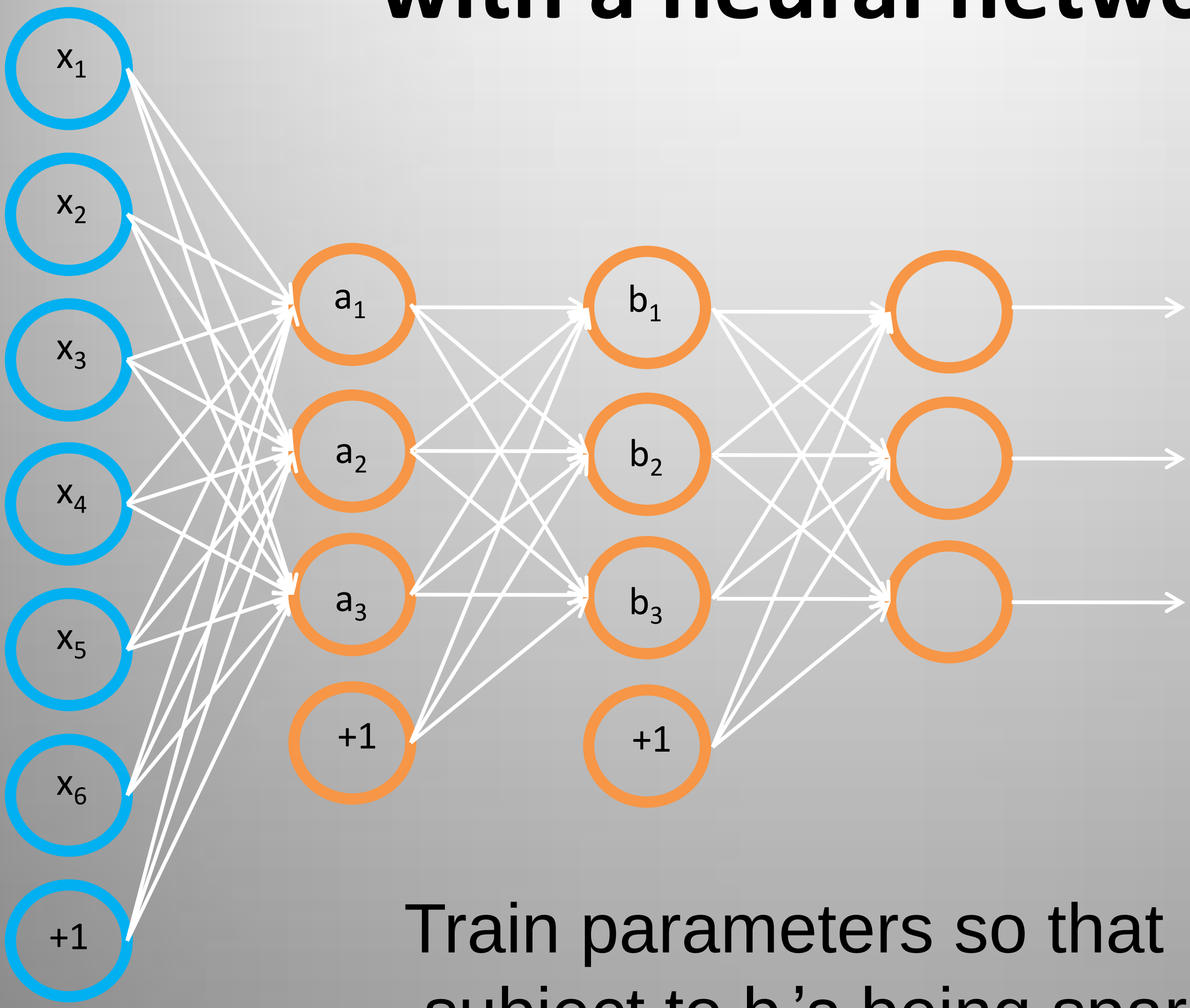
Unsupervised feature learning with a neural network



$h_\theta(x)$

Train parameters so that $h_\theta(x) \approx a$ subject to b_i's being sparse.

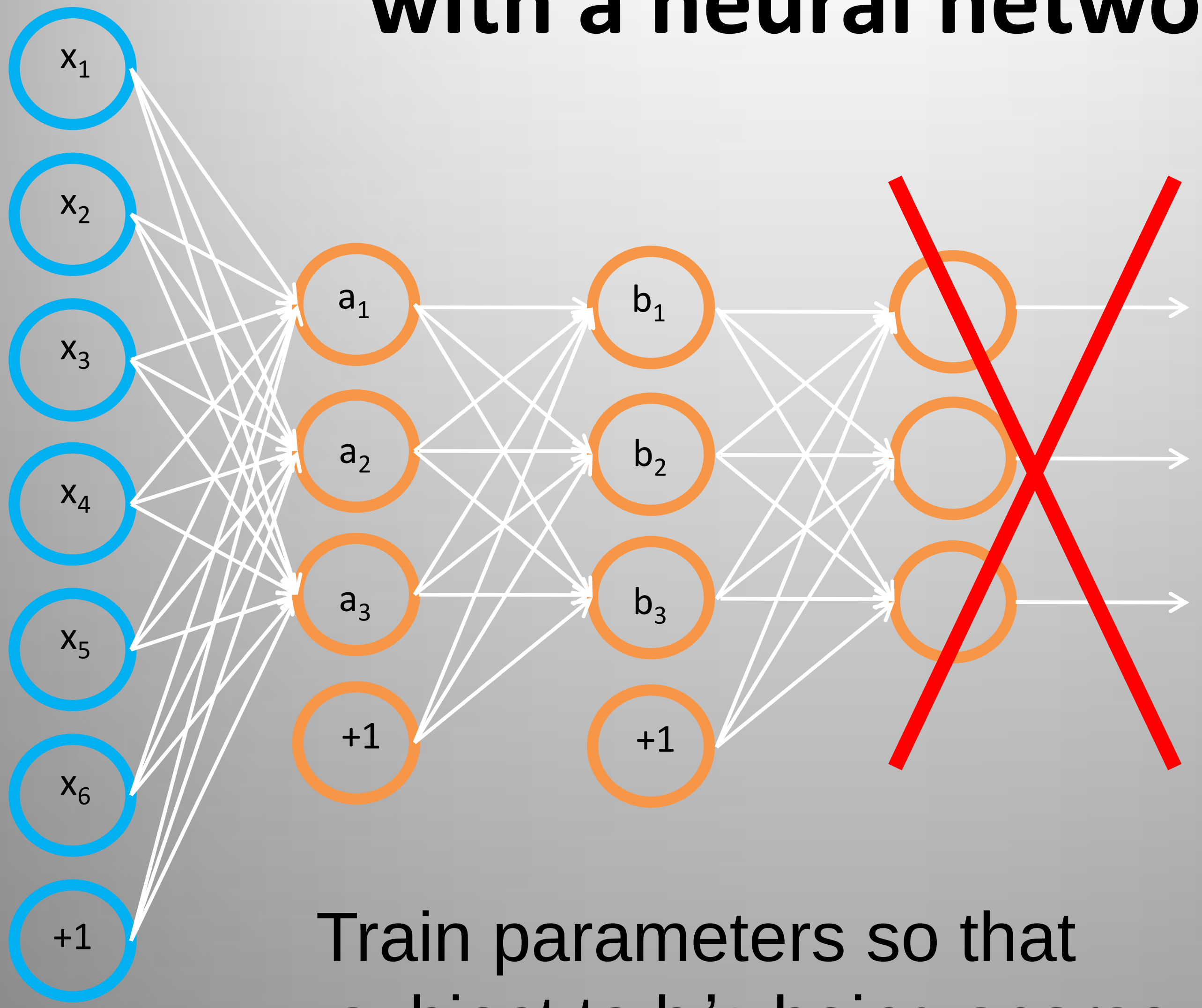
Unsupervised feature learning with a neural network



$h_\theta(x)$

Train parameters so that $h_\theta(x) \approx a$
subject to b_i's being sparse.

Unsupervised feature learning with a neural network

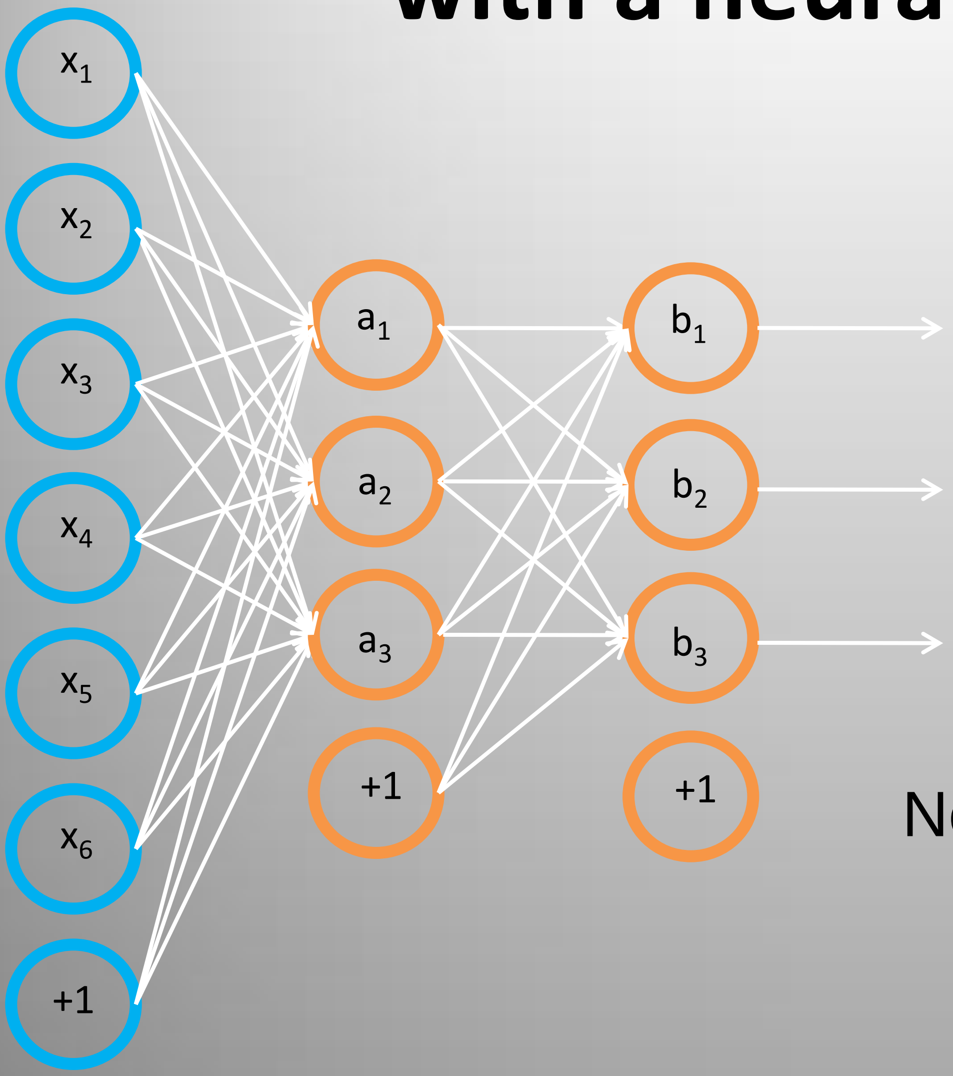


$h_\theta(x)$

Train parameters so that $h_\theta(x) \approx x$

subject to b_i's being sparse.

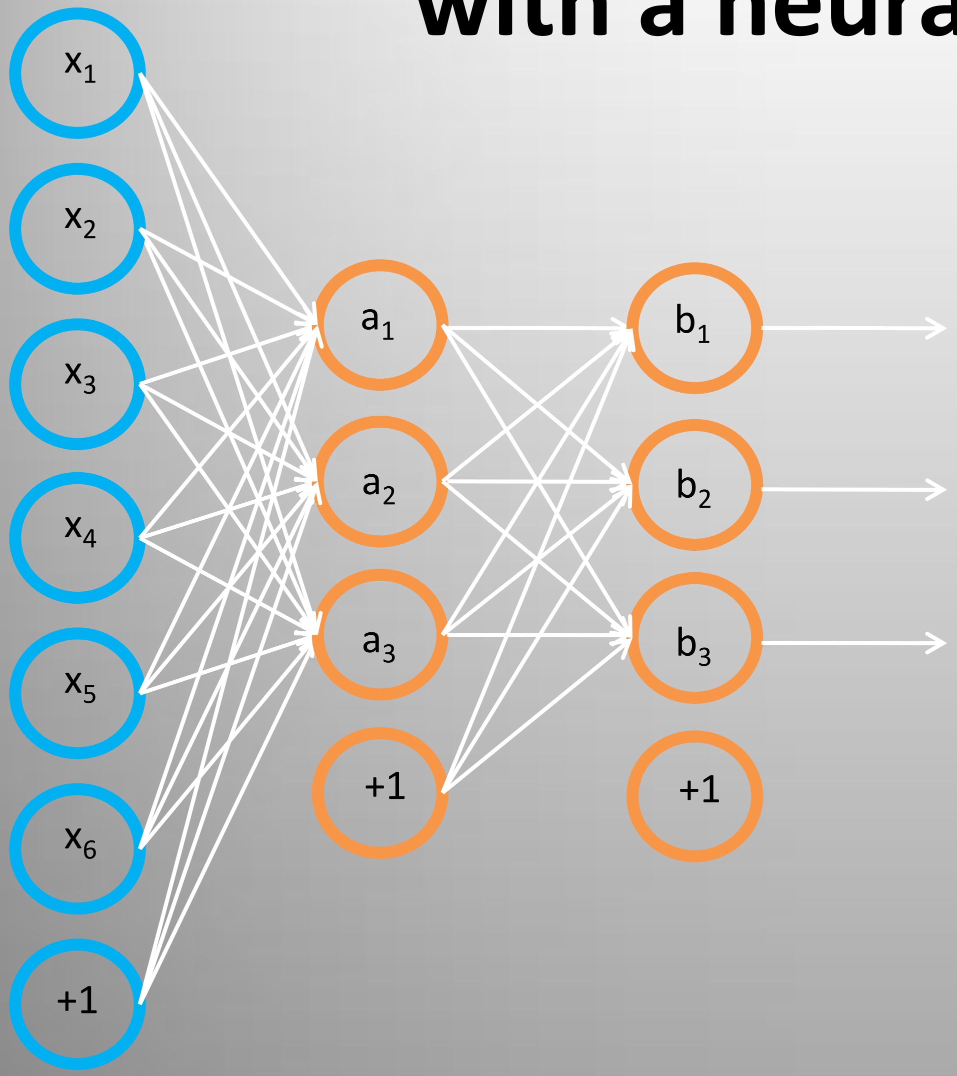
Unsupervised feature learning with a neural network



$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

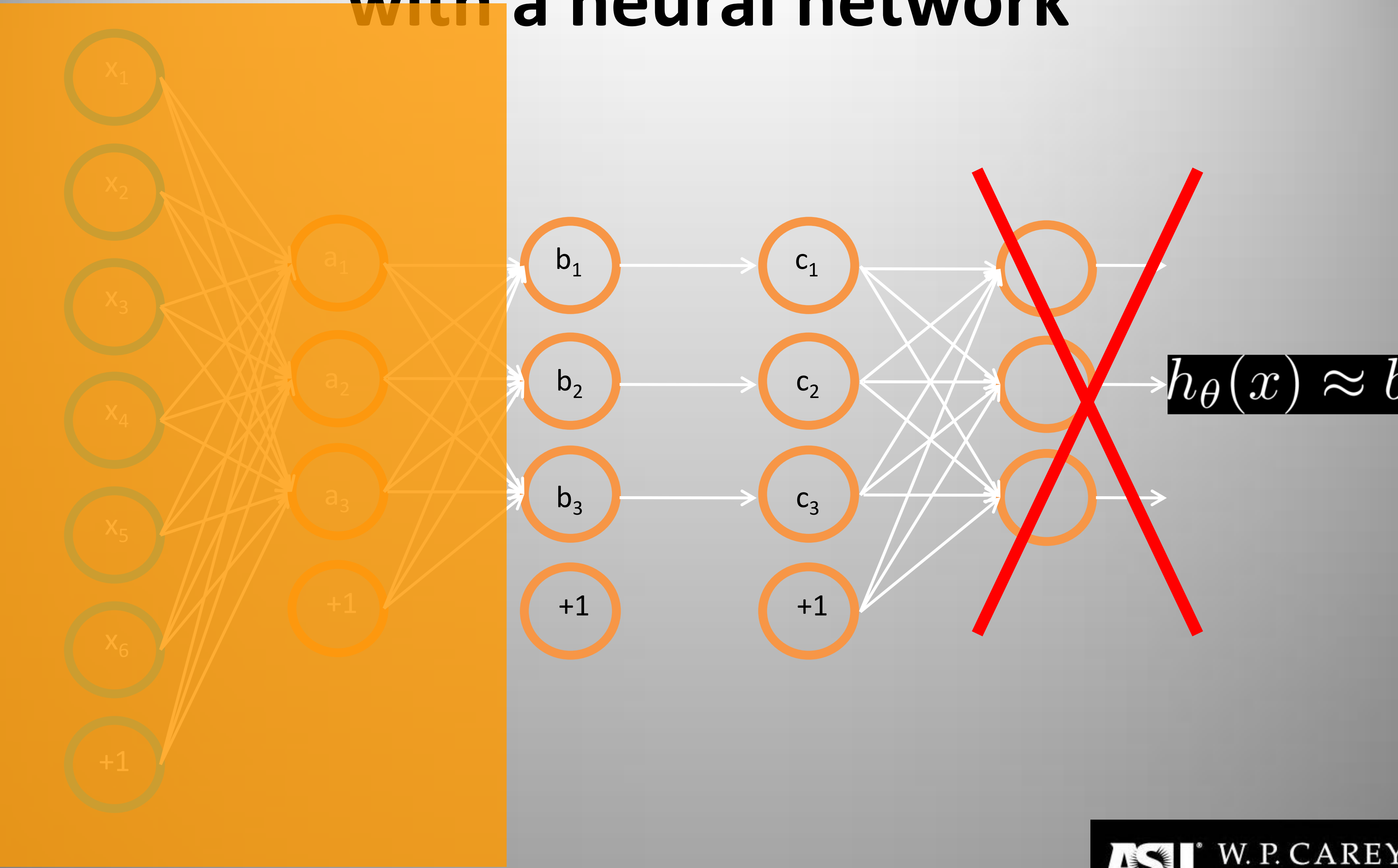
New representation for input.

Unsupervised feature learning with a neural network

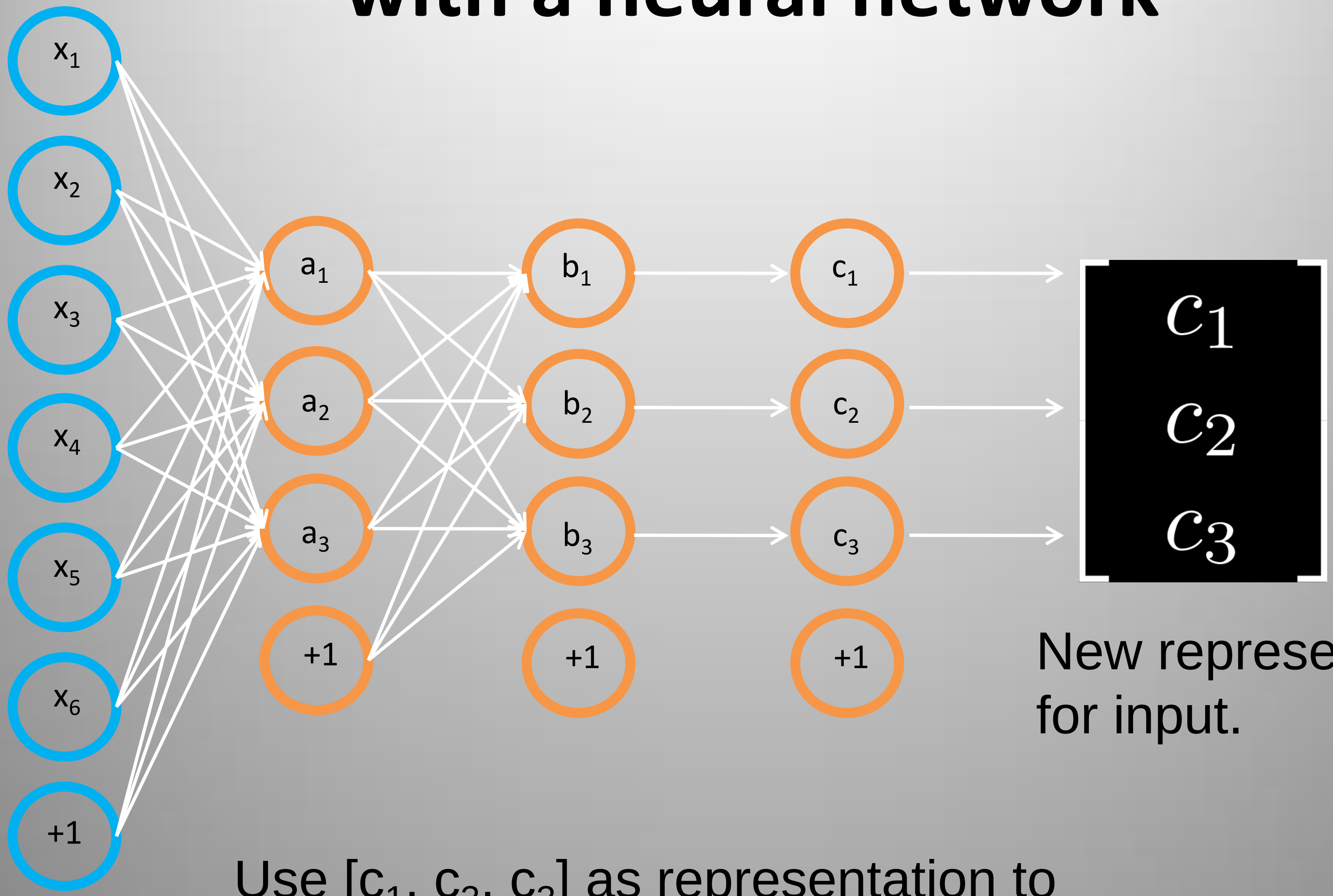


ASU W. P. CAREY SCHOOL of BUSINESS ARIZONA STATE UNIVERSITY

Unsupervised feature learning with a neural network



Unsupervised feature learning with a neural network



New representation for input.

Use [c_1, c_3, c_3] as representation to feed to learning algorithm.

Different types of artificial neural networks

- Autoencoder
- **DNN Deep neural network** & Deep learning
- **MLP Multilayer perceptron**
- RNN (Recurrent neural network)
- RBM Restricted Boltzmann machine
- SOM (Self-organizing map)
- Convolutional neural network

Convolutional Networks

About CNN's

- CNN's are neurobiologically motivated by the findings of locally sensitive and orientation-selective nerve cells in the visual cortex.

- They designed a network structure that implicitly extracts relevant features.

- Convolutional Neural Networks are a special kind of multi-layer neural networks.

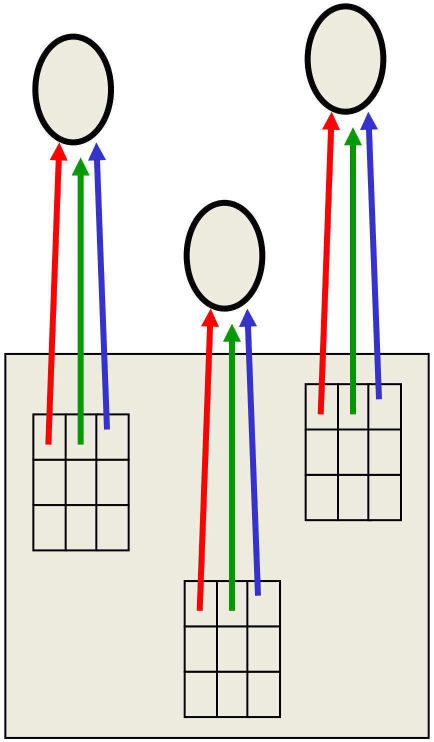
The replicated feature approach

(currently the dominant approach for neural networks)

- Use many different copies of the same feature detector with different positions.
  - Replication greatly reduces the number of free parameters to be learned.

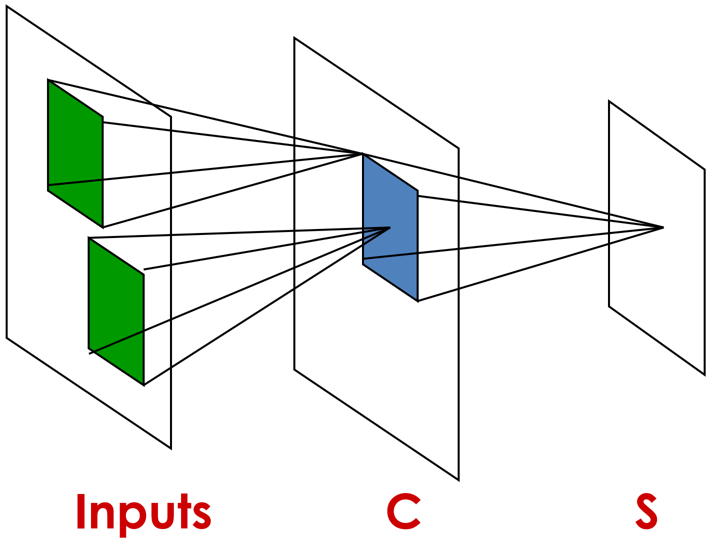
- Use several different feature types, each with its own map of replicated detectors.
  - Allows each patch of image to be represented in several ways.

The red connections all have the same weight.

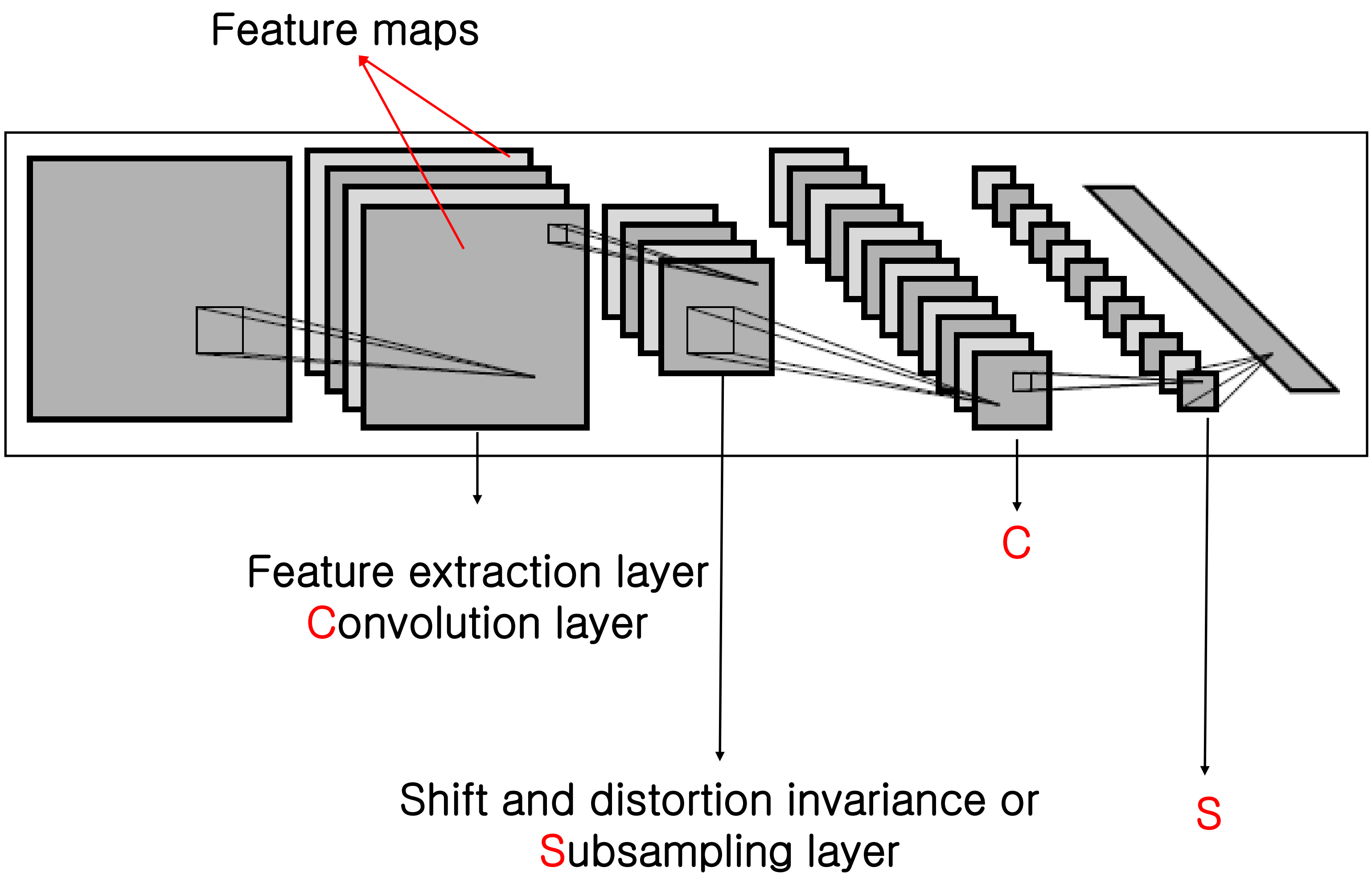


Feature extraction

- Shared weights: all neurons in a feature share the same weights (but not the biases).
- In this way all neurons detect the same feature at different positions in the input image.
- Reduce the number of free parameters.

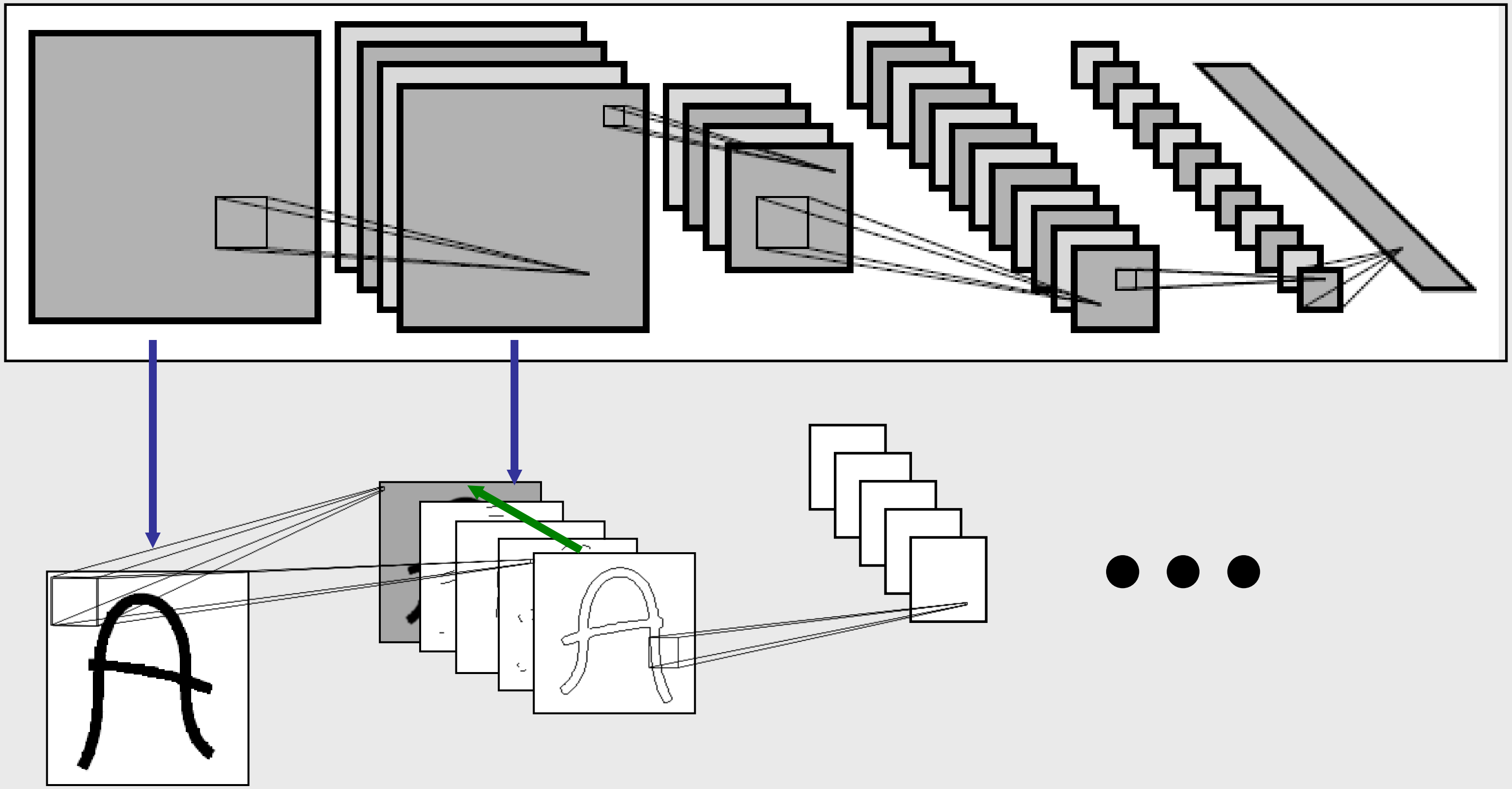


CNN’s Topology



Feature extraction

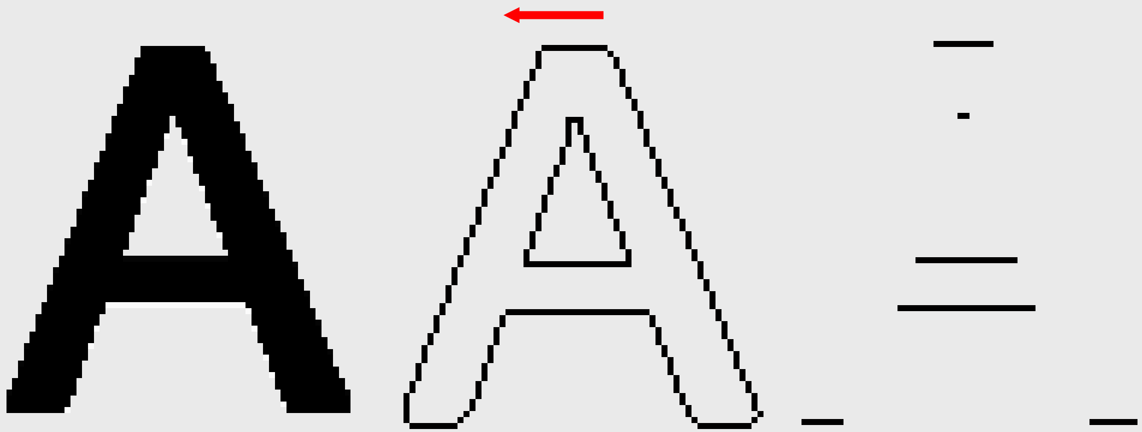
- If a neuron in the feature map fires, this corresponds to a match with the template.



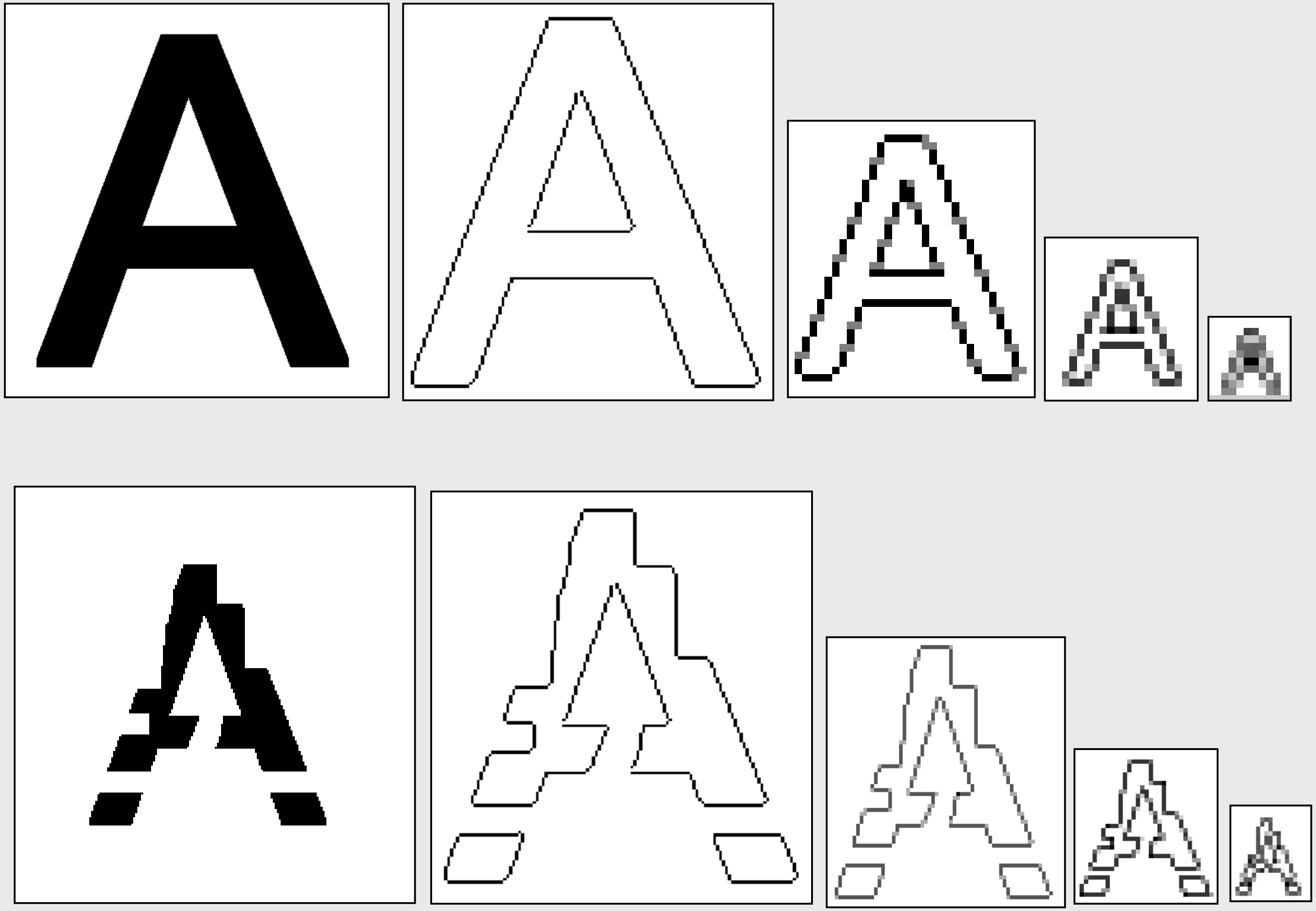
Subsampling layer

- the subsampling layers reduce the spatial resolution of each feature map

- By reducing the spatial resolution of the feature map, a certain degree of shift and distortion invariance is achieved.



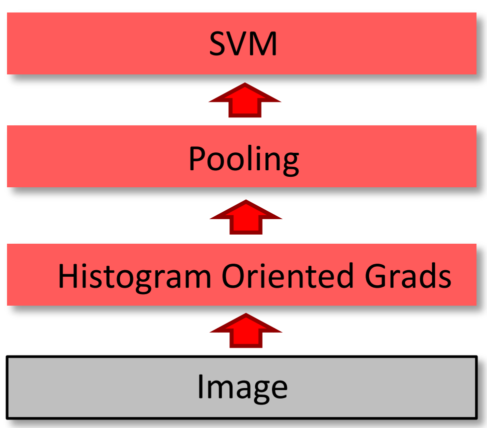
Subsampling layer



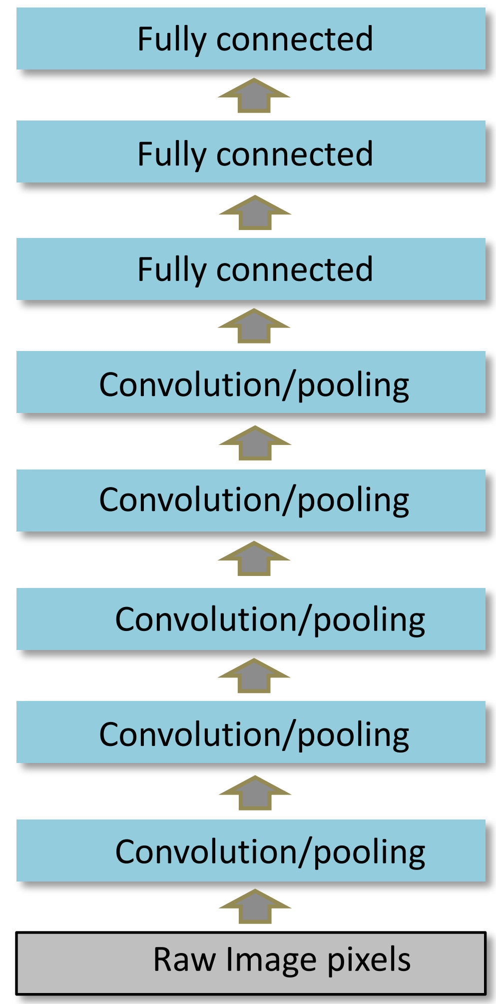
Deep **Convolutional NN** for Image Recognition

CNN: local connections with weight sharing;
pooling for translation invariance

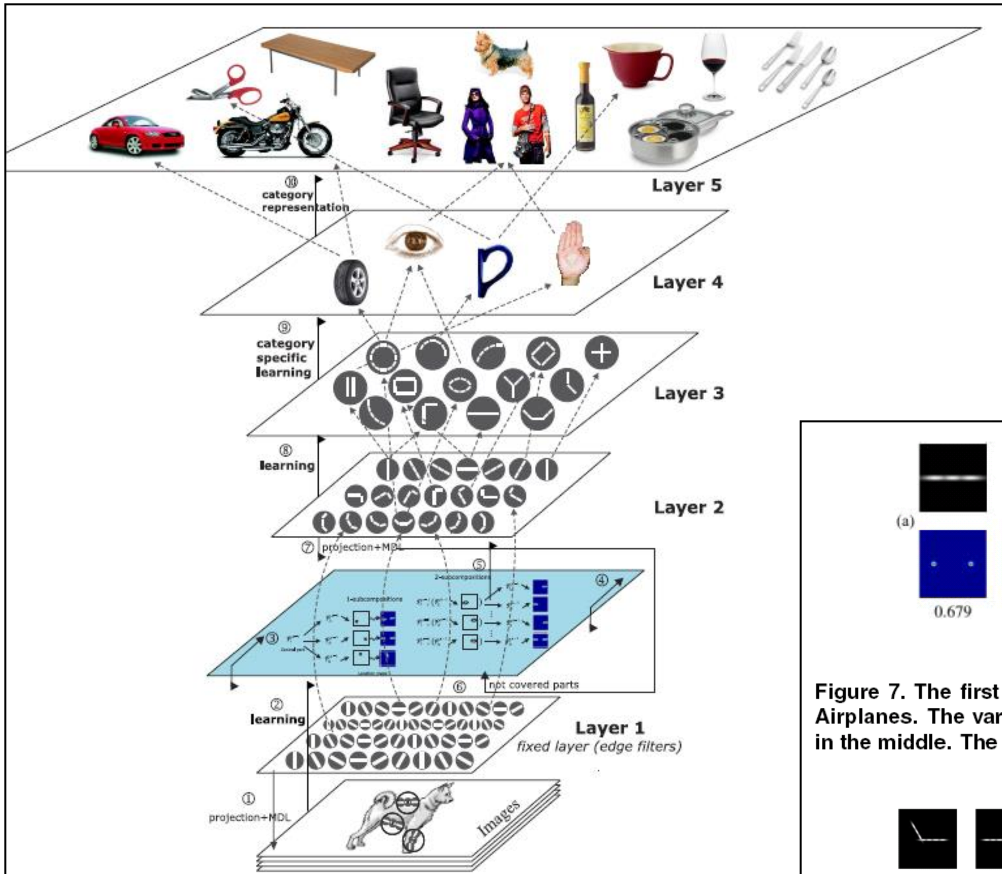
earlier



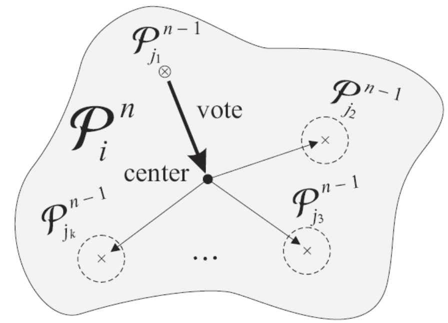
2012-2013



Learning a Compositional Hierarchy ture



The architecture



Parts model

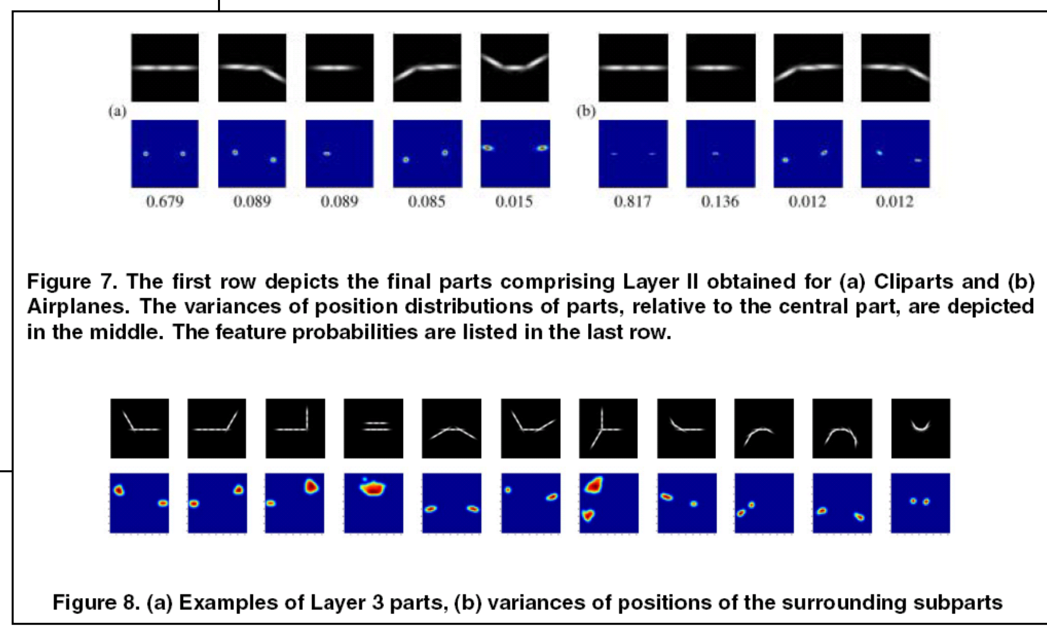


Figure 7. The first row depicts the final parts comprising Layer II obtained for (a) Cliparts and (b) Airplanes. The variances of position distributions of parts, relative to the central part, are depicted in the middle. The feature probabilities are listed in the last row.

Figure 8. (a) Examples of Layer 3 parts, (b) variances of positions of the surrounding subparts

Learned parts