DS501: Deep Learning

Prof. Randy Paffenroth

rcpaffenroth@wpi.edu

Worcester Polytechnic Institute

Announcements

- The grader just finished her part of the grading for Case Study 3.
  - Grades will be posted soon.
- Case Study 4 is due next class!
  - I am excited to hear all of the 90 second pitches :-)
- The final exam is the next class!
- Anything I can help with?
  - We will have a review session today.

Today

- I plan to spend an hour talking about deep learning
  - I know that several of you expressed interest in this topic :-)
  - And this is one of the areas in which my students and I do research.
- Then I plan to spend the rest of the class reviewing for the final.
  - We covered a lot of stuff in the second half of the class and I want to make that we have time to answer all of your questions!

But first, you get to grade me!

- I really appreciate your feedback on the class.
  - It is best way for me to improve the class!
- In particular, there is a place on the form for **written comments**, and I deeply appreciate whatever you have to say!

Feature Engineering

- We have talked about this a few times.
  - Perhaps the data you are given in not the best data for making the predictions you want to make.
- The last question on Case Study 3 was a prime example.
  - How do you find features that make prediction easier?

Slides inspired by Roelof Pieters, KTH & Data Science

Is this problem ubiquitous?

Coming up with features is difficult, time-consuming, requires expert knowledge. "Applied machine learning" is basically feature engineering.

— Andrew Ng, Machine Learning and AI via Brain simulations

https://en.wikipedia.org/wiki/Feature_engineering

Feature engineering

- How do you do it? What are examples?

x^2 x^3 ---

sum results => features

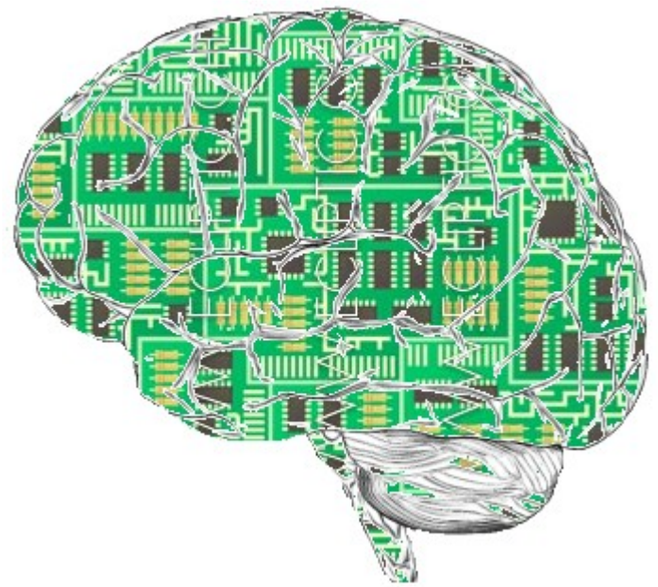
PCA
Factor analysis
normalization

ICA

external data
- dictionaries

Deep learning

- Deep learning (deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to **model high-level abstractions** in data by using **multiple processing layers**, with complex structures or otherwise, composed of **multiple non-linear transformations.**



https://en.wikipedia.org/wiki/Deep_learning

Deep learning: More specifically

1) Deep learning is a **collection** of different techniques, not a single technique.

2) Deep learning aims to automatically learn feature hierarchies from data.

- As opposed to “by hand” feature engineering.

3) Generally, deep learning is based on artificial neural networks.

- Though that is not always true, depending on what an “artificial neural network” really is.

Is deep learning supervised or unsupervised?

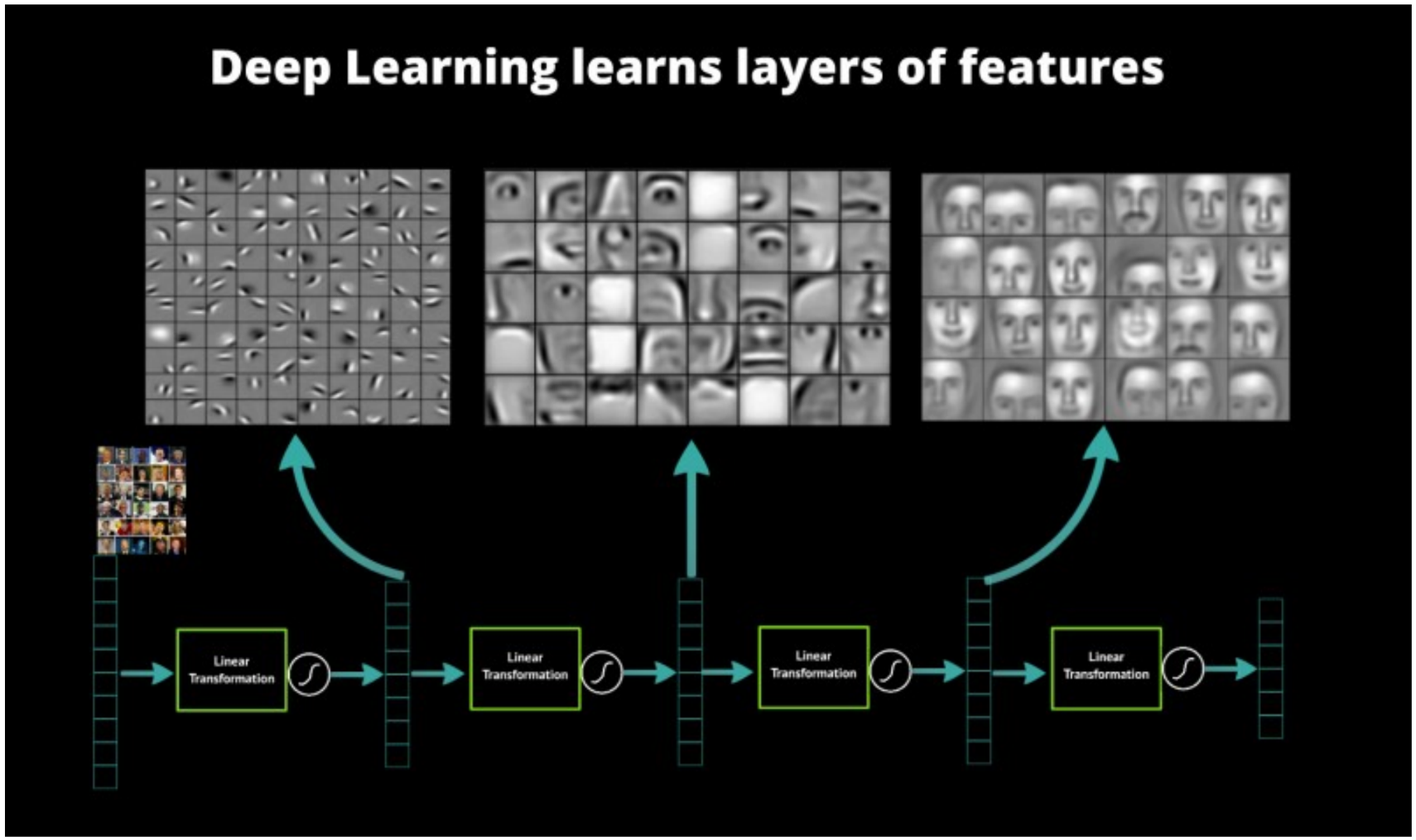
Both!

Joe detection learning

About Faces is unsupervised

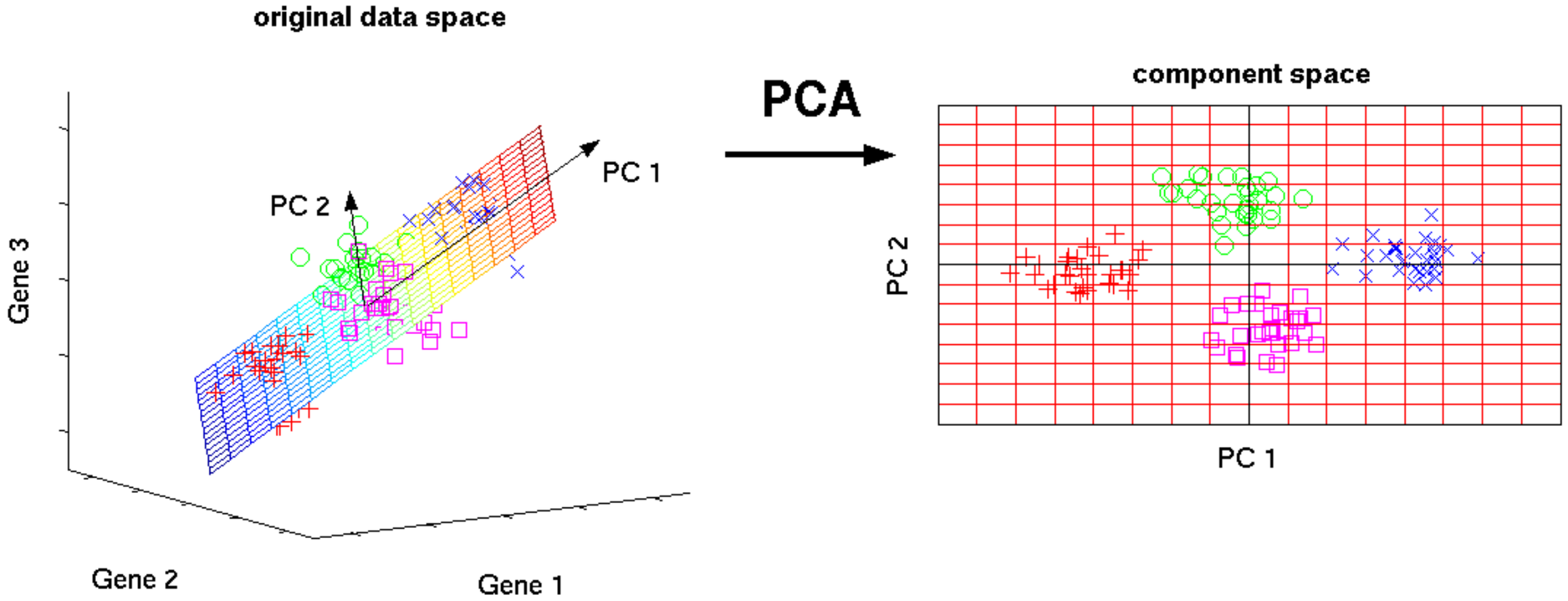
learning about Joes Face is supervised

The idea in one picture

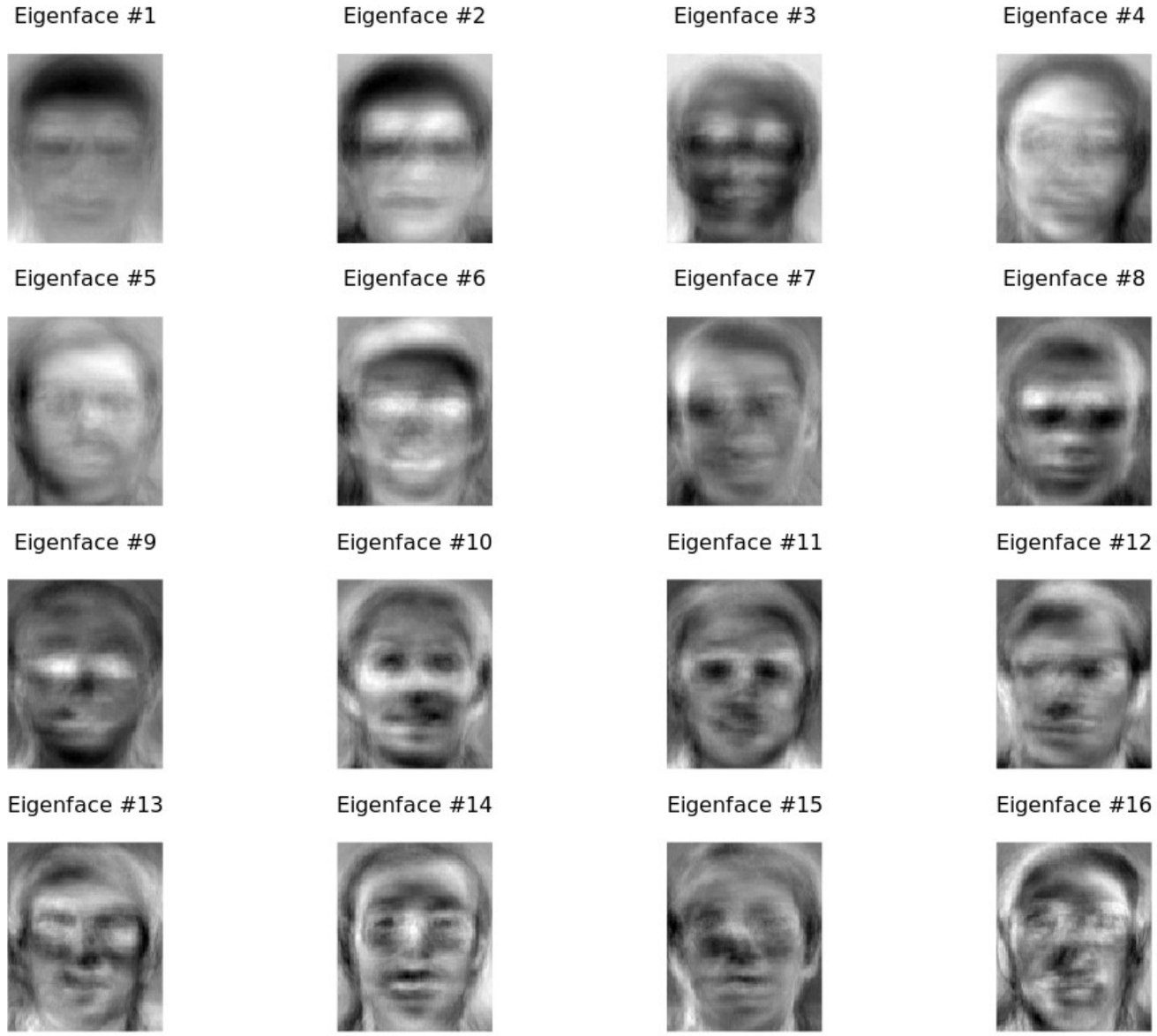


https://www.datarobot.com/blog/a-primer-on-deep-learning/

Probably looks familiar

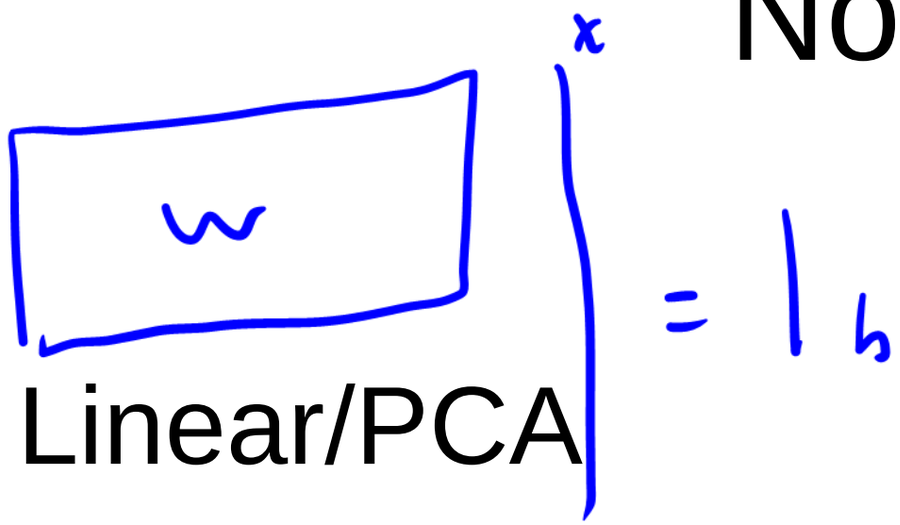


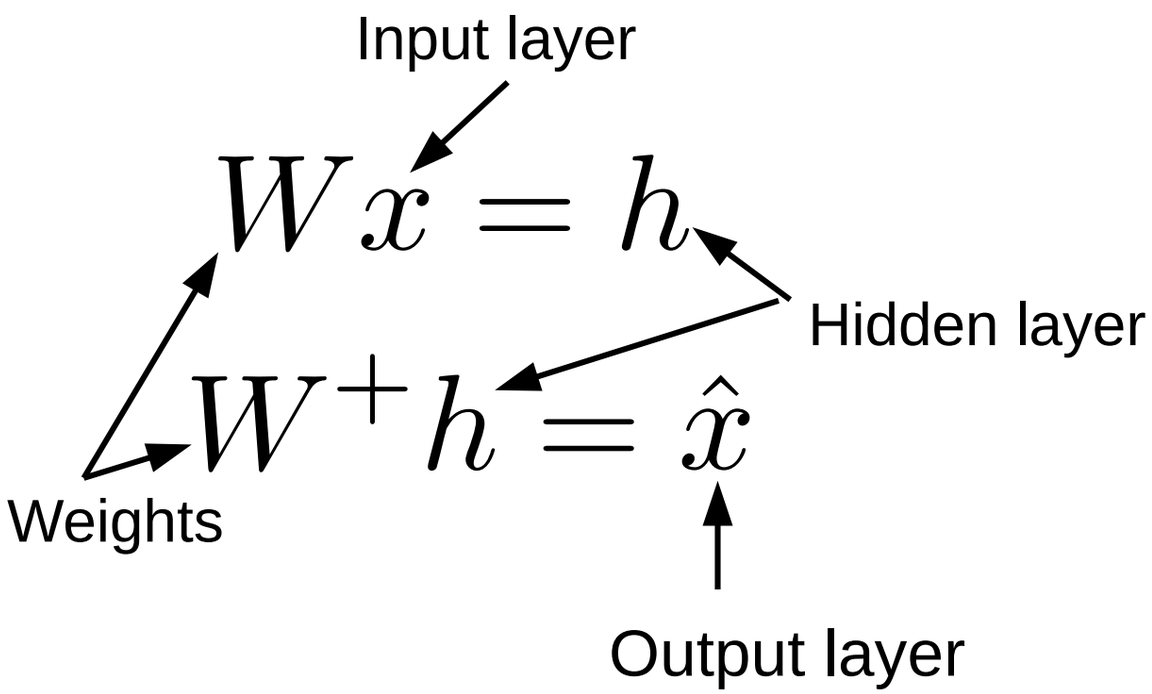
Ok, we learned this already, right?

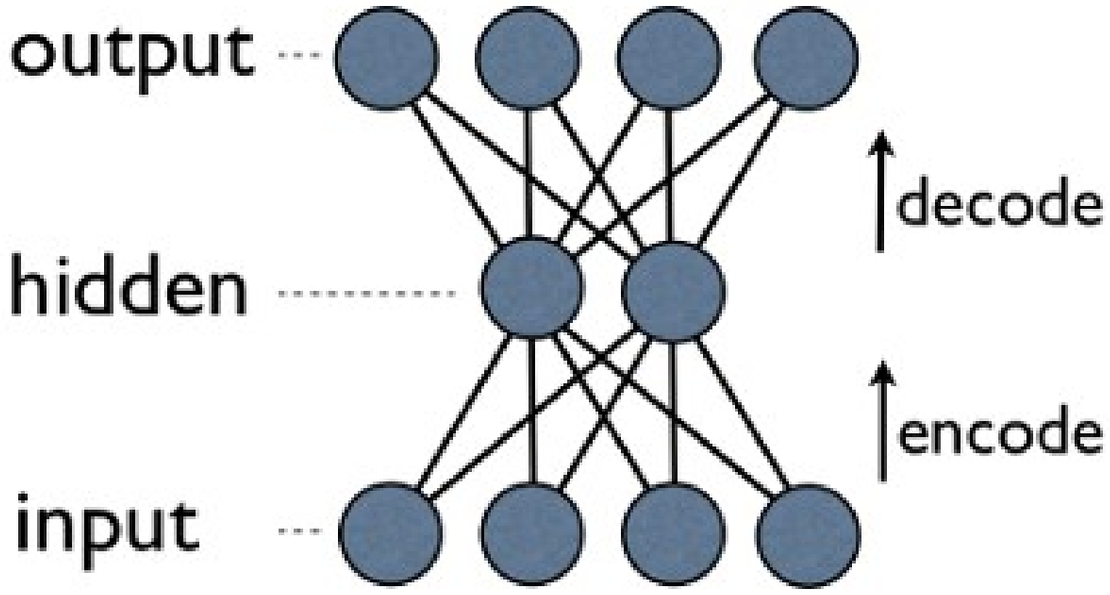


http://bytefish.de/static/images/blog/eigenfaces/subplot_eigenfaces.png

Not quite...


x
w
= h
Linear/PCA


Input layer
$Wx = h$
Hidden layer
$W^+h = \hat{x}$
Weights
Output layer


output
decode
hidden
encode
input

Non-linearity

Non-linear layer

Input layer

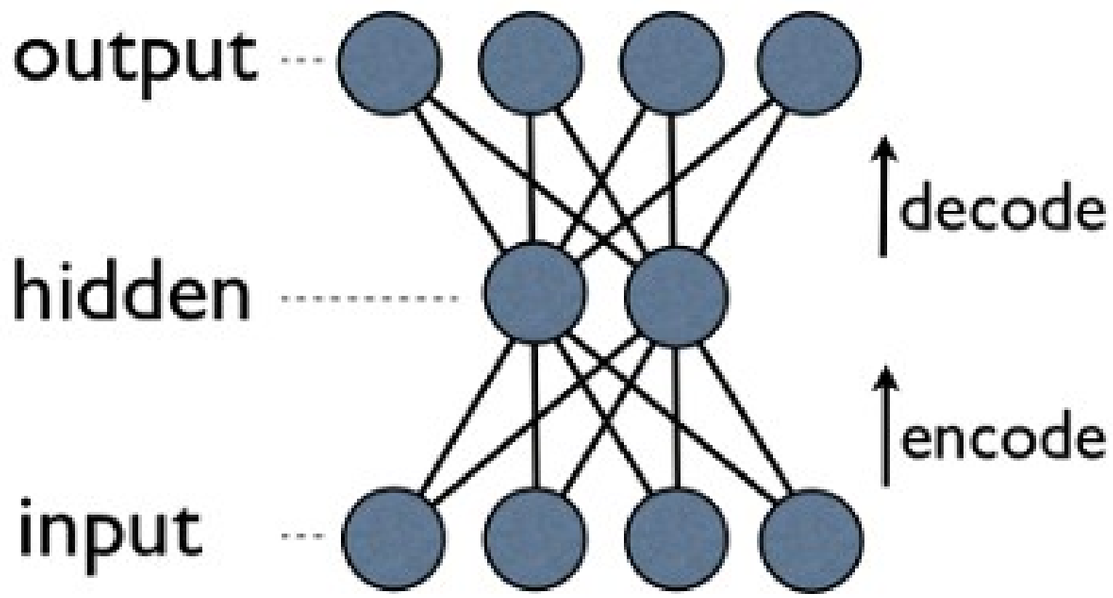
$$f(Wx + b) = h$$

Weights

Hidden layer

$$f(W^T h + b) = \hat{x}$$

Output layer

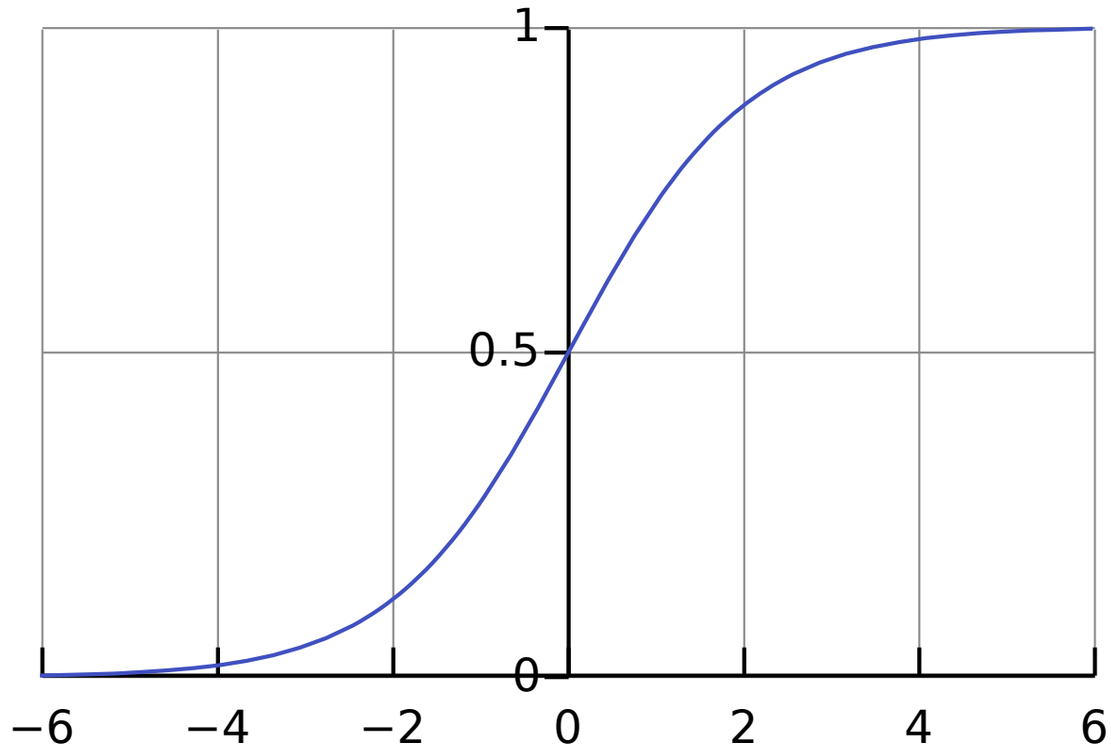


What is f ?

Logistic Regression

Sigmoid (is one popular choice)!

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



Huh!? Why sigmoid?

- One reason – The Universal Approximation Theorem

Let $\varphi(\cdot)$ be a nonconstant, bounded, and monotonically-increasing continuous function. Let I_m denote the *m*-dimensional unit hypercube $[0, 1]^m$. The space of continuous functions on I_m is denoted by $C(I_m)$. Then, given any function $f \in C(I_m)$ and $\varepsilon > 0$, there exists an integer N, real constants $v_i, b_i \in \mathbb{R}$ and real vectors $w_i \in \mathbb{R}^m$, where $i = 1, \cdots, N$, such that we may define:

$$F(x) = \sum_{i=1}^{N} v_i \varphi\left(w_i^T x + b_i\right)$$

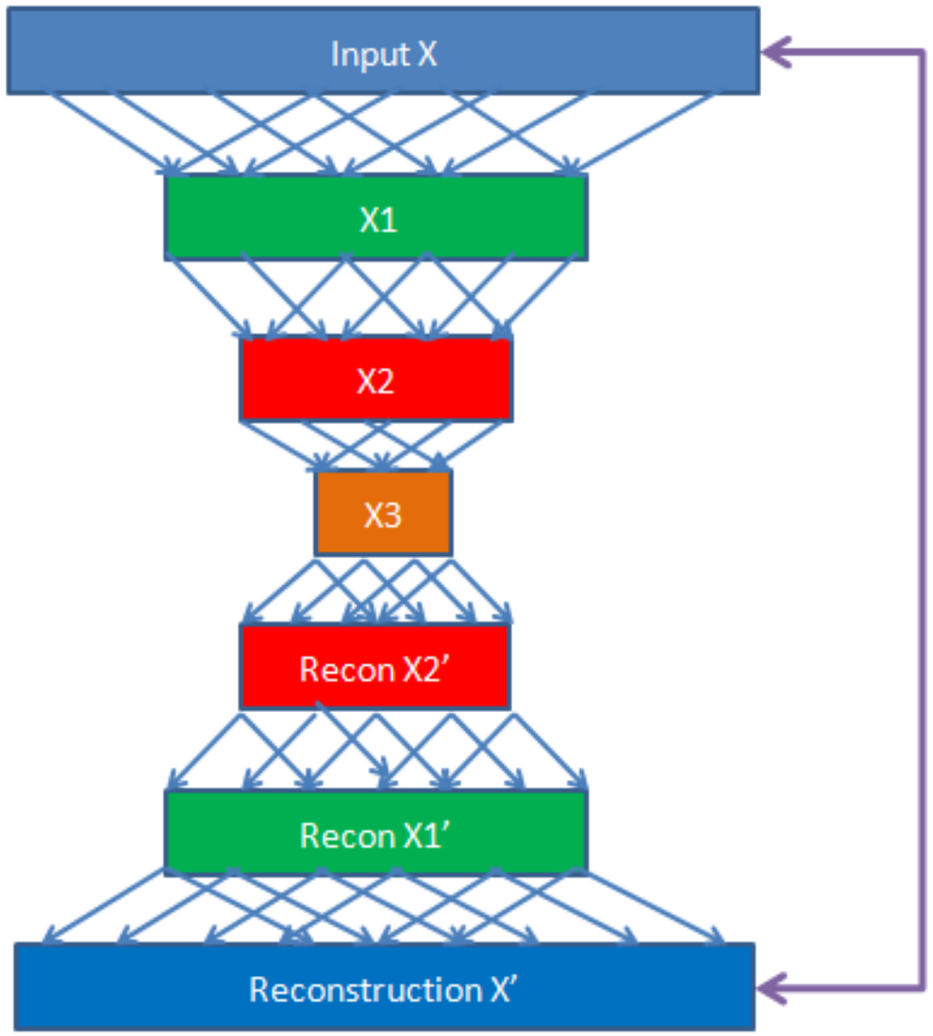
as an approximate realization of the function f where f is independent of φ; that is,

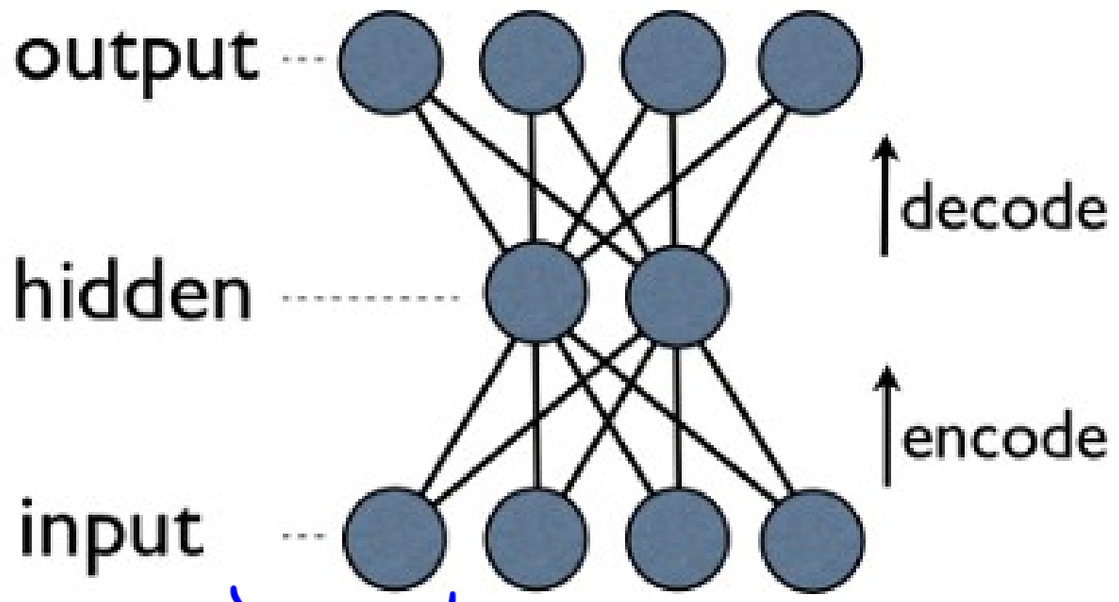
$$|F(x) - f(x)| < \varepsilon$$

for all $x \in I_m$. In other words, functions of the form $F(x)$ are dense in $C(I_m)$.

https://en.wikipedia.org/wiki/Universal_approximation_theorem

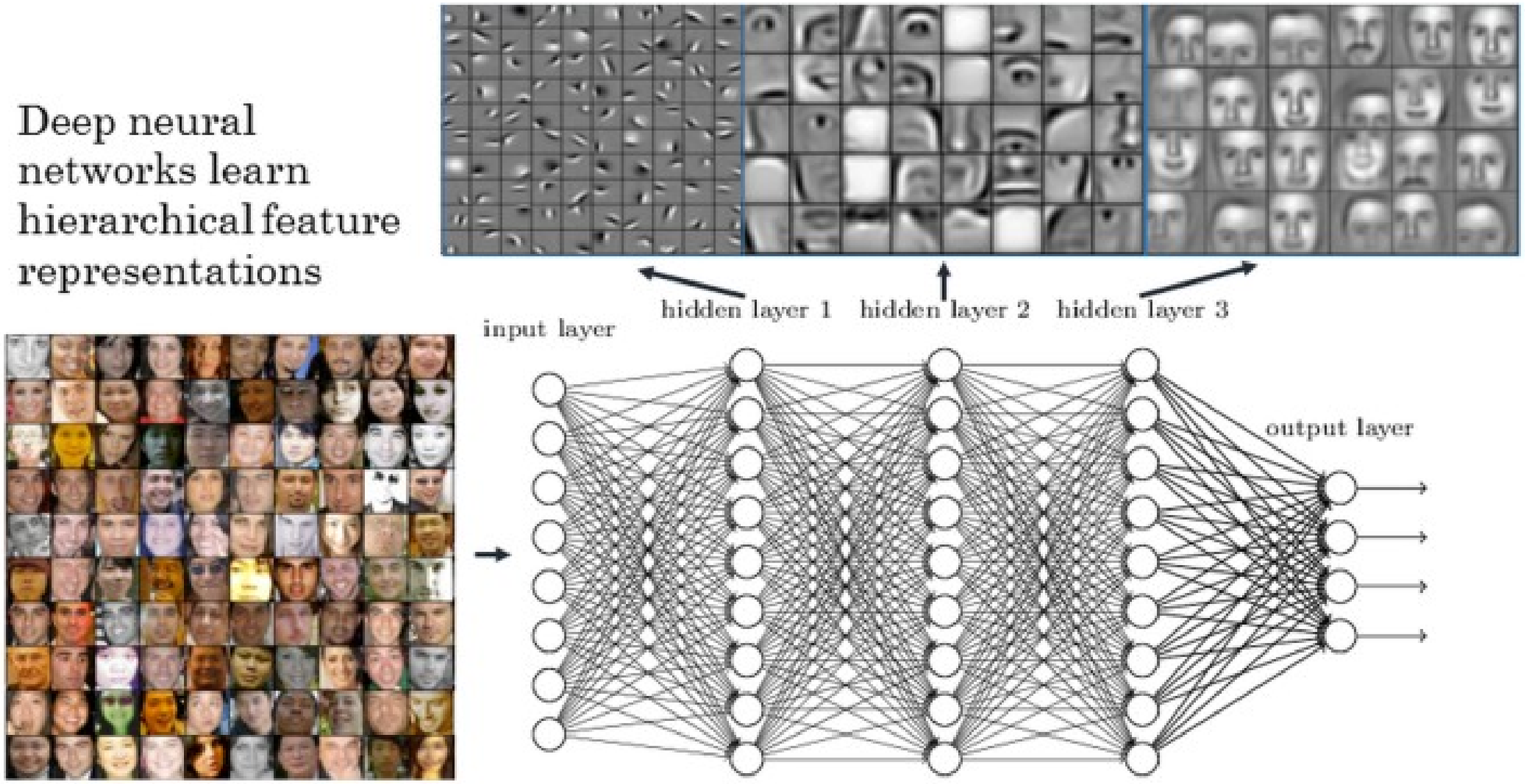
What makes it “deep” learning deep?





Low train layer
By Layer – cheap
Low also train all
at once - expensive

Just do it multiple times!

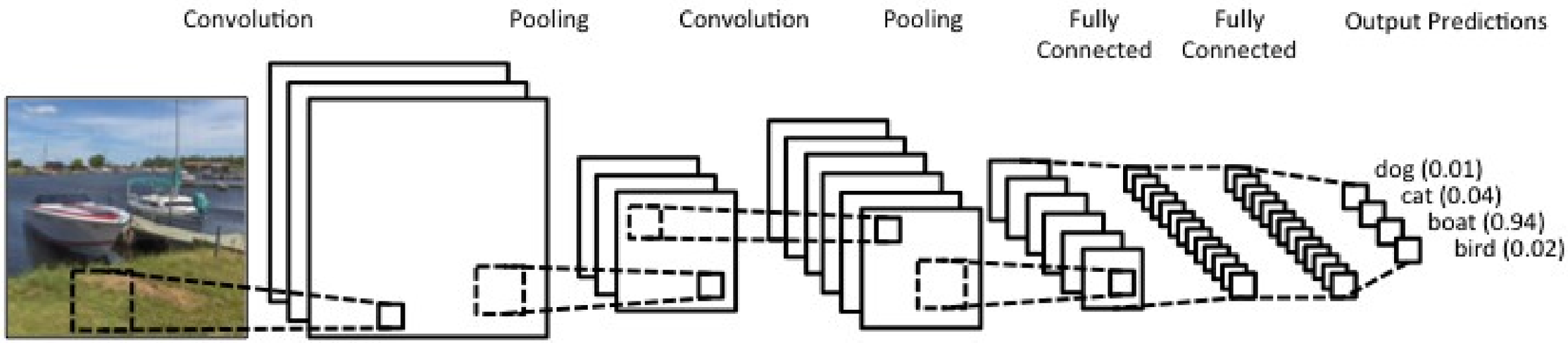


https://nivdul.wordpress.com/2015/11/17/exploring-deep-learning-with-li-zhe/

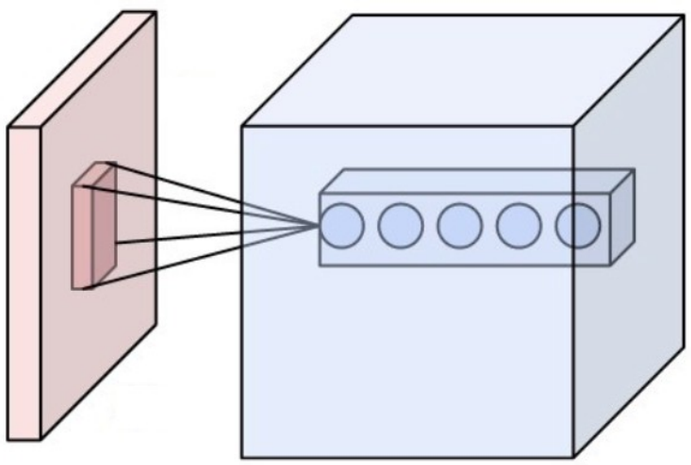
Different types of Deep Learning architectures (there are many :-)

- Convolutional Neural Network (CNN)
- Recurrent Neural Network
- Deep auto-encoders

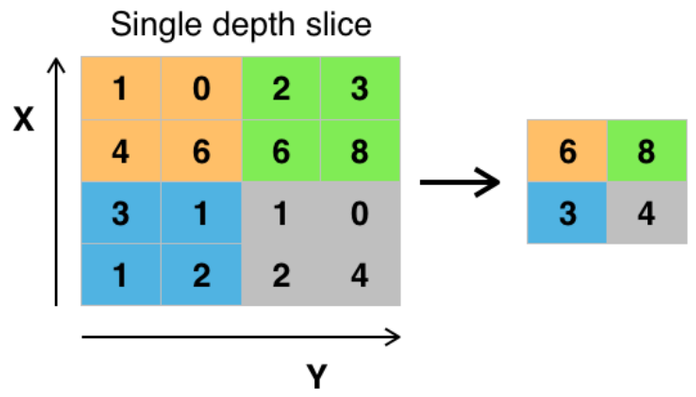
Convolutional Neural Network



https://www.clarifai.com/technology

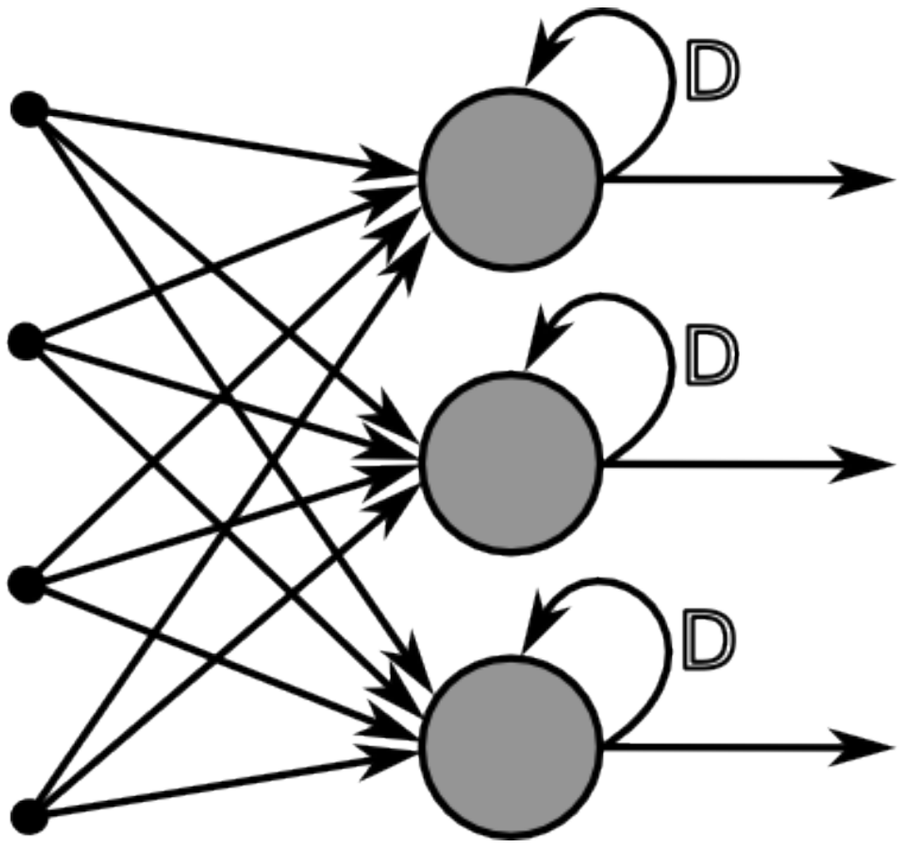


By Aphex34 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=45659236



By Aphex34 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=45673581

Recurrent Neural Network



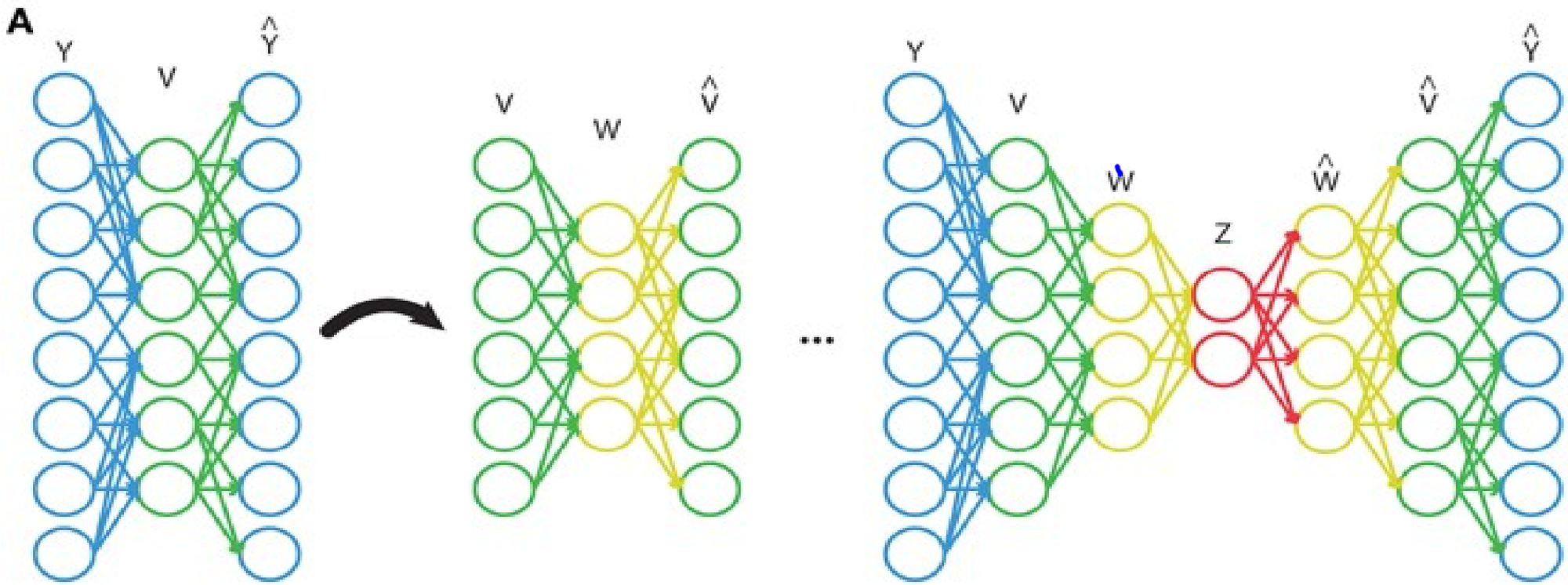
Used for time dependent things such as finance!

By Chrislb - created by Chrislb, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=224513

Auto-encoders… the one I am most interested in.

- Did you ever run a compression algorithm on an already compressed file?
  - That is the idea :-)

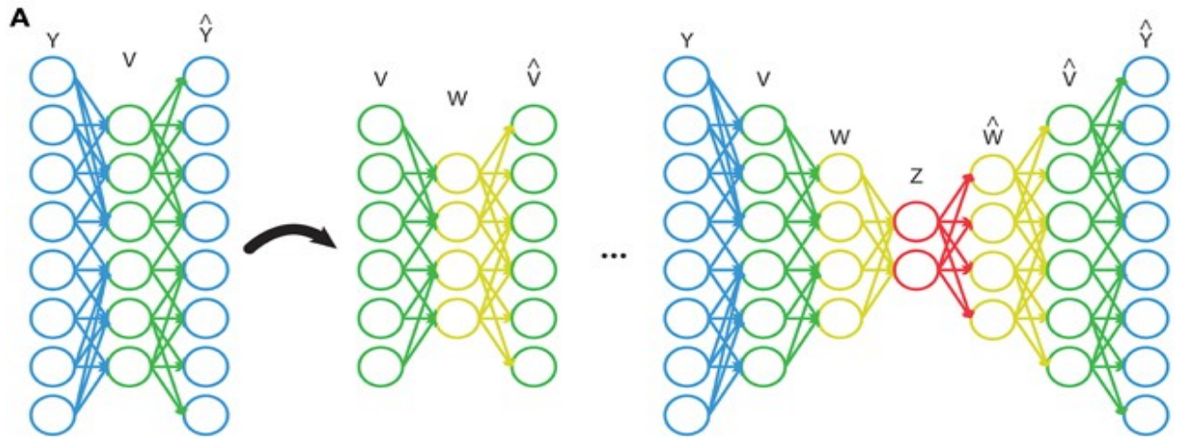
Auto-encoders… the one I am most interested in.



Nice example

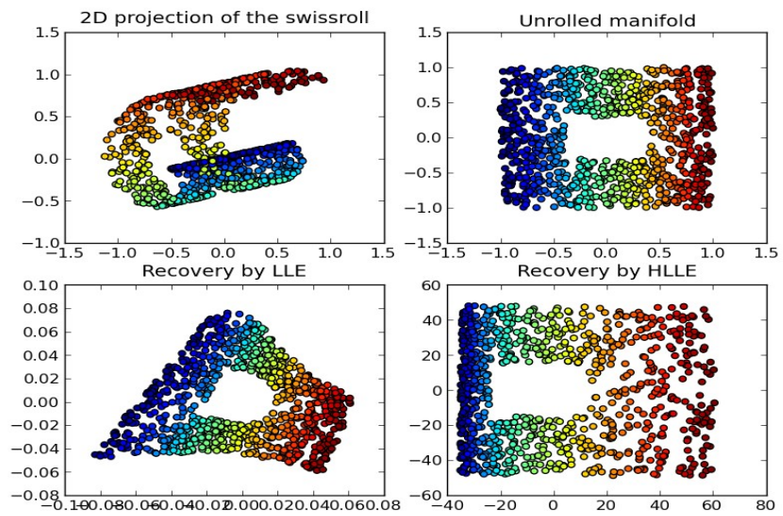
http://acsweb.ucsd.edu/~wfedus/pdf/courses/291_assignment3.pdf

Leading edge… research I am interested in!



$$Wx = h \quad \text{v.s.} \quad f(Wx + b) = h$$

v.s.



"Lle hlle swissroll" by Olivier Grisel - Generated using the Modular Data Processing toolkit and matplotlib.. Licensed under Creative Commons Attribution 3.0 via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Lle_hlle_swissroll.png#mediaviewer/File:Lle_hlle_swissroll.png

Python Libraries

- Low level
  - Theano (http://deeplearning.net/software/theano/)
- High level
  - Pylearn2 (http://deeplearning.net/software/pylearn2/)
  - Keras (http://keras.io/)
  - Lasagne (http://lasagne.readthedocs.org/en/latest/user/tutorial.html#before-we-start)
    - Get it :-)?