האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

Artificial Intelligence in Medicine

Neural networks and real-life data

Nir Friedman and Tommy Kaplan

28/2/24

Why neural networks?

Why neural networks?

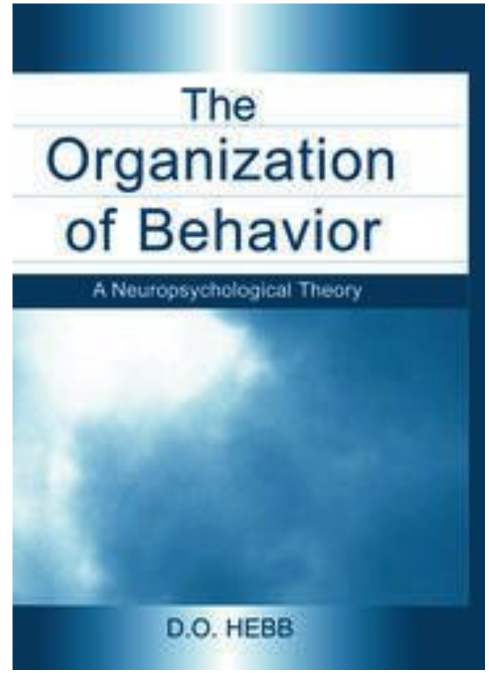
BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE
IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.



1949

J. Physiol. (1959) 148, 574–591

RECEPTIVE FIELDS OF SINGLE NEURONES IN
THE CAT'S STRIATE CORTEX

BY D. H. HUBEL* AND T. N. WIESEL*

*From the Wilmer Institute, The Johns Hopkins Hospital and
University, Baltimore, Maryland, U.S.A.*

J. Physiol. (1962), **160**, *pp.* 106–154

RECEPTIVE FIELDS, BINOCULAR INTERACTION
AND FUNCTIONAL ARCHITECTURE IN
THE CAT'S VISUAL CORTEX

BY D. H. HUBEL AND T. N. WIESEL

*From the Neurophysiology Laboratory, Department of Pharmacology
Harvard Medical School, Boston, Massachusetts, U.S.A.*

Why neural networks?

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

J. Physiol. (1959) 148, 574–591

RECEPTIVE FIELDS OF SINGLE NEURONES IN THE CAT'S STRIATE CORTEX

BY D. H. HUBEL* AND T. N. WIESEL*

From the Wilmer Institute, The Johns Hopkins Hospital and University, Baltimore, Maryland, U.S.A.

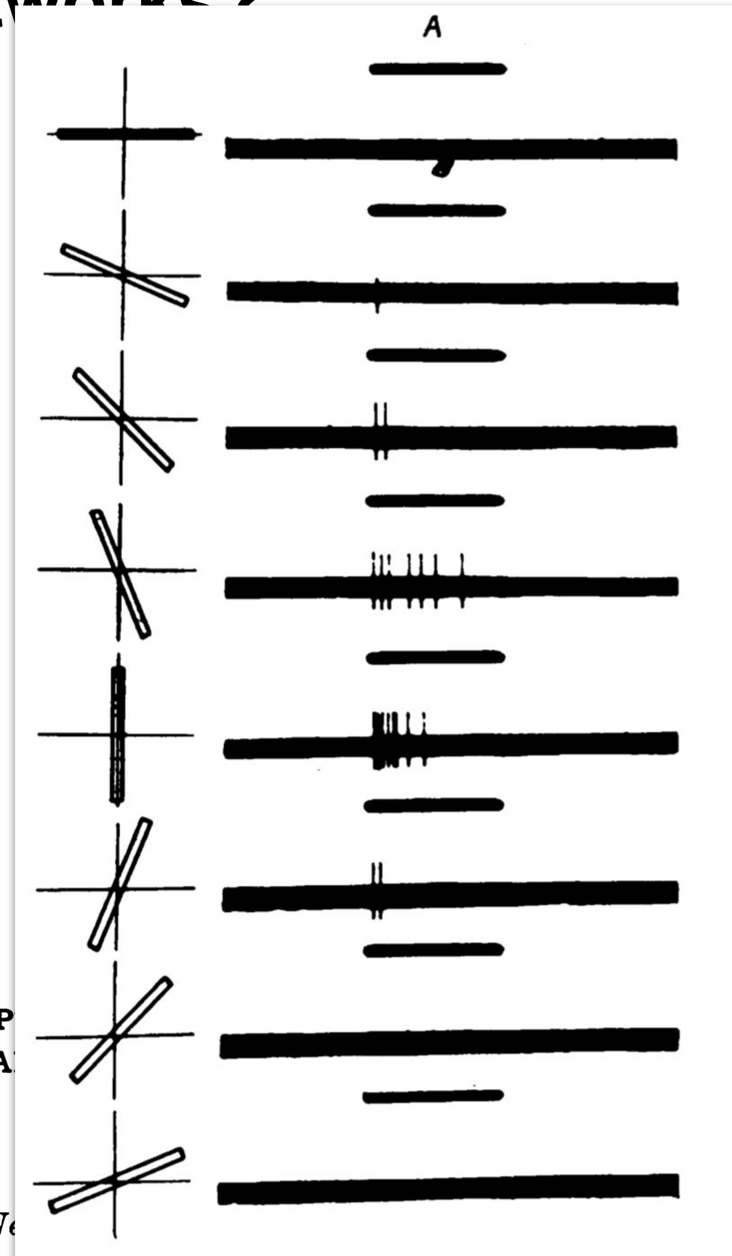
-154

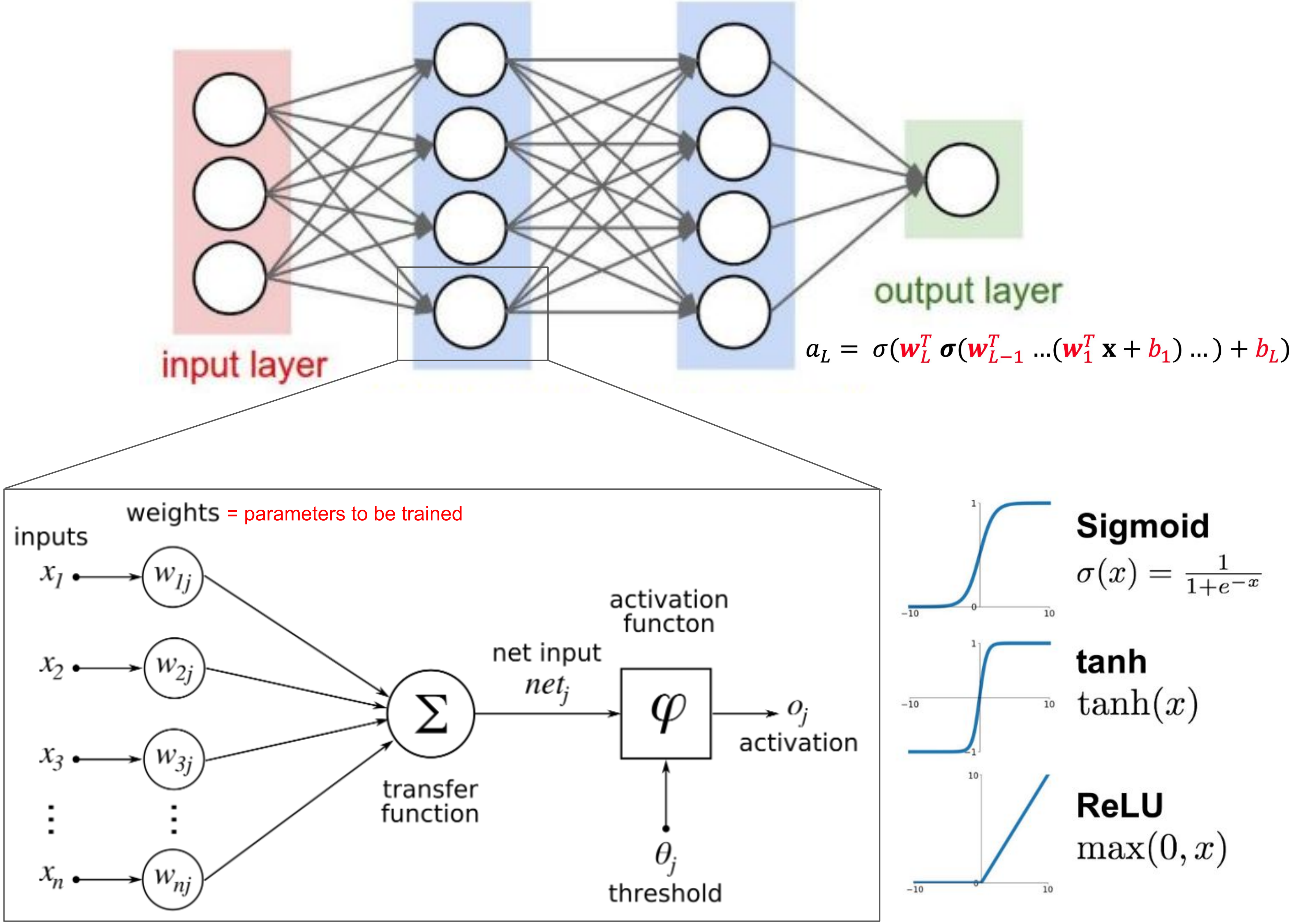
RECEP

A

From the Ne ... *gy*

Harva



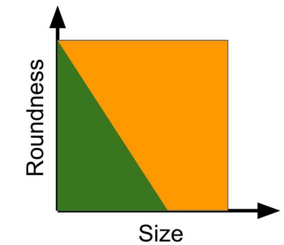

input layer
output layer
$a_L = \sigma(\boldsymbol{w}_L^T\, \boldsymbol{\sigma}(\boldsymbol{w}_{L-1}^T \ldots (\boldsymbol{w}_1^T\, \mathbf{x} + b_1) \ldots) + b_L)$
weights = parameters to be trained
inputs
x_1
x_2
x_3
x_n
w_{1j}
w_{2j}
w_{3j}
w_{nj}
Σ
transfer function
net input
net_j
activation functon
φ
o_j
activation
θ_j
threshold
Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$
tanh
$\tanh(x)$
ReLU
$\max(0, x)$

“Neuron” ≈ linear classifier

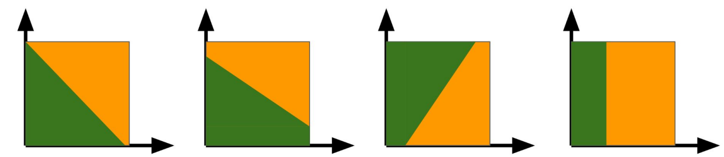
- Smoothed non-binary output
- No temporal dynamics
- Parameter learning ≈ Hebbian learning

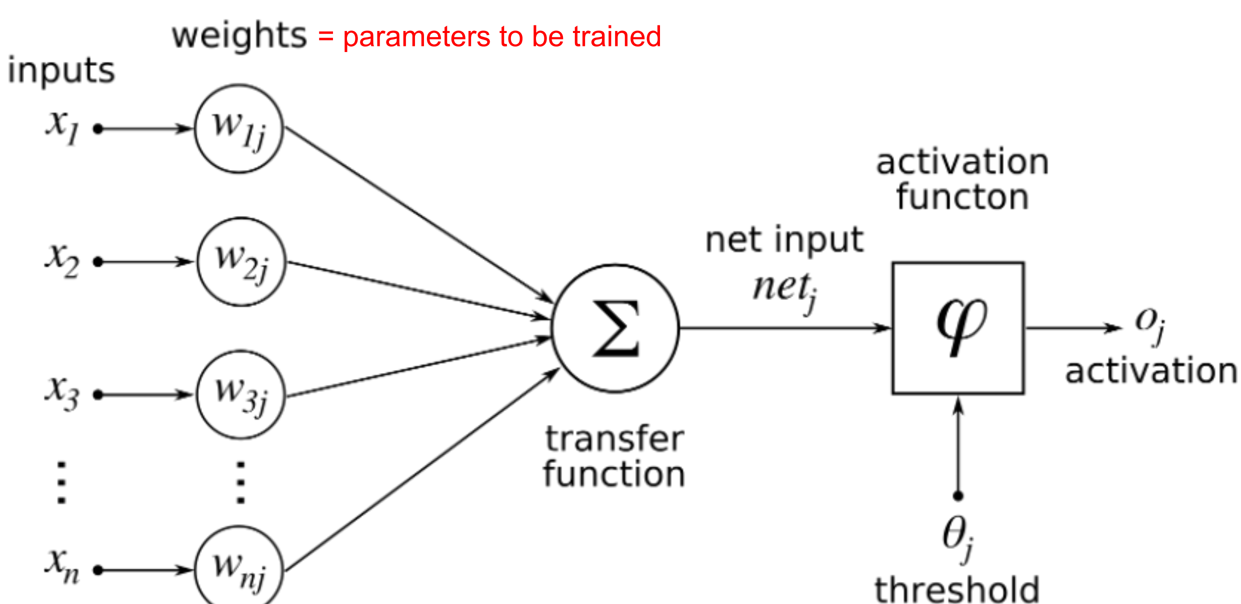
Linear classifier

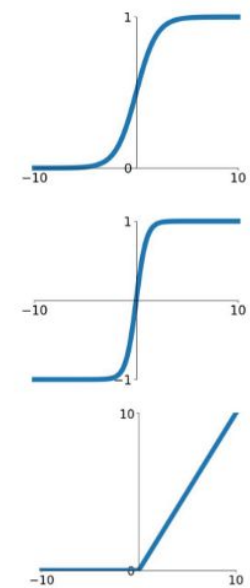
```
if 2*size + 5*round > 100 then
   “orange”
else
   “apple”
```



Wider range of hypotheses





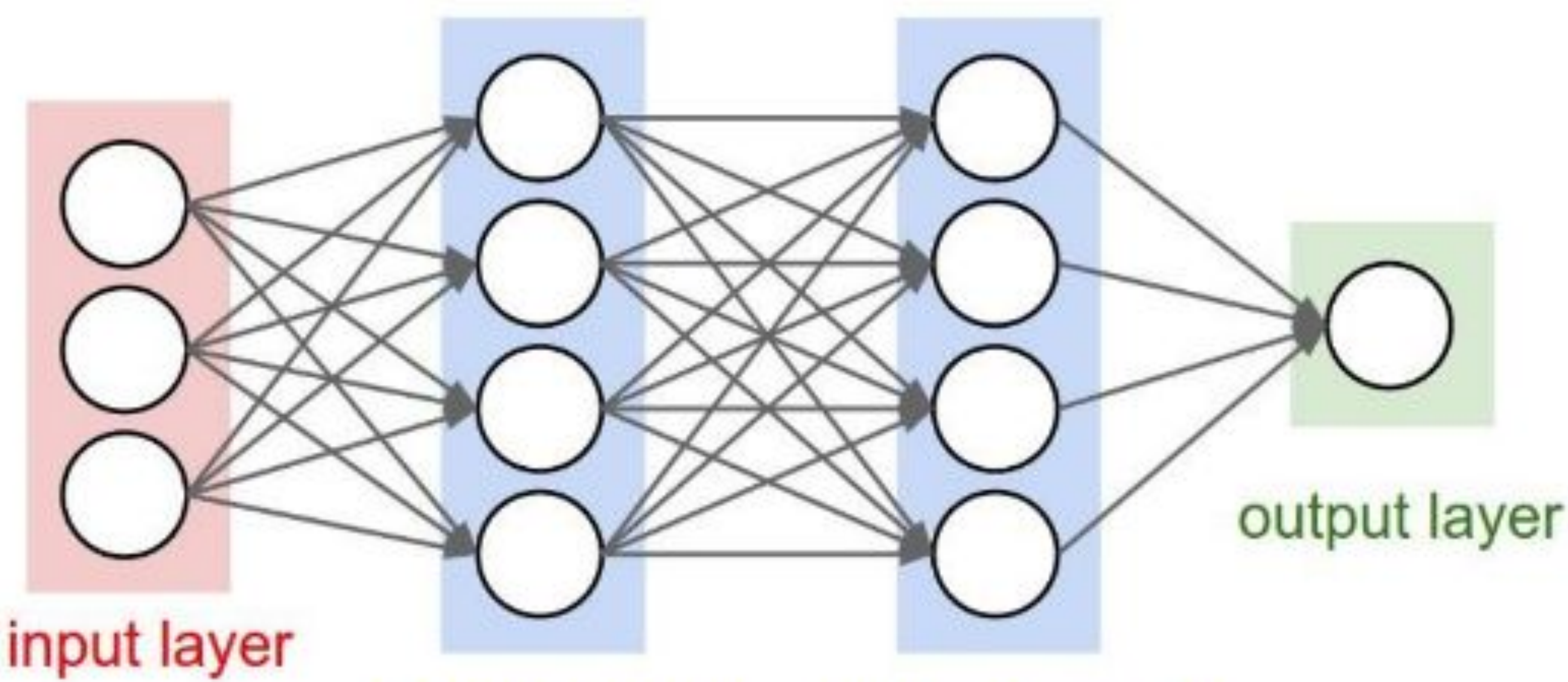


Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$

tanh
$\tanh(x)$

ReLU
$\max(0, x)$

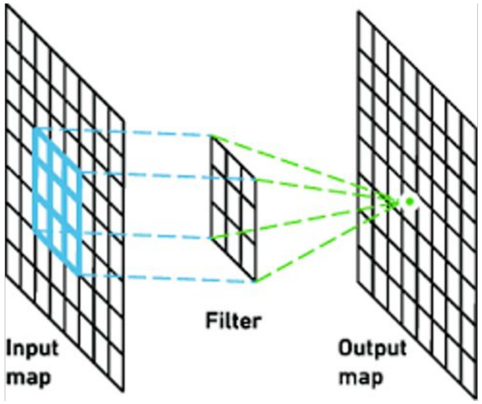
Fully connected network



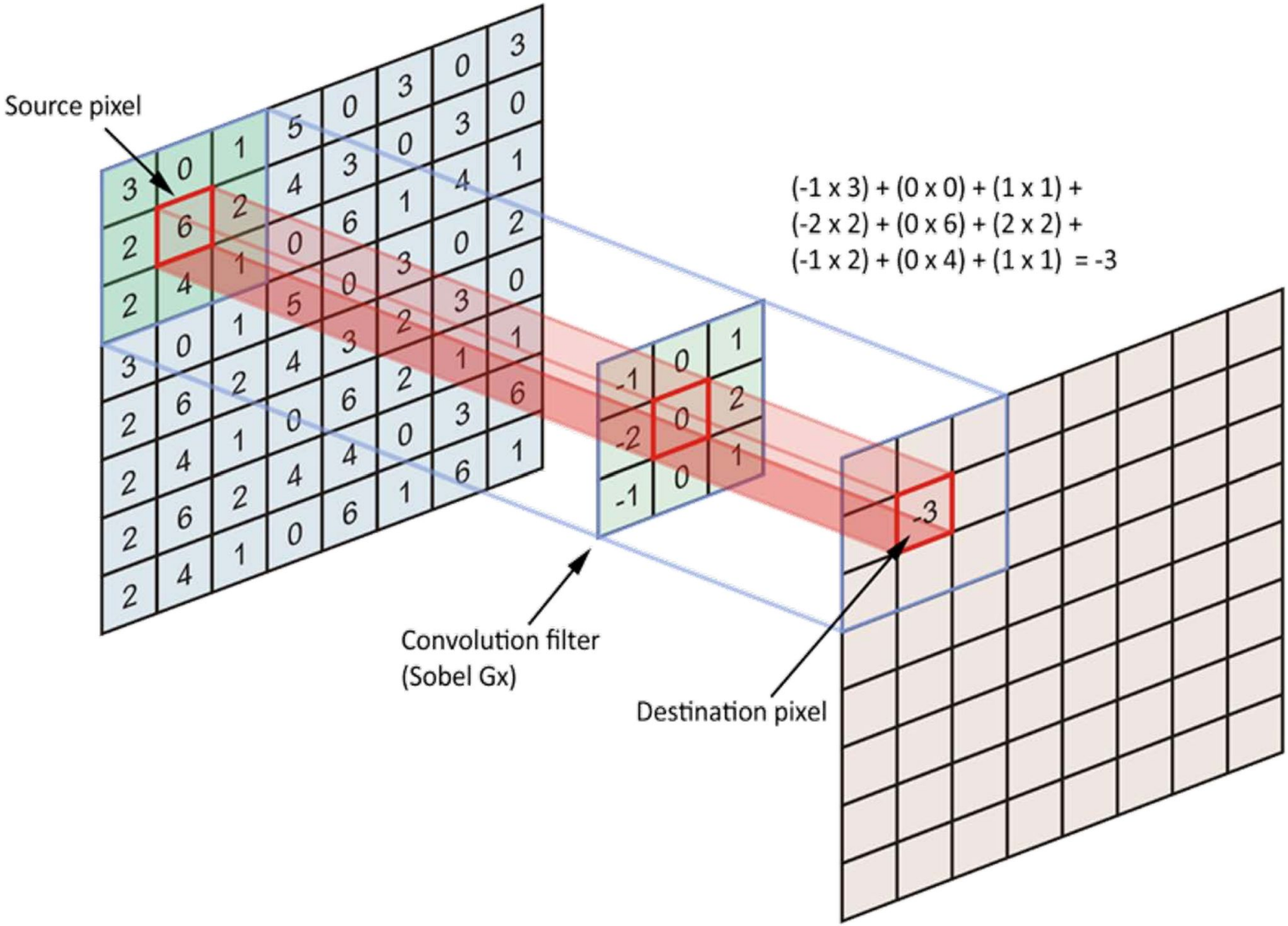
Convolutional neural network

Few changes from fully connected networks

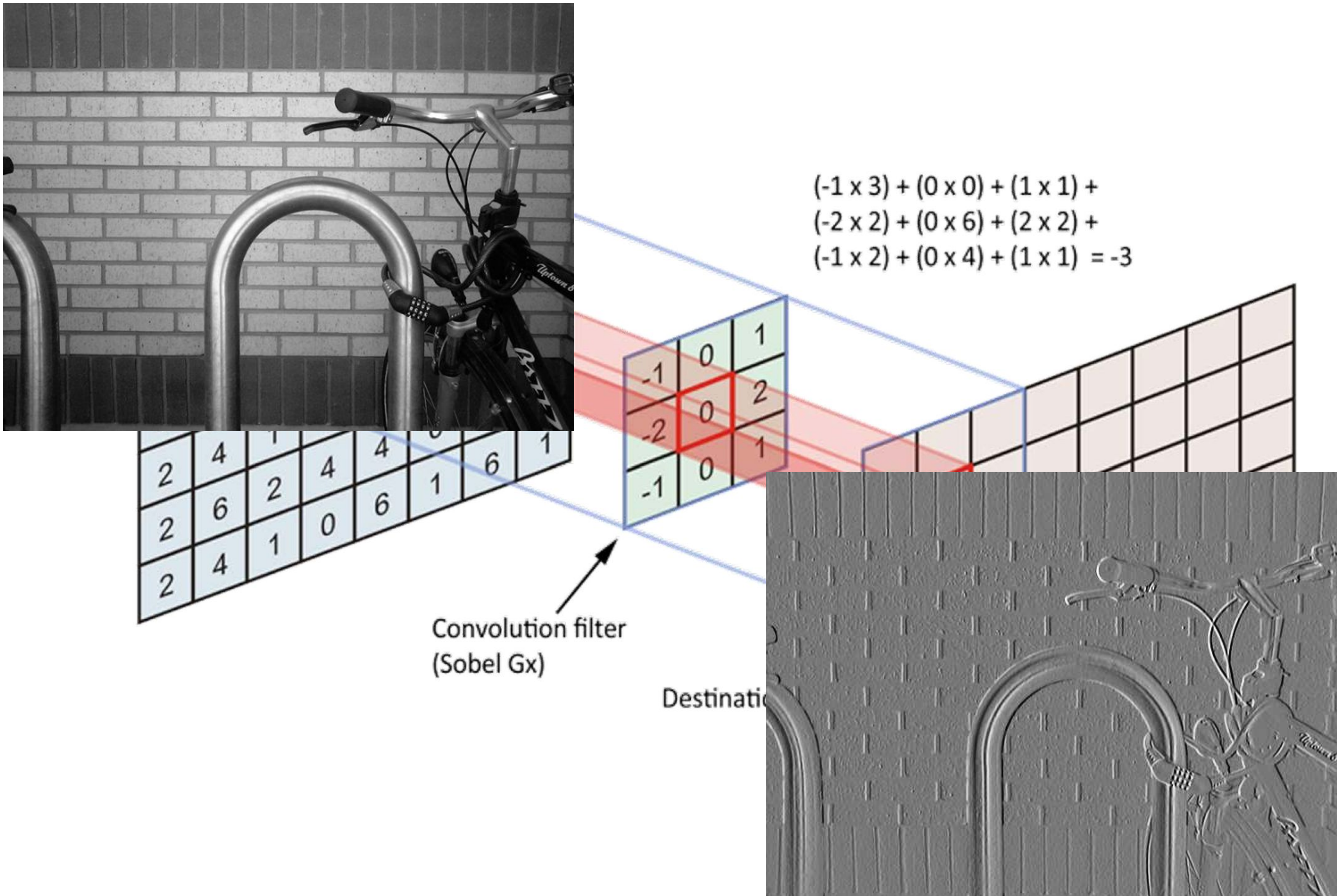
1. A neuron is not connected to all neurons (in prev. layer)
2. Keep it local
3. Use the same filter across all regions



Convolutional neural network - filters



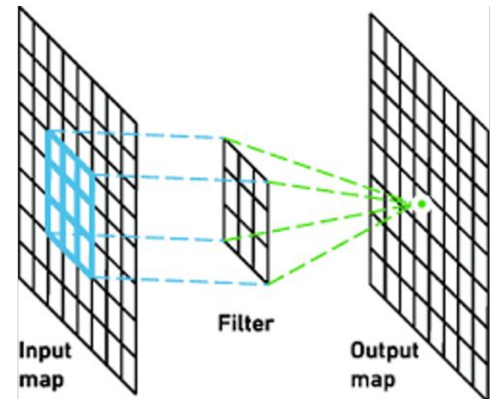
Convolutional neural network - filters

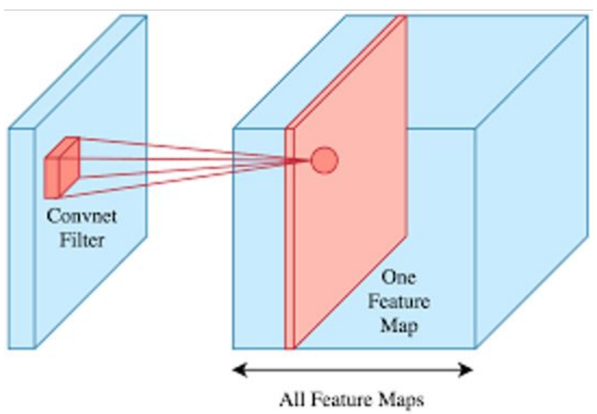


Convolutional neural network

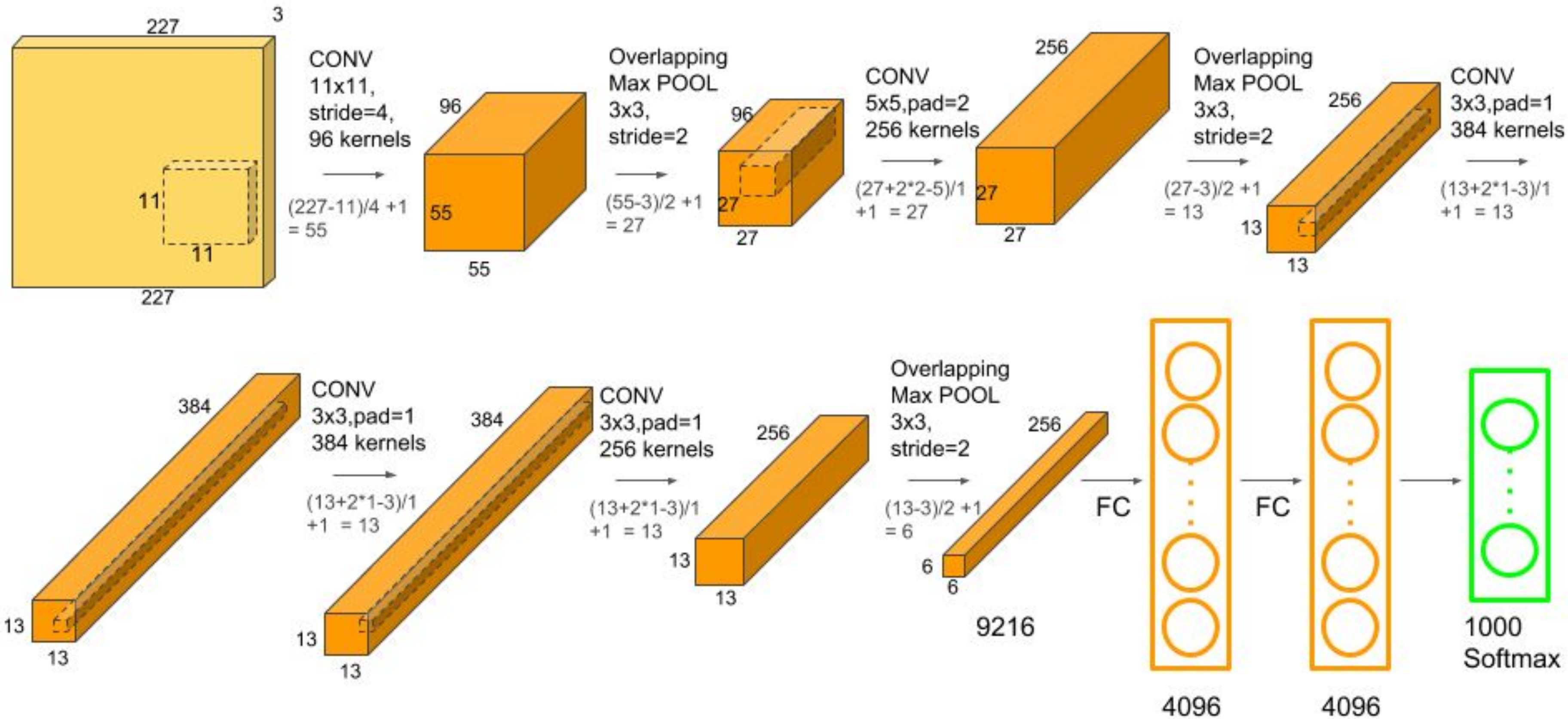
Few changes from fully connected networks

1. A neuron is not connected to all neurons (in prev. layer)
2. Keep it local
3. Use the same filter across all regions
4. Use more than one filter





Convolutional neural network - architecture



AlexNet, 2012

From engineering to learning to deep learning

Manual modeling

Features
intensity, texture
shape, location

MANUAL

Relations
Boundary
differences, …

MANUAL

Machine learning

Features
intensity, texture
shape, location

MANUAL

Relations
Derived by
regression, SVM,…

AUTOM

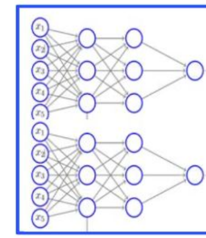
Deep learning

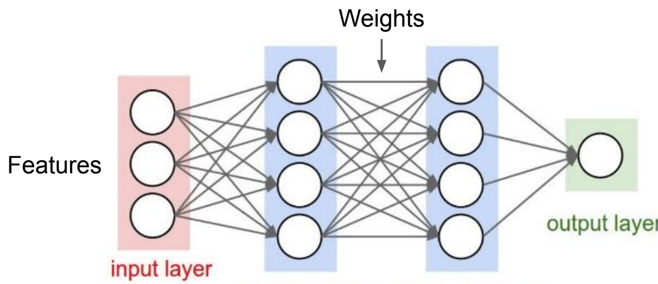
Features
Derived from data
Implicit

AUTOM

Relations
Derived from data
Implicit

AUTOM



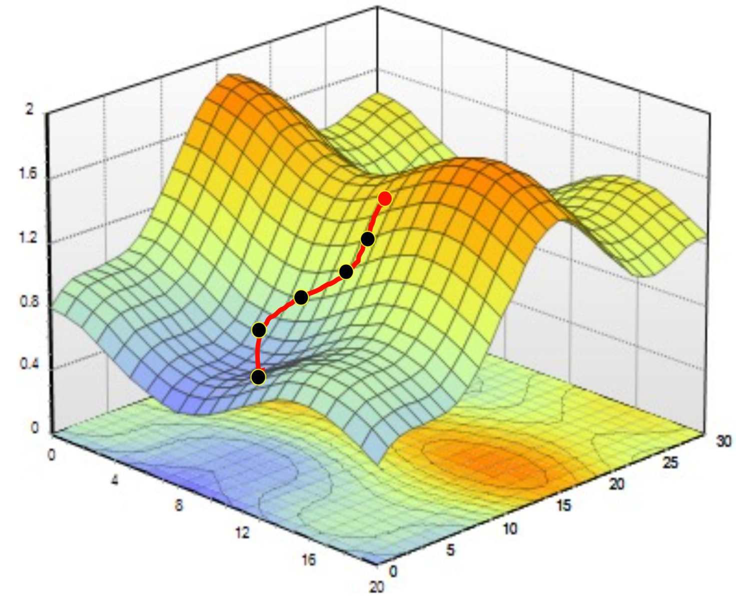


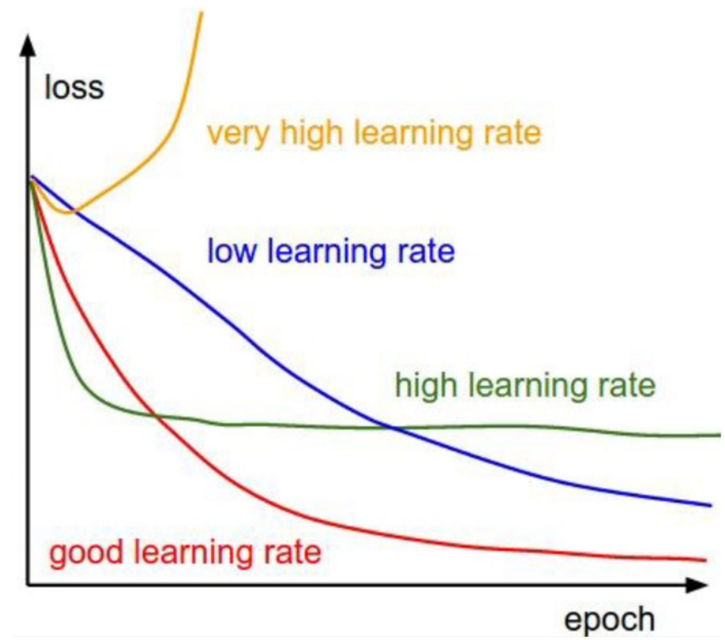
Stochastic Gradient Ascent

- Iteratively, approximate the direction at each optimization step using a small subset of samples (mini-batch)

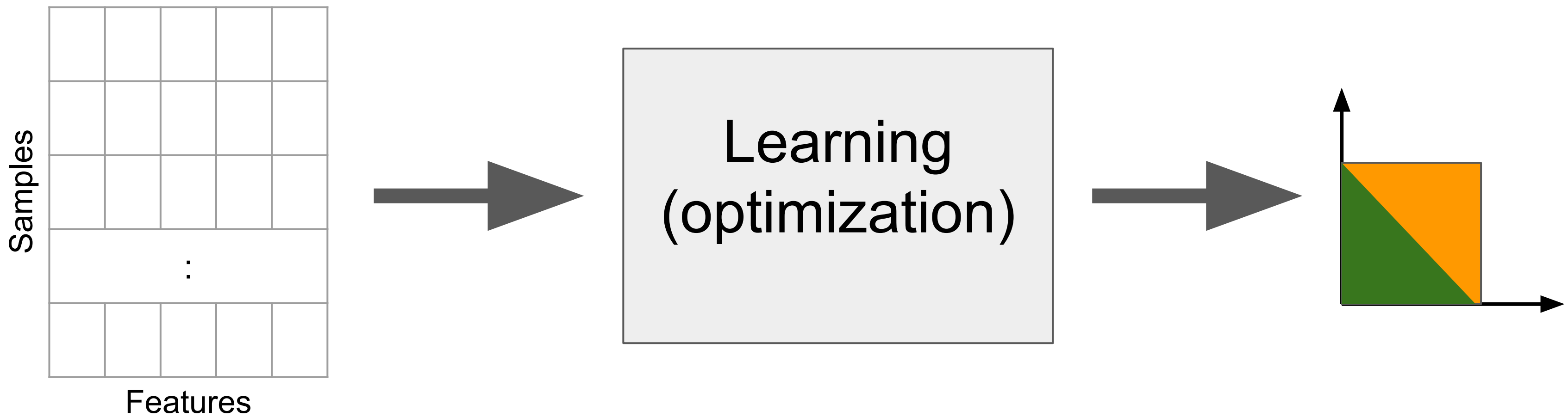
- Epoch: a series of steps, using all training data

- Learning rate = step size
- Too large and you're over the mountain
- Too small and you won't get far





Revisiting our assumptions



Training set

- Training set - Samples X Features
- Every sample has value for all the features

Missing Data

Training set contains “?”

1	5	**?**	23	**?**
12	7	8	**?**	10
5	4	12	20	12
		:		
1	**?**	10	21	8

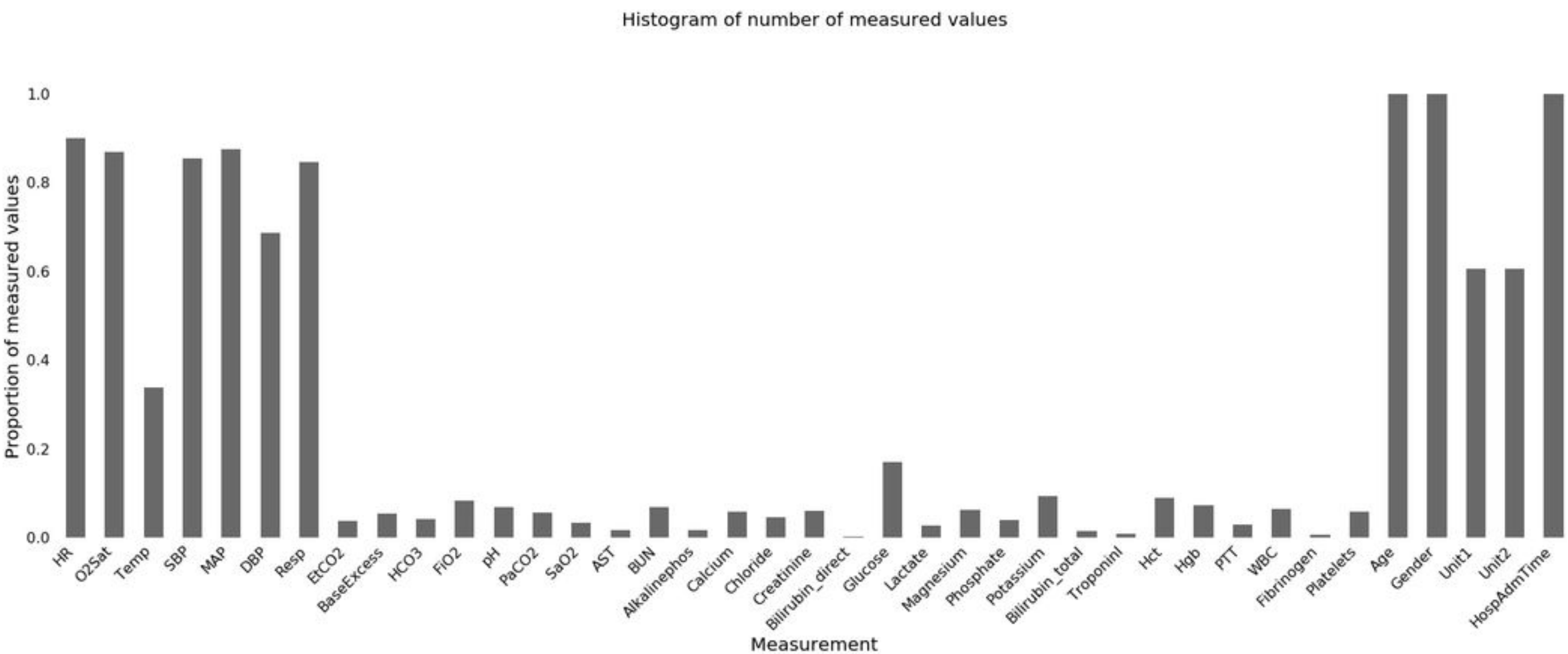
Samples

Features

Missing Data

Example:

- Basic parameters & blood works of patients in ER

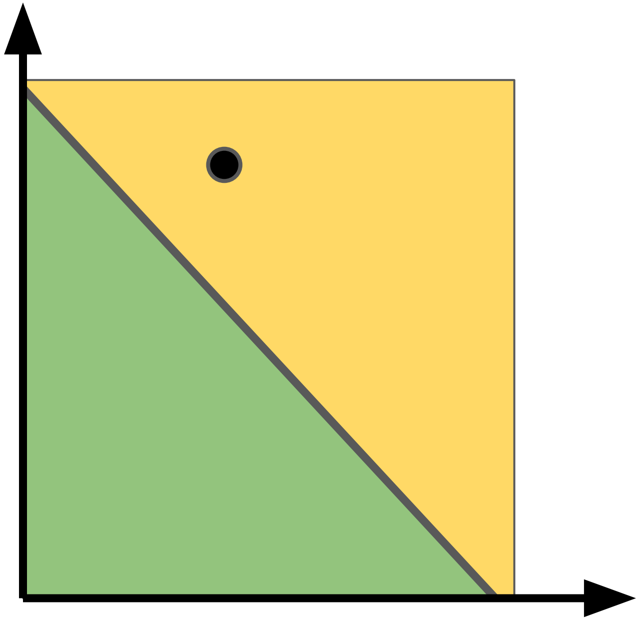


https://www.biosymetrics.com/blog/missing-values-healthcare-data

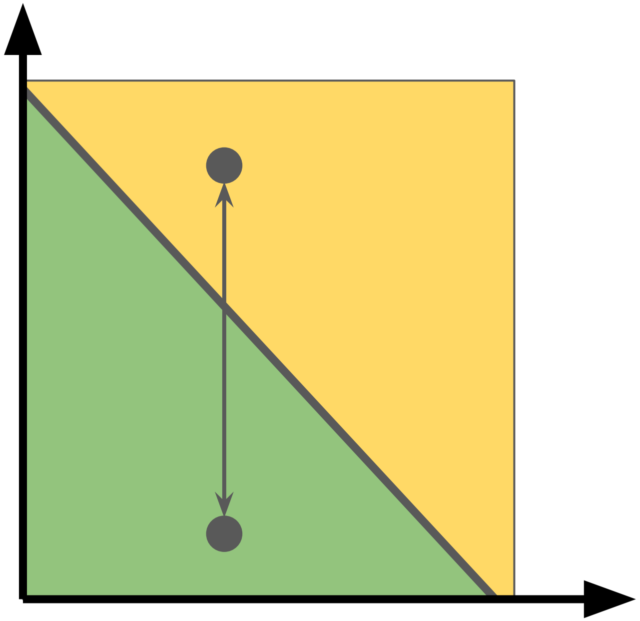
Missing Data

Why is that a problem?

Complete data - (10, 24)



Missing data - (10, ?)



Sources of Missing Data

- Missing completely at random
  Random “mechanism” removes values
  - Patients miss ~5-10% of questions on the form, each person different ones
  - Measurement device is flaky and not all results are measured

Sources of Missing Data

- Missing completely at random
- Missing specific values

  Hiding mechanism depends on actual value

  - Overweight people often do not report their weight
  - Only abnormal temperatures are recorded

Sources of Missing Data

- Missing completely at random
- Missing specific values
- Missing specific cases
  Other aspects of the sample determine whether the value is observed
  - Pathology report only when colonoscopy had positive findings

Sources of Missing Data

- Missing completely at random
- Missing specific values
- Missing specific cases
- Complex mechanism
  - Creatinine is typically measured for patients with potential kidney problems
  - Employees who fear their manager do not report their job satisfaction

Issues to consider

- Observed/missing status - is it informative?
  - Should we count it as another feature?
- Distribution of “missing values” - different than observed?
-

Approaches to Missing Data

New value that stands for “missing” or “unknown”

- Enables reasoning about the implications of not observing the values
- Can complicates the learning procedure

Approaches to Missing Data

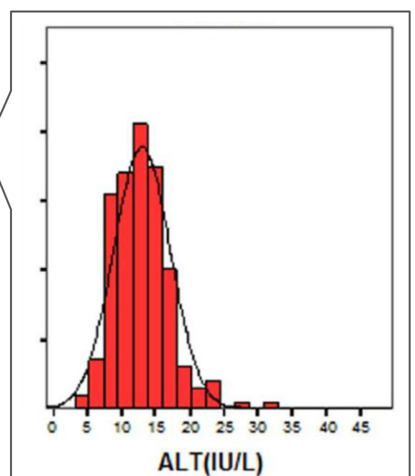
Special value: Danger of artifacts

- Missing weight value denoted as 0
- Mean / variance estimates are skewed
- Regression model treats it as another number

Samples

1	5	**?**	23	**?**
12	7	8	10	**?**
5	4	12	20	**?**
		:		
1	**?**	10	21	8

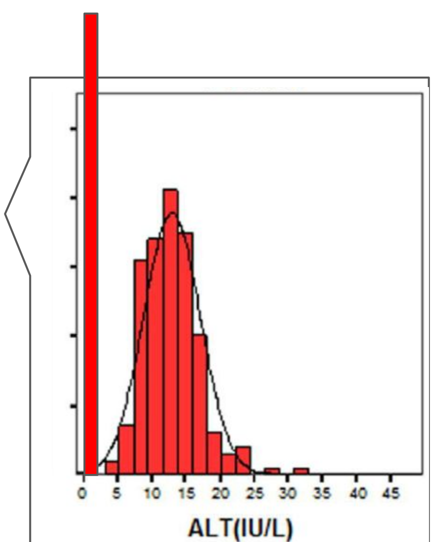
Features



Samples

1	5	**?**	23	**0**
12	7	8	10	**0**
5	4	12	20	**0**
		:		
1	**?**	10	21	8

Features



Approaches to Missing Data

Fill in missing values

- Use existing algorithms and procedures
- “Shields” the learning procedure from missing data
- Ignores information in observed/missing status

Issue - what values to fill in?

Imputation - Fill in the blanks

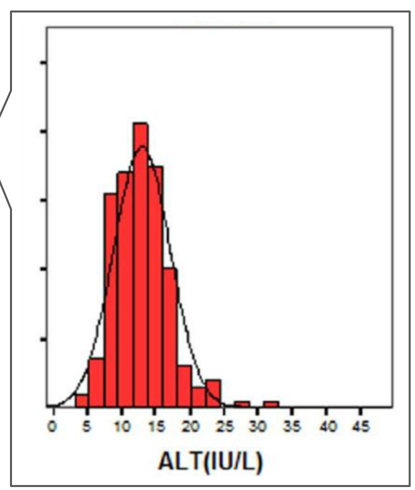
Default value

- Skews the distribution of values
- Underestimation of variance

Samples

1	5	**?**	23	**?**
12	7	8	10	**?**
5	4	12	20	**?**
		:		
1	**?**	10	21	8

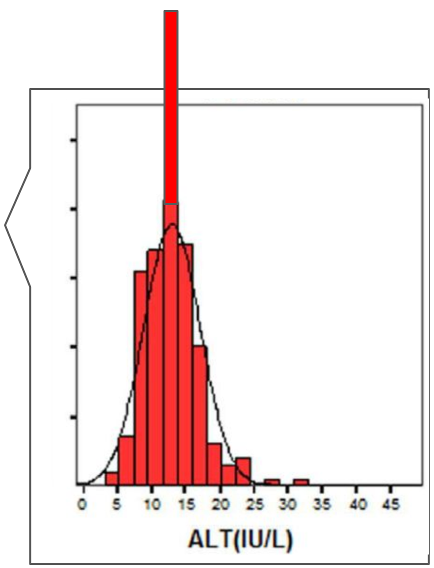
Features



Samples

1	5	**?**	23	**10**
12	7	8	10	**10**
5	4	12	20	**10**
		:		
1	**?**	10	21	8

Features



Imputation - Fill in the blanks

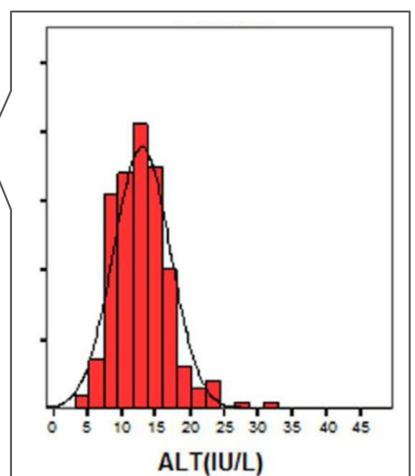
Use randomization

- Fixed distribution
- Empirical distribution

Samples

1	5	**?**	23	**?**
12	7	8	10	**?**
5	4	12	20	**?**
		:		
1	**?**	10	21	8

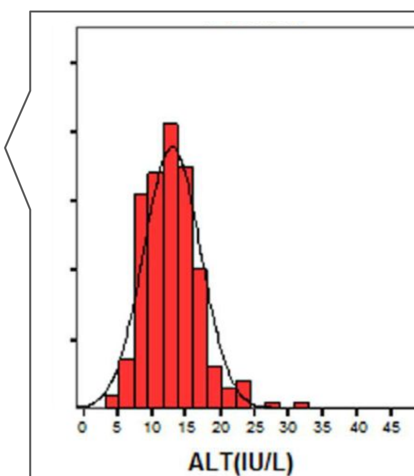
Features



Samples

1	5	**?**	23	**10**
12	7	8	10	**12**
5	4	12	20	**9**
		:		
1	**?**	10	21	8

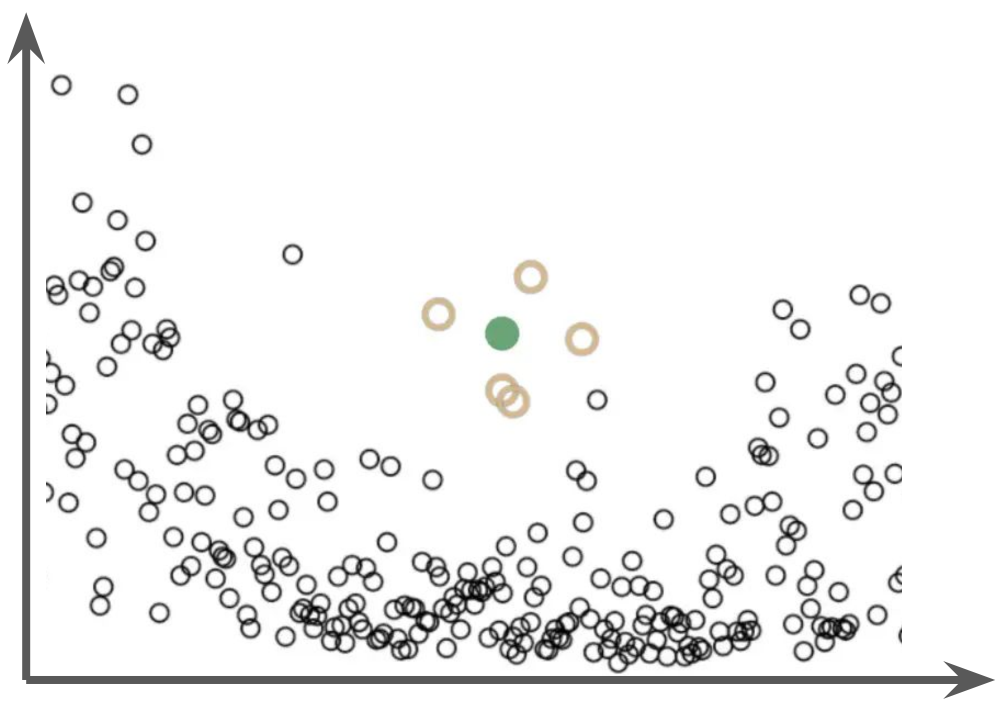
Features



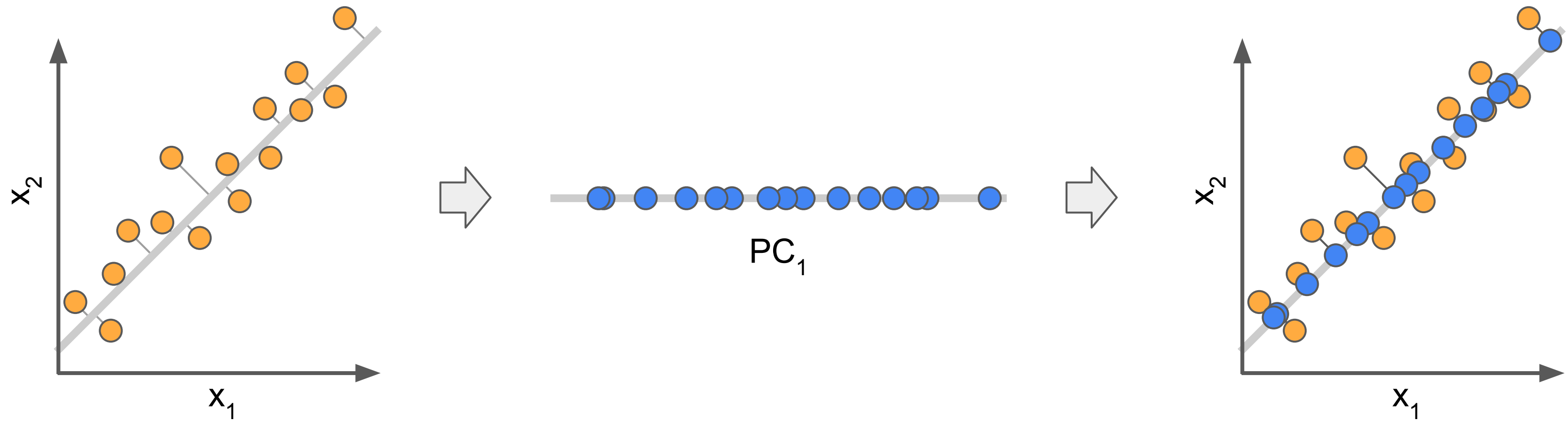
Imputation - Fill in the blanks

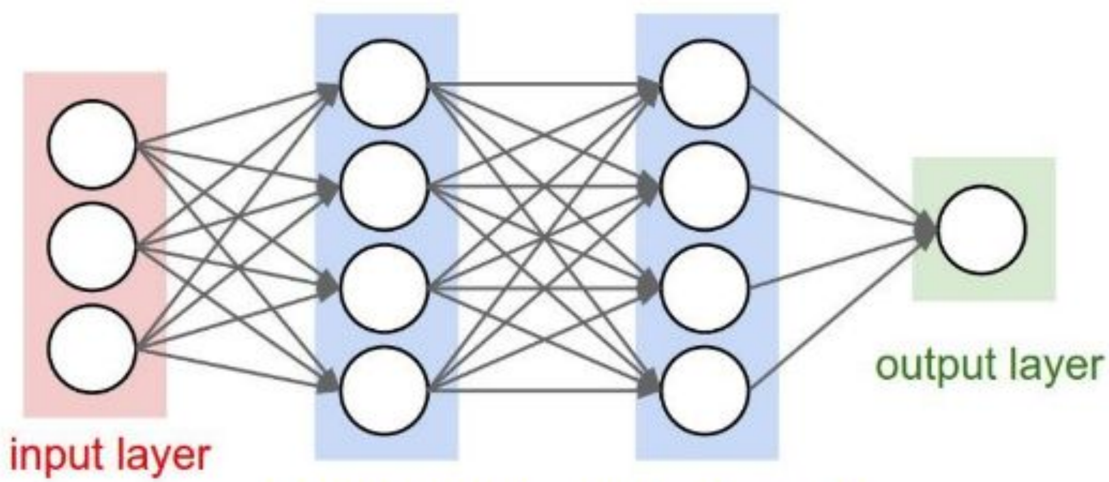
More advanced methods

- Classifier to predict based on other examples
- Use nearest neighbors to predict missing values

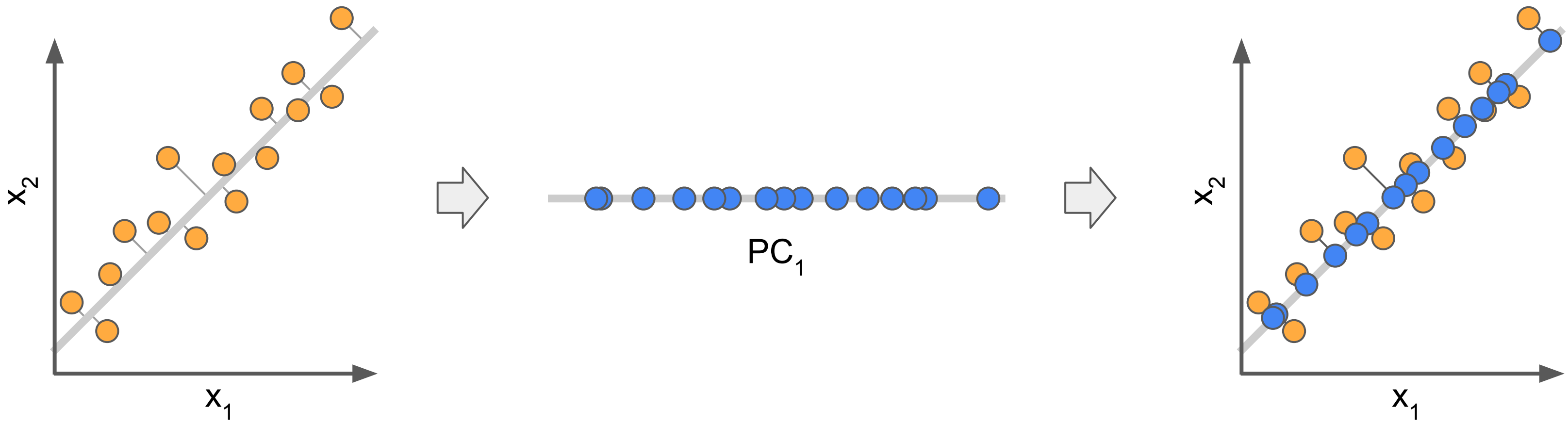


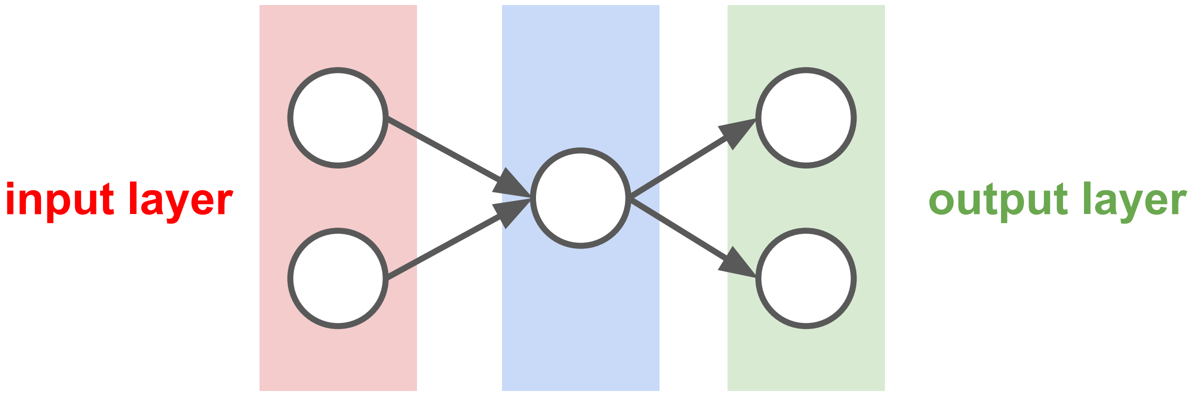
Remember PCA?





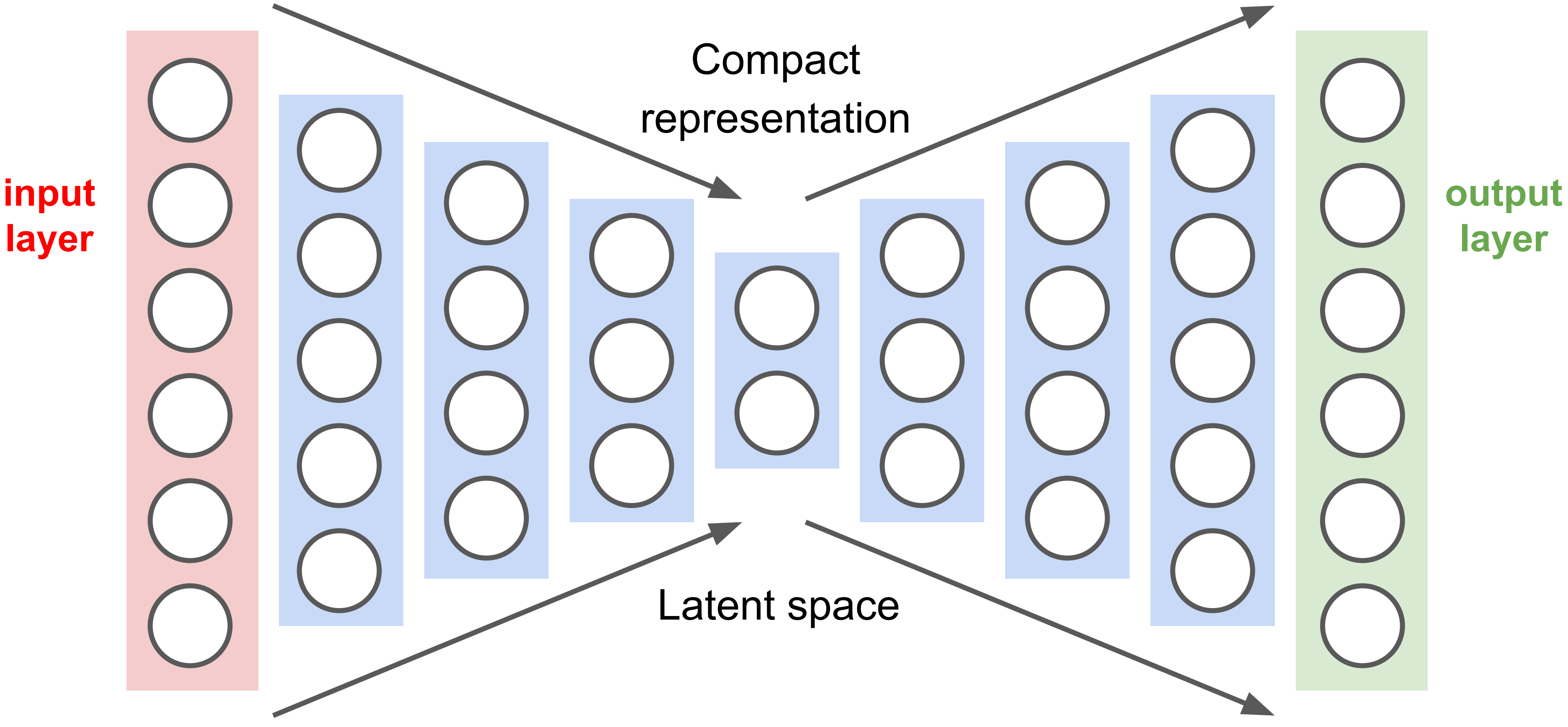
Remember PCA?





Auto-encoders

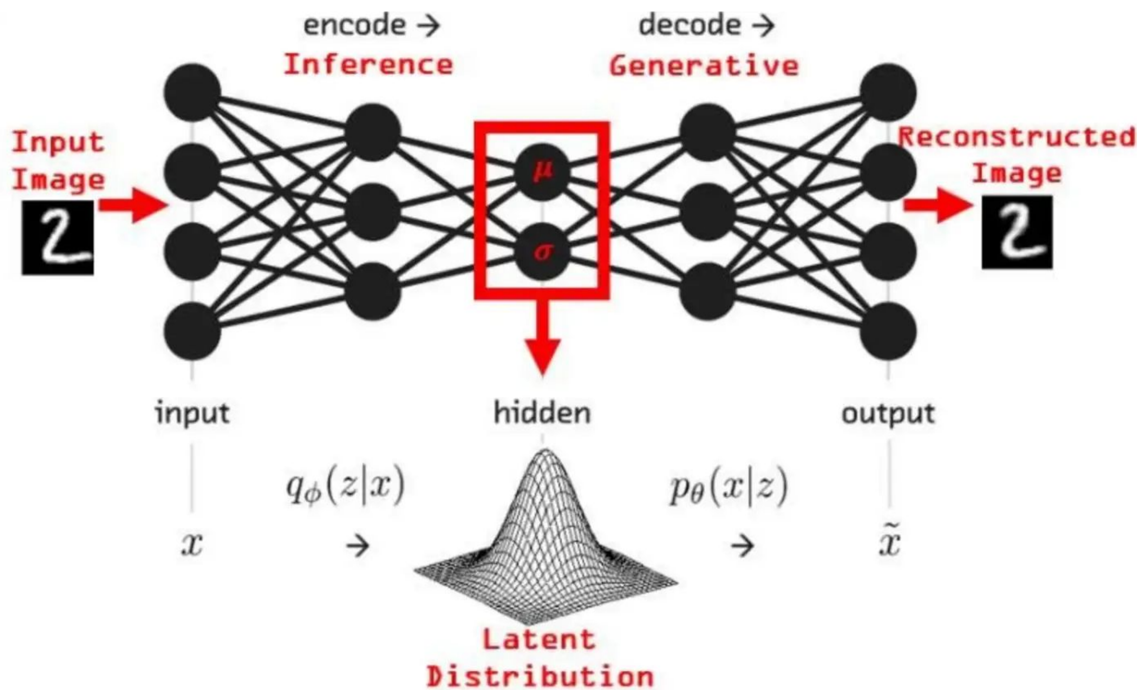
“Compress” data to lower dimension



then, “decompress” back to original dimension

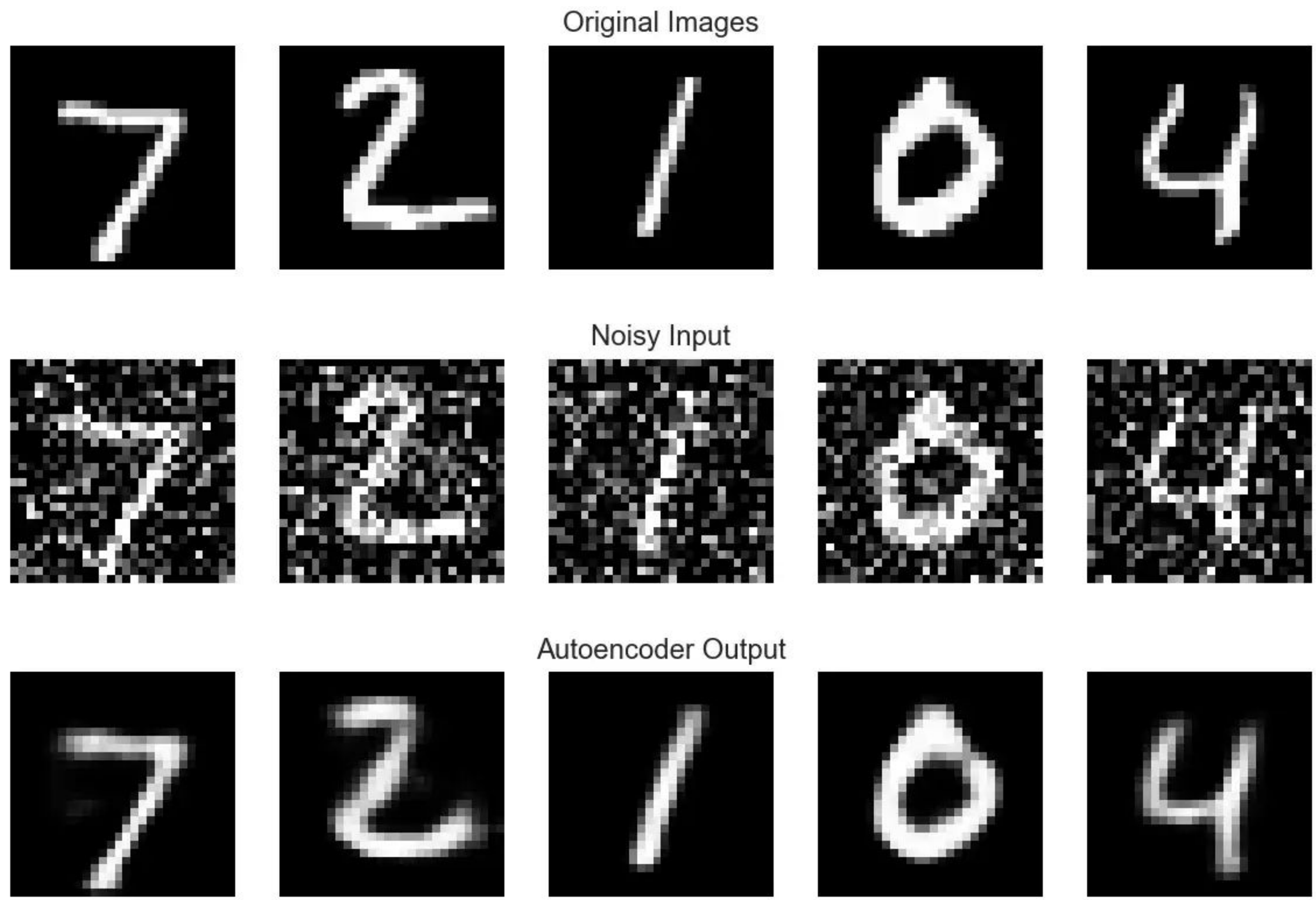
Why?

- Learn hidden dependencies or patterns in data
- Denoise



Why?

- Learn hidden dependencies or patterns in data
- Denoise



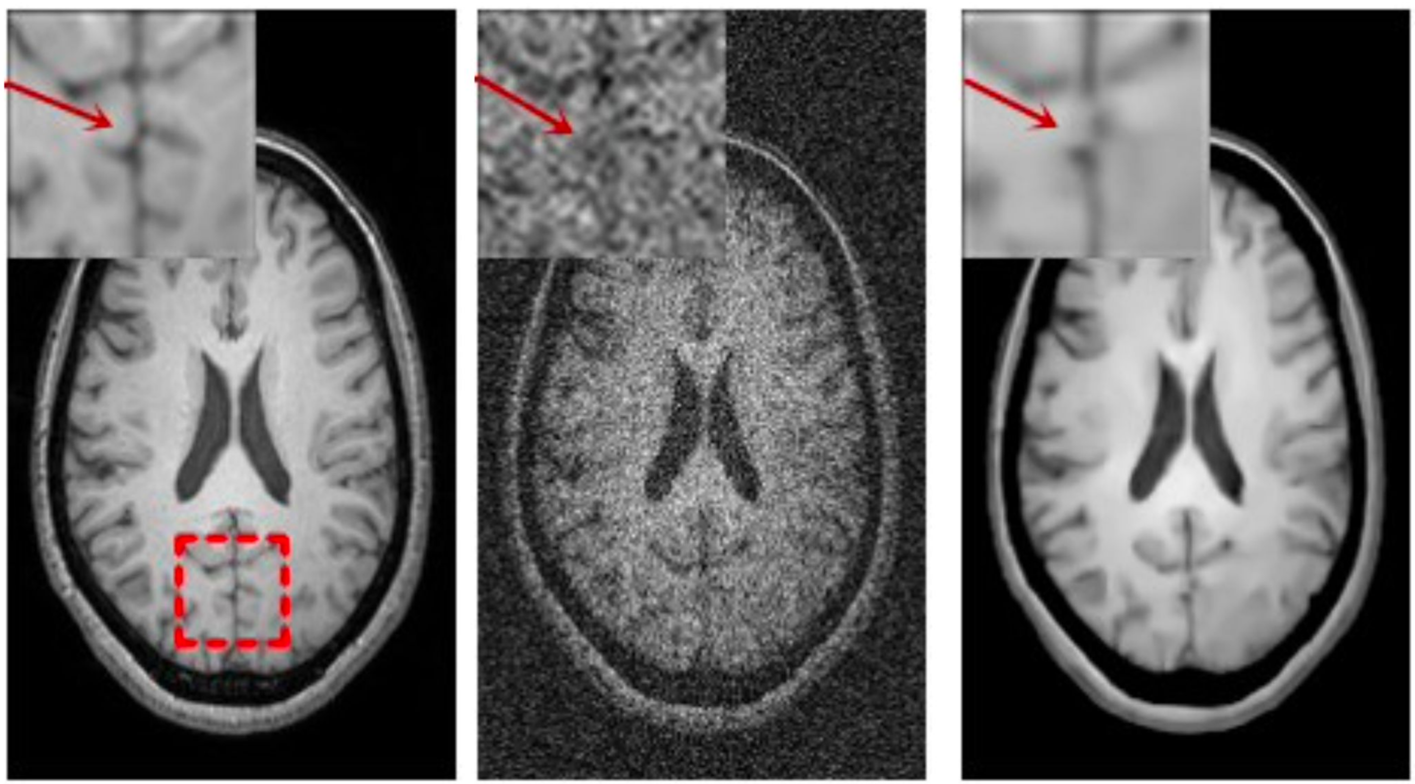
https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

Why?

- Learn hidden dependencies or patterns in data
- Denoise

Denoising of 3D Magnetic Resonance Images Using a Residual Encoder-Decoder Wasserstein Generative Adversarial Network

Maosong Ran[1], Jinrong Hu[2], Yang Chen[3,4,5], Hu Chen[1], Huaiqiang Sun[6], Jiliu Zhou[1], Yi Zhang[1,7,*]



(a) (b) (c)

Why?

- Learn hidden dependencies or patterns in data
- Denoise
- Impute

Extracting and Composing Robust Features with Denoising Autoencoders

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, Pierre-Antoine Manzagol

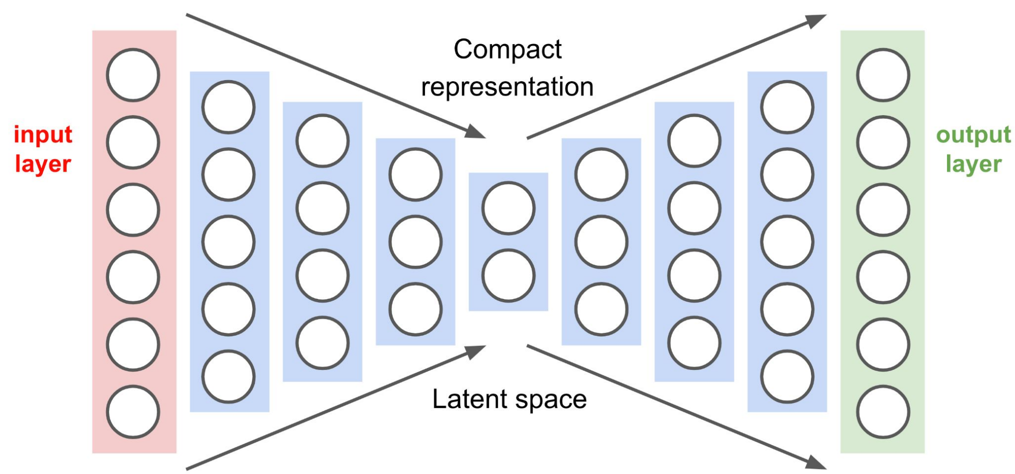
MISSING DATA IMPUTATION IN THE ELECTRONIC HEALTH RECORD USING DEEPLY LEARNED AUTOENCODERS*

BRETT K. BEAULIEU-JONES

Genomics and Computational Biology Graduate Group, Computational Genetics Lab, Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia PA, 19104
Email: brettbe@med.upenn.edu

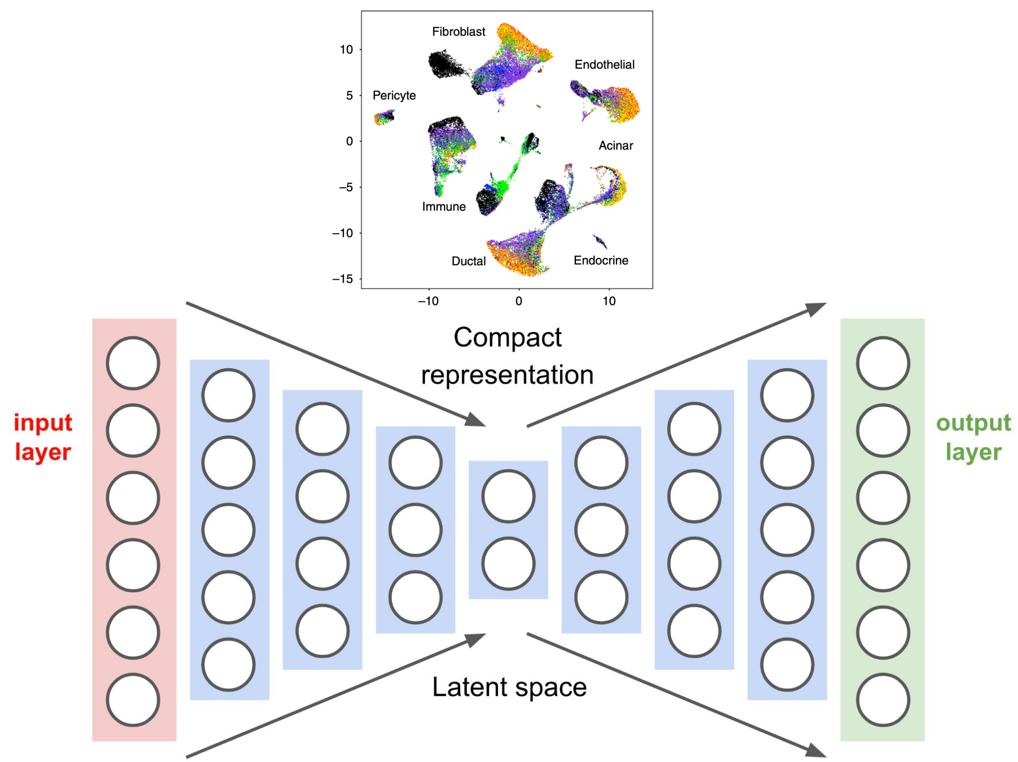
JASON H. MOORE

Computational Genetics Lab, Institute for Biomedical Informatics, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia PA, 19104
Email: jhmoore@exchange.upenn.edu



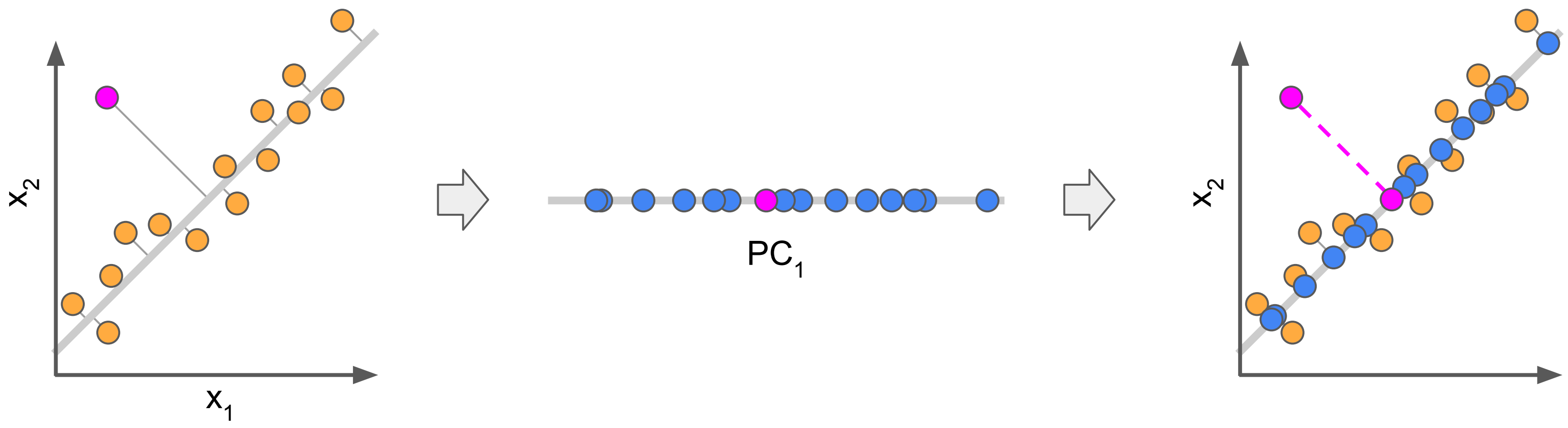
Why?

- Learn hidden dependencies or patterns in data
- Denoise
- Impute
- Visualize / cluster data in latent space



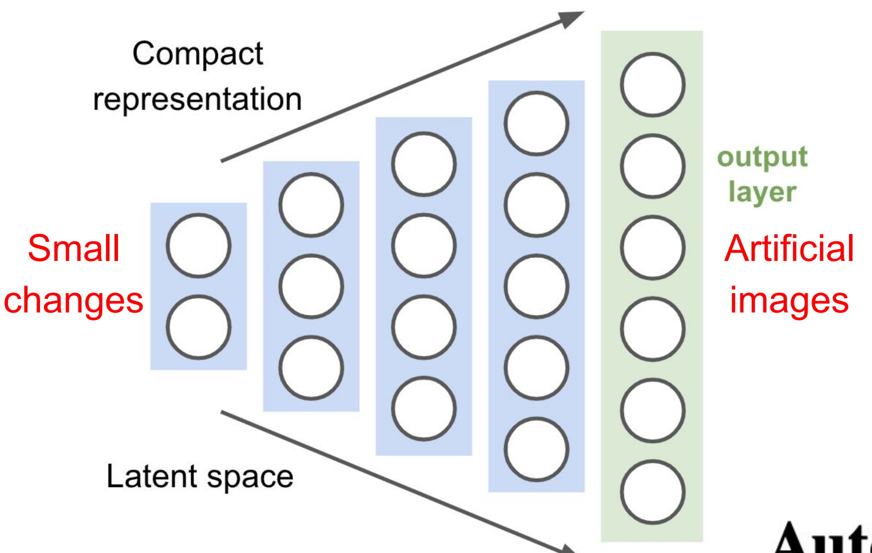
Why?

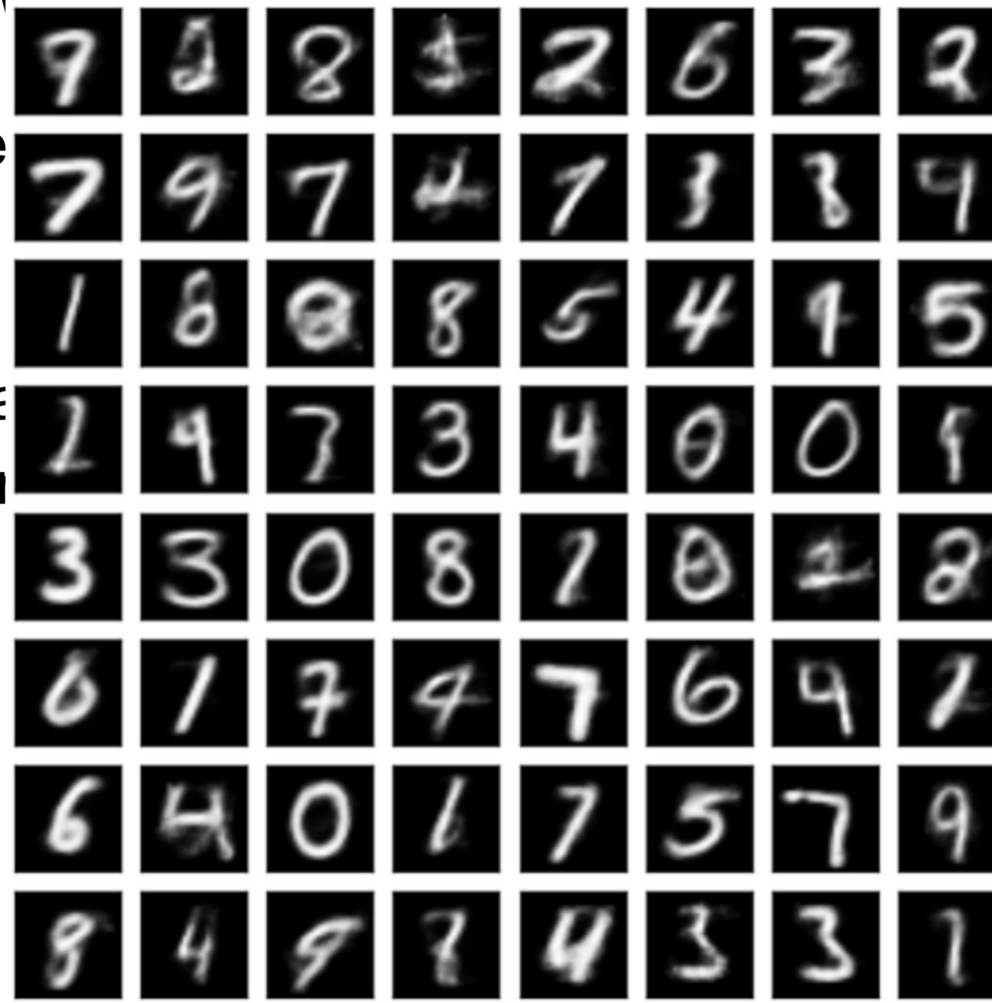
- Learn hidden dependencies or patterns in data
- Denoise
- Impute
- Visualize / cluster data in latent space
- Anomaly detection (= reconstruction failures)



Why?

- Learn hidden dependencie
- Denoise
- Impute
- Visualize / cluster data in la
- Anomaly detection (= recon
- Data generation



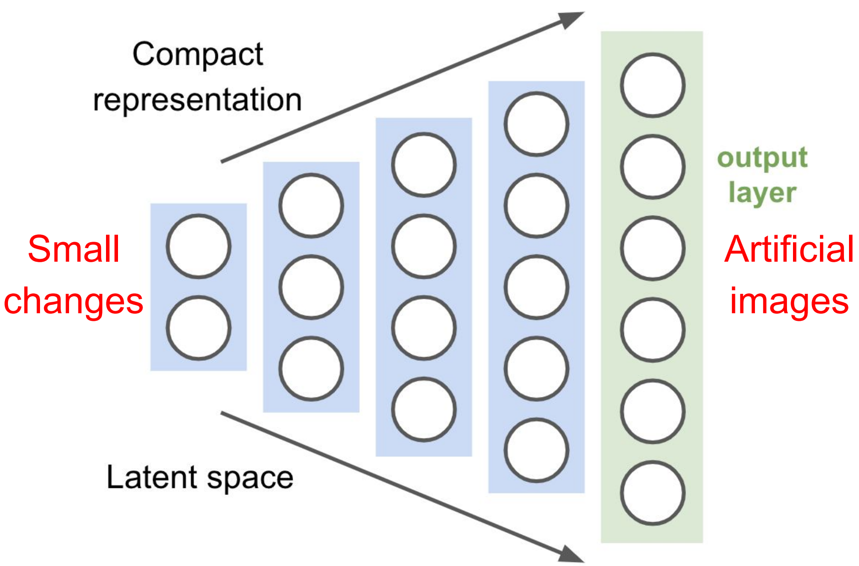


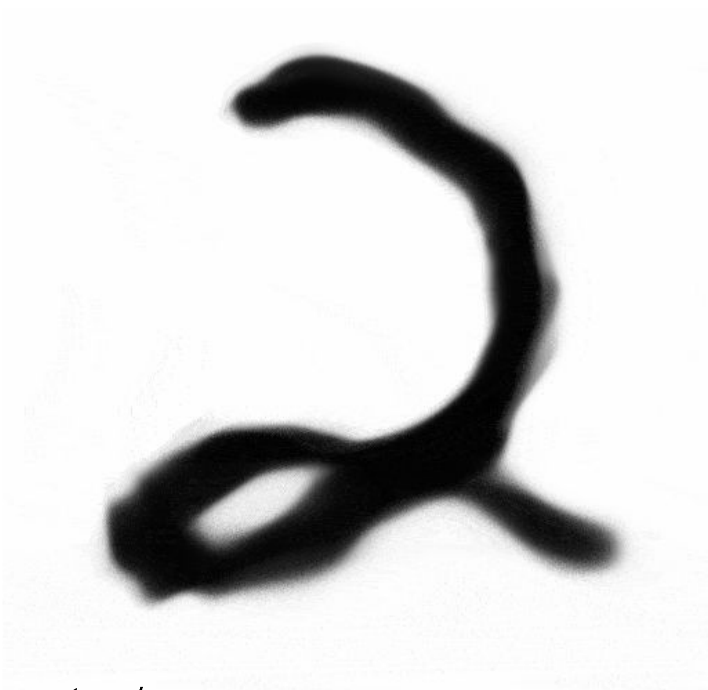
(b) VAE generated MNIST images.

Autoencoders Dor Bank, Noam Koenigstein, Raja Giryes

Why?

- Learn hidden dependencies or patterns in data
- Denoise
- Impute
- Visualize / cluster data in latent space
- Anomaly detection (= reconstruction failures)
- Data generation

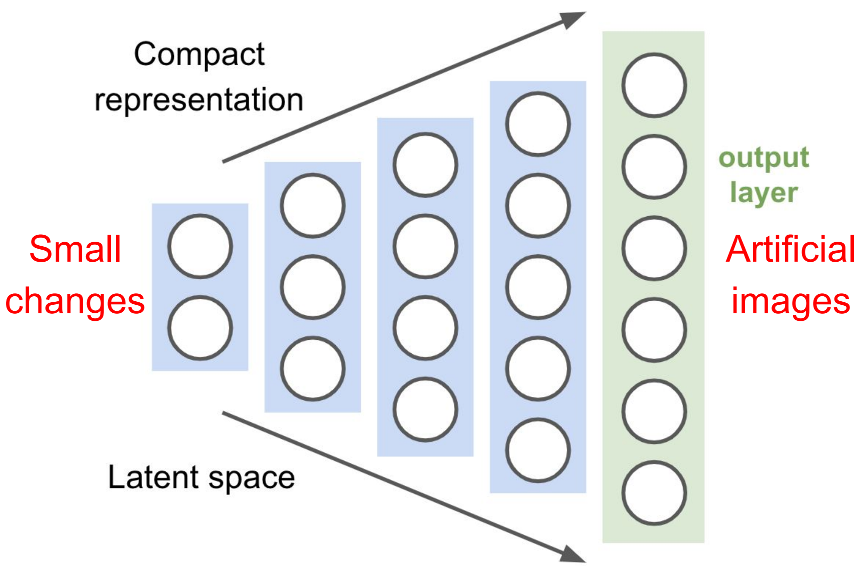


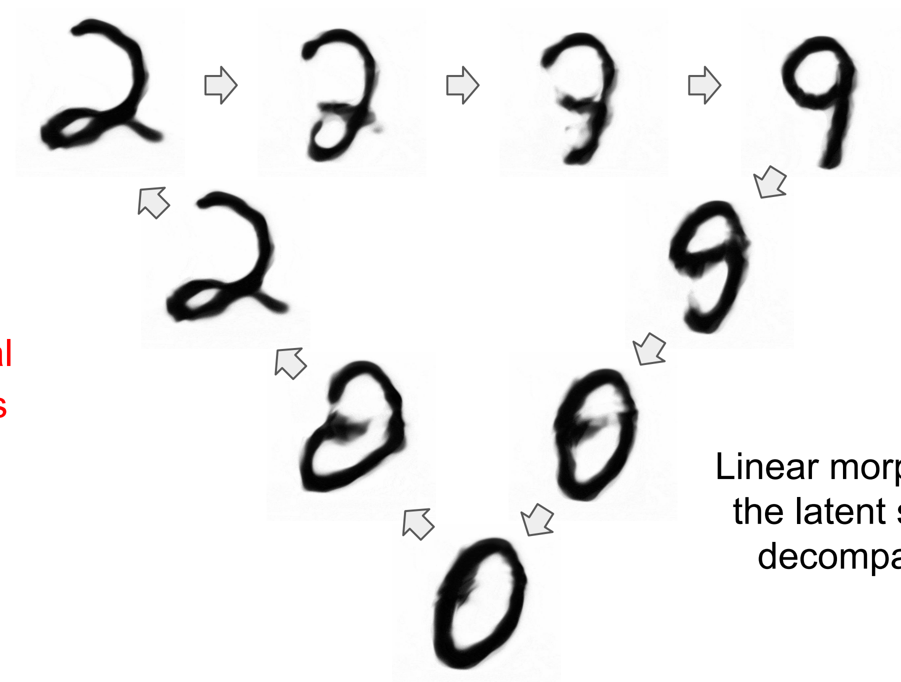


https://blog.otoro.net/2016/04/01/generating-large-images-from-latent-vectors/

Why?

- Learn hidden dependencies or patterns in data
- Denoise
- Impute
- Visualize / cluster data in latent space
- Anomaly detection (= reconstruction failures)
- Data generation



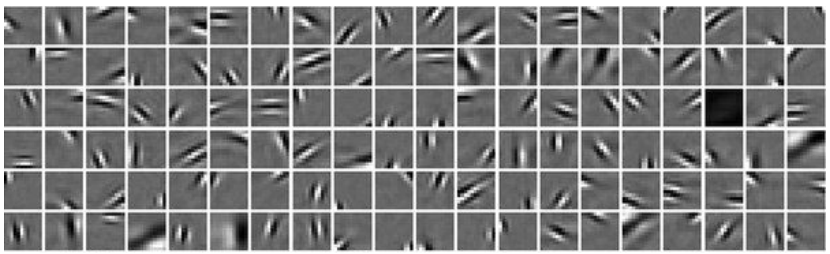


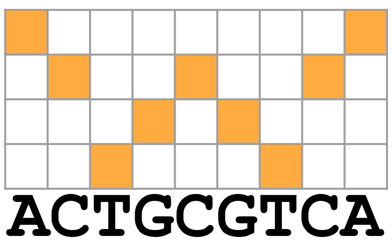
Linear morphing in the latent space, decompacted

https://blog.otoro.net/2016/04/01/generating-large-images-from-latent-vectors/

Transfer learning

- Re-use parts of a trained network (e.g., early filters/features)





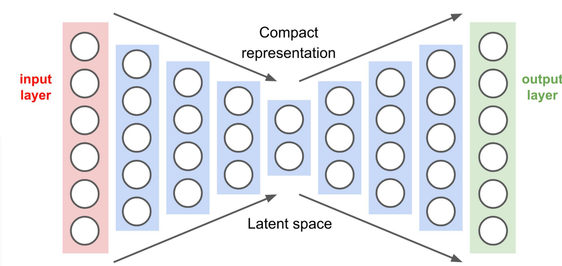
Transfusion: Understanding Transfer Learning for Medical Imaging

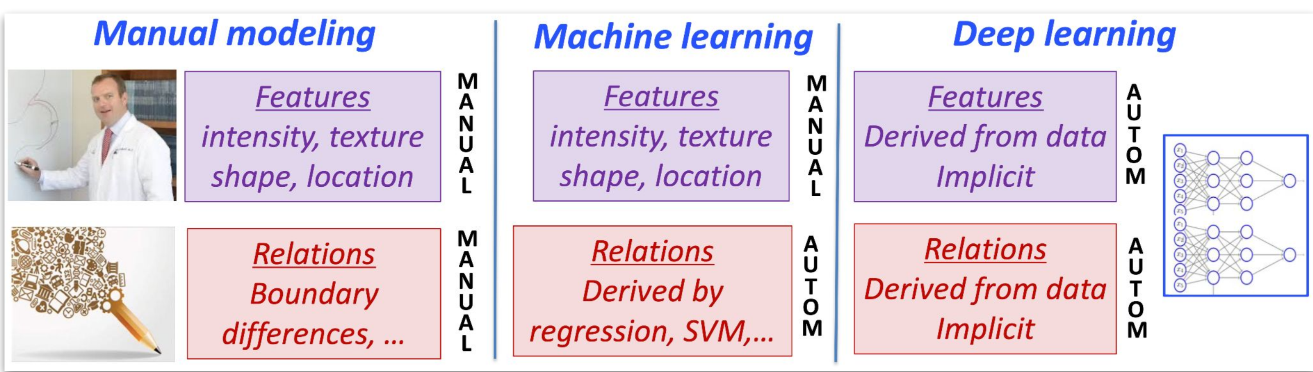
Maithra Raghu*
Cornell University and Google Brain
maithrar@gmail.com

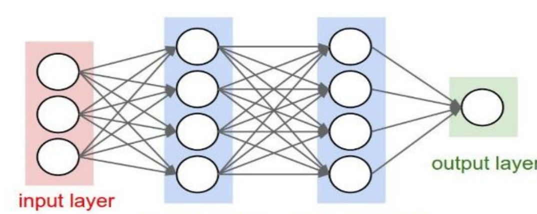
Chiyuan Zhang*
Google Brain
chiyuan@google.com

Jon Kleinberg†
Cornell University
kleinber@cs.cornell.edu

Samy Bengio†
Google Brain
bengio@google.com

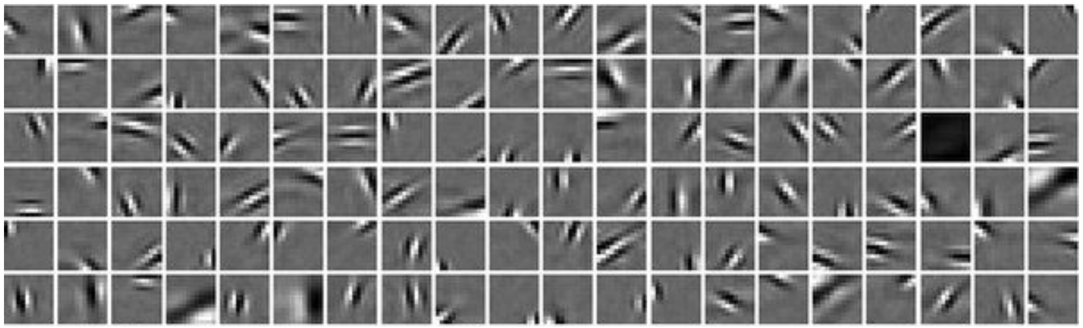






Transfer learning

- Re-use parts of a pre-trained network (early filters = basic visual features)



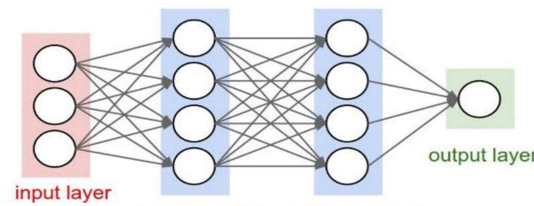
Transfusion: Understanding Transfer Learning for Medical Imaging

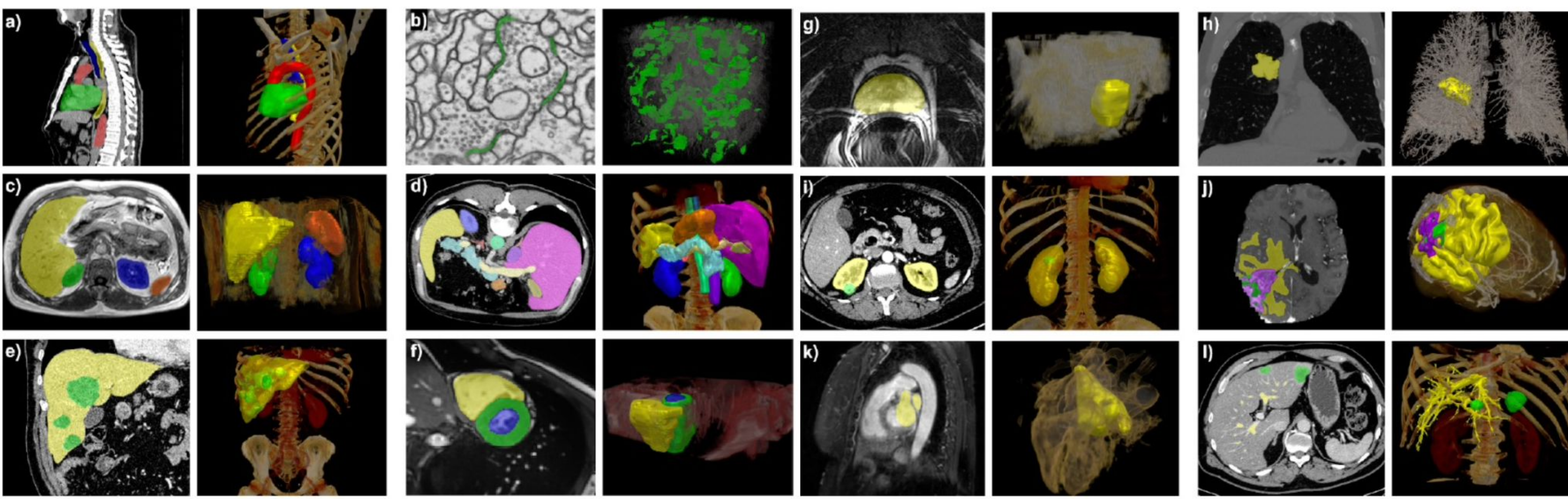
Maithra Raghu*
Cornell University and Google Brain
maithrar@gmail.com

Chiyuan Zhang*
Google Brain
chiyuan@google.com

Jon Kleinberg†
Cornell University
kleinber@cs.cornell.edu

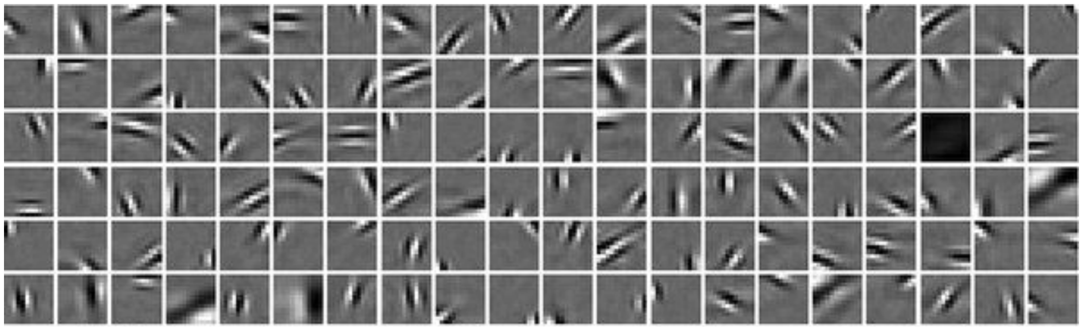
Samy Bengio†
Google Brain
bengio@google.com





Transfer learning

- Re-use parts of a pre-trained network (early filters = basic visual features)



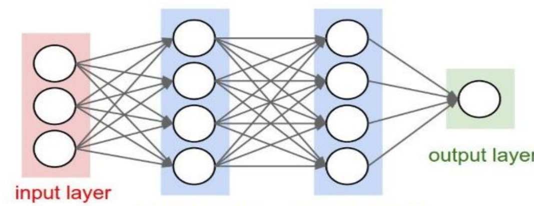
Transfusion: Understanding Transfer Learning for Medical Imaging

Maithra Raghu[*]
Cornell University and Google Brain
maithrar@gmail.com

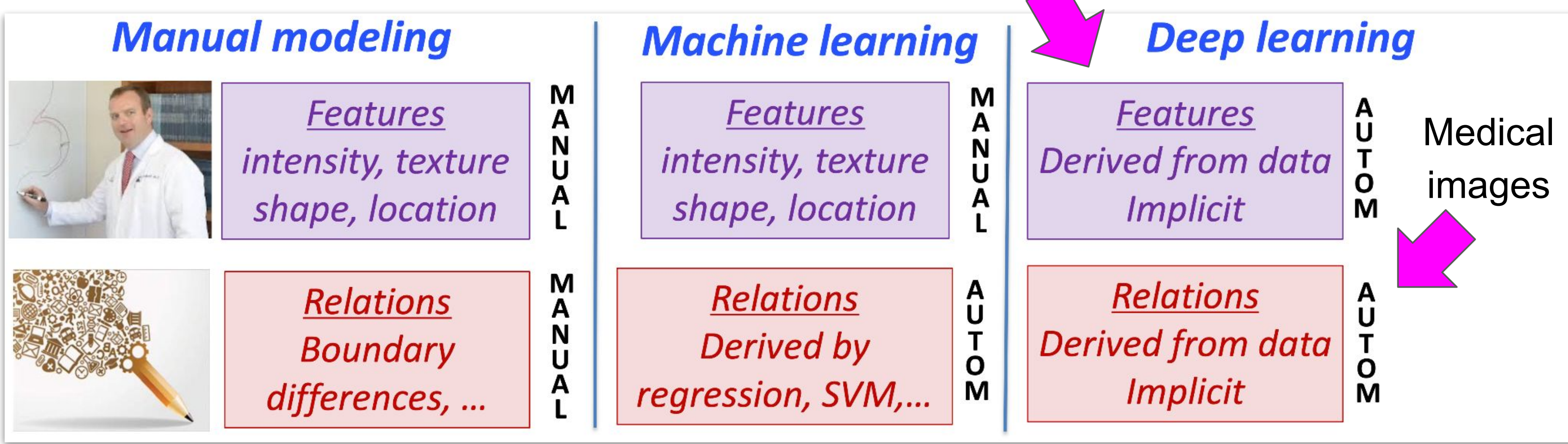
Chiyuan Zhang[*]
Google Brain
chiyuan@google.com

Jon Kleinberg[†]
Cornell University
kleinber@cs.cornell.edu

Samy Bengio[†]
Google Brain
bengio@google.com



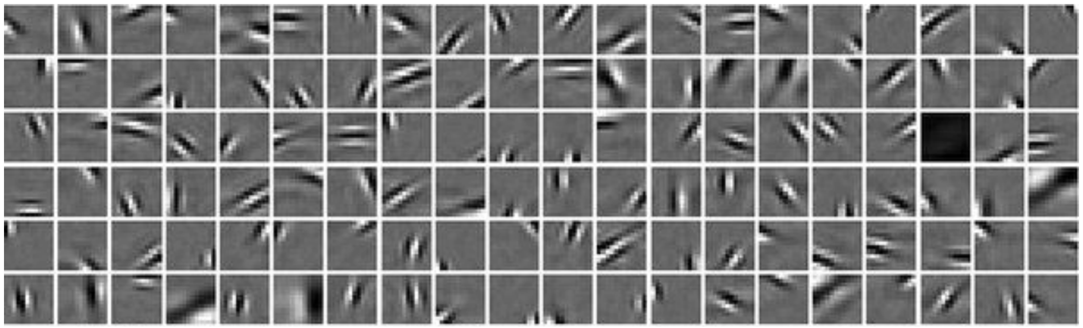
"Vanilla" images



Medical images

Transfer learning

- Re-use parts of a pre-trained network (early filters = basic visual features)



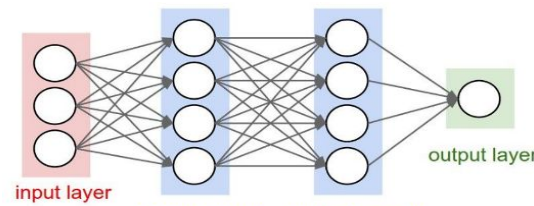
Transfusion: Understanding Transfer Learning for Medical Imaging

Maithra Raghu* Cornell University and Google Brain maithrar@gmail.com

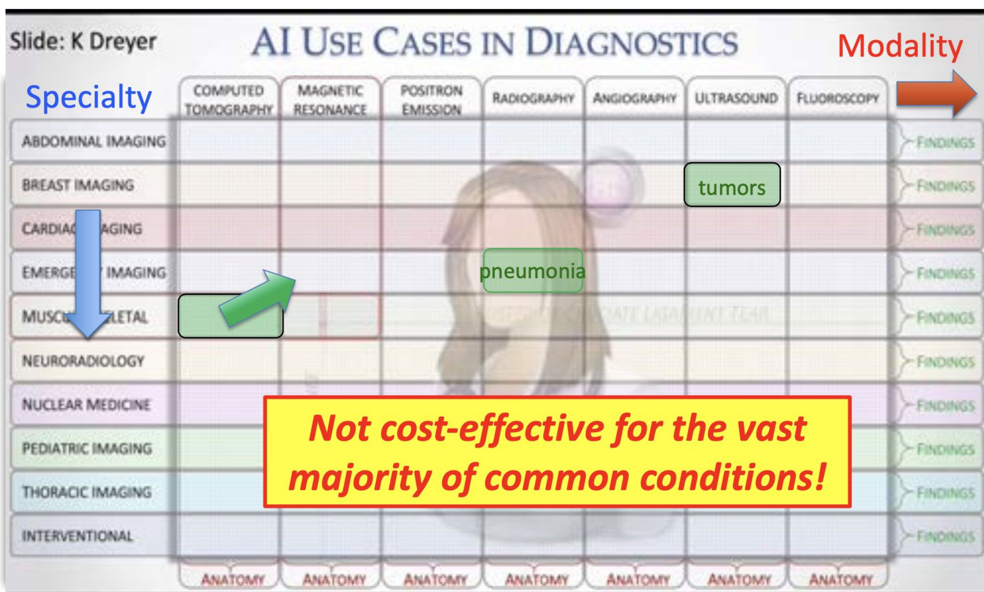
Chiyuan Zhang* Google Brain chiyuan@google.com

Jon Kleinberg† Cornell University kleinber@cs.cornell.edu

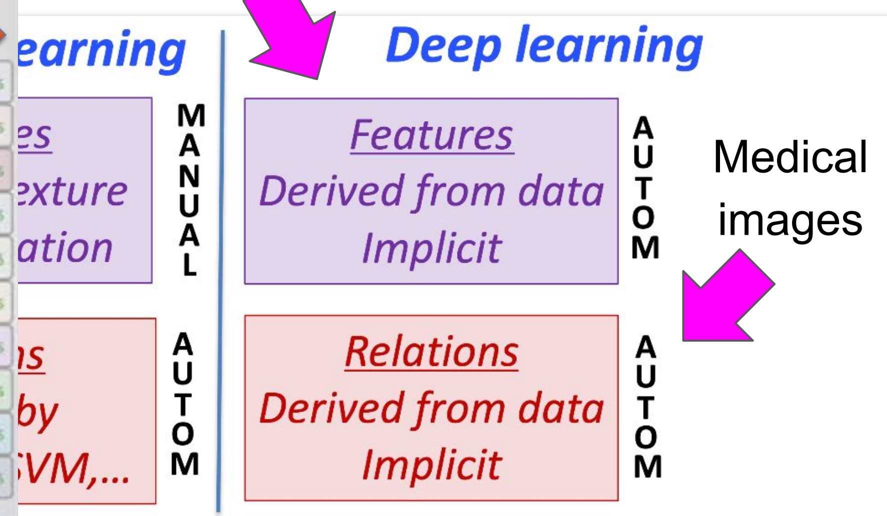
Samy Bengio† Google Brain bengio@google.com



Narrow AI – long time and high costs!



"Vanilla" images



Medical images

What have we learned?

- Convolutional neural networks
- Missing data
- Imputation
- Auto-encoders
- Learning latent representations
- Multiple uses of latent representations
- Transfer learning