MANCHESTER 1824

The University of Manchester

Data Representation by Deep Learning

SOCO/CISIS/ICEUTE 2018

Hujun Yin
The University of Manchester

MANCHESTER 1824

The University of Manchester

Outline

- Linear Representation (PCA)
- Nonlinear Representation (NLPCA, MDS, PC/S, etc.)
- Neural Networks
- Deep Neural Networks
- Deep Representation (ConvNet Features)
- Manifold & Meaning
- Autoassociative NNs & Deep Autoencoder
- Unsupervised ConvNets Features
- Summary

MANCHESTER 1824

The University of Manchester

Linear Data Representation

- **Data matrix:** $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, $\mathbf{x}_i \in \mathrm{R}^n$ *column vector,*

  N: number of samples

$\mathbf{x}_i = [x^1, x^2, \ldots, x^n]^T$

Gene expressions,

Images,

Patient records,

Surveys,

Documents,

……

Microarray Data

Ranking	*University*	*F1*	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*Total*
1	Cambridge	241	182	247	97	88	100	50	**1005**
2	Oxford	214	175	244	81	99	100	30	**941**
3	LSE	200	175	233	75	51	86	50	**923**
4	Imperial	203	154	232	89	44	100	40	**864**
5	York	206	143	208	94	56	76	60	**850**
6	UCL	172	152	210	95	72	100	30	**830**
7	St Andrews	139	131	194	96	73	91	100	**824**
8	Warwick	153	155	215	97	68	86	20	**795**
9	Bath	132	142	211	79	62	93	60	**791**
9	Nottingham	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**791**
11	Bristol	145	131	218	90	75	94	20	**779**
11	Durham	163	132	207	91	64	72	50	**779**
11	Edinburg	106	145	218	96	74	100	40	**779**
14	Lancaster	156	144	186	95	62	63	50	**756**
15	UMIST	135	141	188	91	61	64	50	**752**
16	Birmingham	146	124	193	88	62	82	20	**747**
17	Loughborough	162	115	177	93	57	66	60	**732**
18	Southampton	143	124	180	93	55	71	50	**716**
19	King's College	135	126	204	96	63	100	-10	**714**
20	Newcastle	134	117	193	97	60	87	20	**708**
21	Manchester	125	134	198	96	66	98	-10	**707**
22	Leeds	122	127	199	97	61	74	20	**700**
23	Sheffield	143	125	213	94	51	72	-20	**691**
24	East Anglia	125	127	176	96	63	60	40	**687**
24	Leicester	125	120	183	94	52	93	20	**687**

h.yin@manchester.ac.uk

MANCHESTER 1824

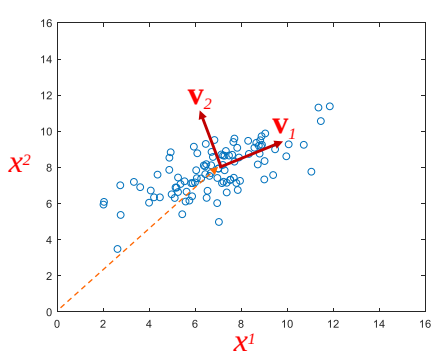
The University of Manchester

Linear Data Representation

- **Data matrix:** $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, $\mathbf{x}_i \in \mathrm{R}^n$ *column vector,*

  N: number of samples

$\mathbf{x}_i = [x^1, x^2, \ldots, x^n]^T$



Assume: n is too large, and $x^1, x^2, \ldots, x^n$ are correlated.

Can we de-correlate and reduce these variables?

e.g. $n \rightarrow 1$?

$\rightarrow$ *find* $\mathbf{v}_1$ *so that* $\mathbf{v}_1^T\mathbf{X}$ *or* $\mathbf{v}_1^T\mathbf{X}\mathbf{X}^T\mathbf{v}_1$ *largest.*

If $n \rightarrow 2$*: in addition to* $\mathbf{v}_1$*, find* $\mathbf{v}_2$ *so that* $\mathbf{v}_1 \perp \mathbf{v}_2$*, and* $\mathbf{v}_2^T\mathbf{X}\mathbf{X}^T\mathbf{v}_2$ *largest.*

h.yin@manchester.ac.uk

MANCHESTER 1824

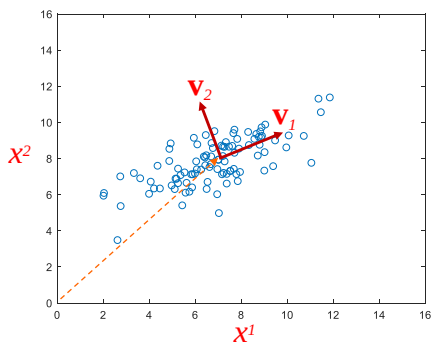
The University of Manchester

Linear Data Representation: PCA

- **PCA:** *A linear coordinate transformation*

$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$, $\mathbf{x}_i$ *zero mean*, *Covariance:* $\mathbf{X}\mathbf{X}^T$

$\mathbf{x}_i = [x^1, x^2, \ldots, x^n]^T$



$$max\{\mathbf{v}_i^T\mathbf{X}\mathbf{X}^T\mathbf{v}_i\} = \sigma_i^2, \mathbf{v}_i^T\mathbf{v}_j = 0, i \neq j$$

$$min \sum_{\mathbf{x}} \left\| \mathbf{x} - \sum_{i=1}^{k} \mathbf{v}_i^T\mathbf{x}\mathbf{v}_i \right\|^2$$

Eigenvalue problem:

$$(\mathbf{X}\mathbf{X}^T - \lambda_i\mathbf{I})\mathbf{v}_i = 0$$

$$\mathbf{V}^T\mathbf{X}\mathbf{X}^T\mathbf{V} = \Lambda$$

- $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$
- $\mathbf{\Lambda} = \text{diag}\,[\lambda_1, \lambda_2, \ldots \quad \lambda_n]$
- $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ eigenvalues or variances

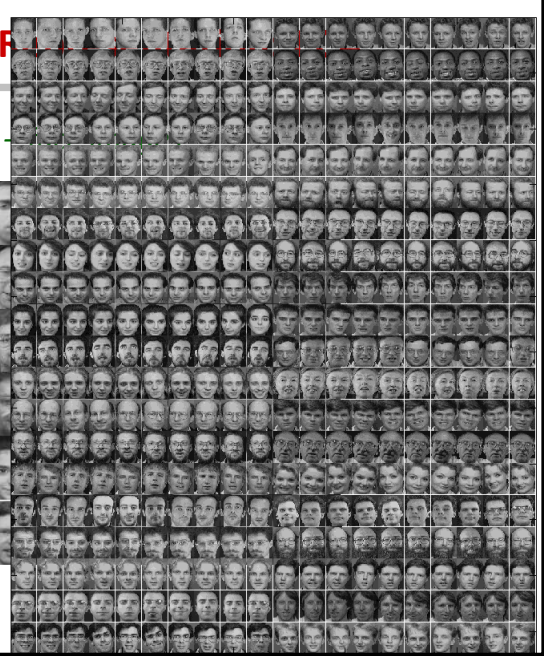
h.yin@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

Linear Data R

- **PCA:** *eigenface example*



MANCHESTER 1824

The University of Manchester

Linear Data Representation: PCA

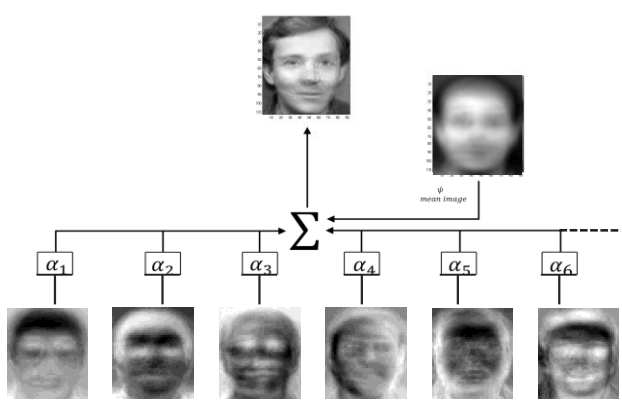
- **PCA:** *eigenface example – first 50 eigenvectors (eigenfaces)*

MANCHESTER 1824

The University of Manchester

Linear Data Representation: PCA

- **PCA:** *eigenface example – representation or reconstruction*



Reconstruction of an image from the mean image and a number of weighted eigenfaces, calculated from the ORL database.

MANCHESTER 1824

The University of Manchester

Why Study Images or Vision?

All our knowledge has its origins in our perceptions

– Leonardo da Vinci

MANCHESTER 1824

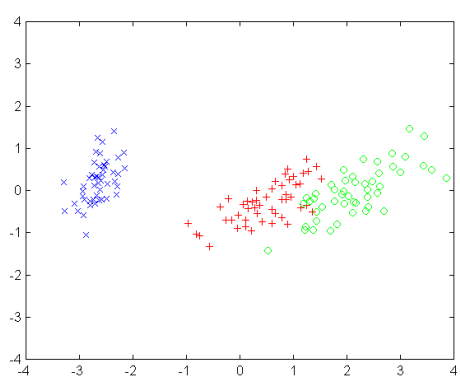
The University of Manchester

Nonlinear Representation

MDS (Multidimensional Scaling)

$$S = \frac{1}{\sum d_{ii}^2} \sum_{i>j} [d_{ij} - D_{ij}]^2$$

4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
……			
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
5.7	2.8	4.5	1.3
……			
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
7.6	3.0	6.6	2.1
……			
……			



MANCHESTER 1824

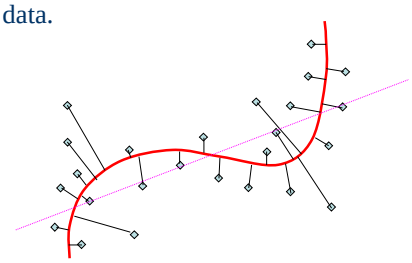
The University of Manchester

Nonlinear Representation

Principal Curve/Surface

-Hastie and Stuetzle (1989)

A smooth and self-consistent curve passing through the "middle" of the data.



Projection:

$$\rho_f(\mathbf{x}) = \sup_{\rho\in\Lambda}\{\rho : \|\mathbf{x} - f(\rho)\| = \inf_{\vartheta}\|\mathbf{x} - f(\vartheta)\|\}$$

Expectation:

$$f(\rho) = E[\mathbf{X} \mid \rho_f(\mathbf{X}) = \rho]$$

Kernel smoothing:

$$F(\rho) = \frac{\sum_i^S \mathbf{x}_i \kappa(\rho, \rho_i)}{\sum_i^S \kappa(\rho, \rho_i)}$$

MANCHESTER 1824

The University of Manchester

Nonlinear Representation

- **Kernel PCA:** *Shölkopf, Smola & Müller (1998)* for nonlinear PCA.
  - **Kernel method has become popular.**

$$\Phi : \mathrm{X} \to \mathrm{F},$$
$$\kappa : \mathrm{X}\times\mathrm{X} \in \Re,$$

  - **PCA**

$$\mathbf{Cq} = \lambda\mathbf{q}, \qquad \mathbf{C} = \frac{1}{N}\sum_i \mathbf{x}_i\mathbf{x}_i^T, \ \mathbf{q} = \sum_i \alpha_i \mathbf{x}_i,$$

$$\min \sum_{\mathbf{x}} \left(\mathbf{x} - \sum_{j=1}^m (\mathbf{q}_j^T\mathbf{x})\mathbf{q}_j\right)^2$$

MANCHESTER 1824

The University of Manchester

Nonlinear Representation

- **Kernel PCA:** *Shölkopf, Smola & Müller (1998)* for nonlinear PCA.

$$\mathbf{K}\boldsymbol{\alpha} = \lambda\boldsymbol{\alpha},$$

$$K_{ij} := \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle, \qquad \boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots\ldots \alpha_N]^T,$$

$$\mathbf{q} = \sum_i \alpha_i \Phi(\mathbf{x}_i), \qquad \langle \Phi(\mathbf{x}_k), \mathbf{q} \rangle = \sum_i \alpha_i \kappa(\mathbf{x}_k, \mathbf{x}_i),$$

$$Cov = \frac{1}{N} \sum_{i=1}^{N} \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_i)^T$$

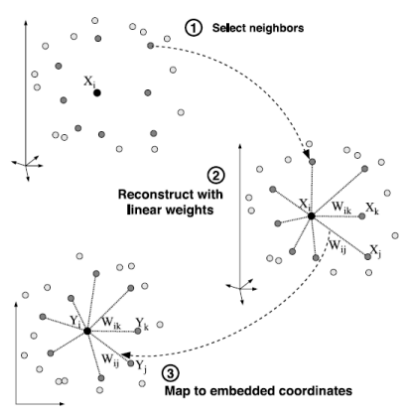
MANCHESTER 1824

The University of Manchester

Nonlinear Representation

- **LLE (Local Linear Embedding):**
  - Roweis & Saul (2000)
  for nonlinear dimensionality reduction

  - **Select neighbourhood graph:**
    k nearest neighbours or ε ball.
  - **Reconstruct linear weights:**
    $$\varepsilon(W) = \min \sum_i \| X_i - \sum_j W_{ij} X_j \|^2$$
  - **Compute embedding coordinates Y:**
    $$\Phi(Y) = \min \sum_i \| Y_i - \sum_j W_{ij} Y_j \|^2$$



MANCHESTER 1824

The University of Manchester

Nonlinear Representation

- **Grouping of linear/nonlinear mapping (Yin, FEEE, 2011)**

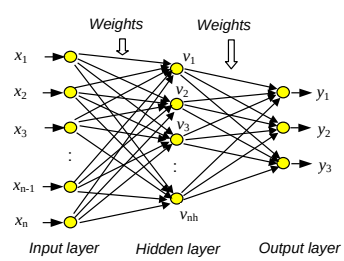
Eigen decomp. based	*MDS* based	*Principal manifold based*
PCA, KPCA	MDS	Principal Curve/Surface
LLE	Isomap	SOM/ViSOM/GTM
HLLE	CCA	…
Laplacian eigenmap	…	
Spectral clustering		
…		

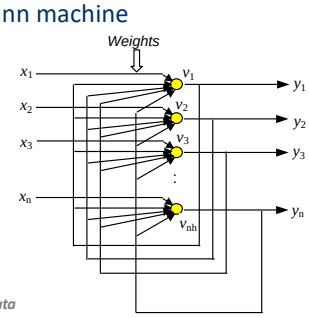
MANCHESTER 1824

The University of Manchester

Neural Networks

- Feed-forward Networks
  - Perceptron and multilayer perceptron
  - Radial basis function
  - Support vector machine
- Recurrent Networks
  - Hopfield networks
  - Boltzmann machine

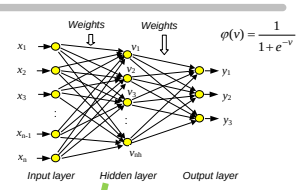




Neural Networks

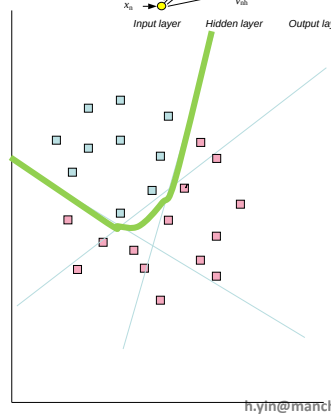
- **Multiplayer perceptron**

How does MLP form nonlinear separations?



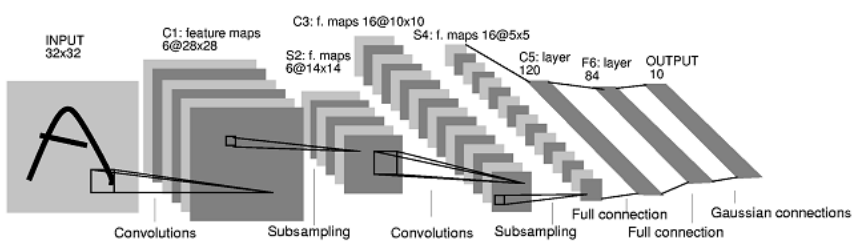
Key points:

- Each hidden node forms a linear separate boundary;
- An output node is a combination of all hidden nodes, in effect forming a piecewise linear (or nonlinear) separation boundary.



Deep Neural Networks

- **CNN** (*Convolutional Neural Network*) or **ConvNet**:
  feature layers (convolutional filters to extract features)
  pooling/subsampling layer (summarize or abstract responses of the filters, e.g. mean or max pooling)
  typical CNNs: LetNet5, AlexNet, VGG16, GoogLeNet



CNN LeNet5 Architecture (LeCun & Bottou 1998)

MANCHESTER 1824

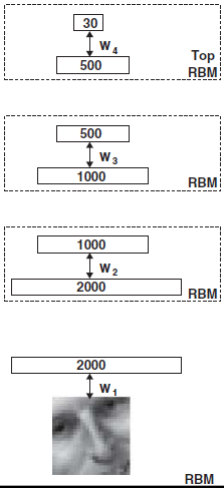
The University of Manchester

Deep Neural Networks

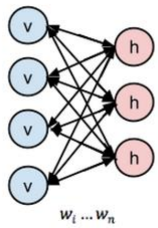
- **Deep Recurrent or Belief Networks**

  RBM (Restricted Boltzmann Machine): Stochastic NN with layers of both visible and invisible (latent) nodes to model probabilistic relations of inputs and latent variables.

Hinton & Salakhutdinov, 2006



A Symmetrical, Bipartite, Bidirectional Graph with Shared Weights



$$\Delta w_{ij} = \varepsilon(\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{reconstr})$$

Image courtesy of deeplearning4j.org

MANCHESTER 1824

The University of Manchester

Deep Neural Networks

- Deep NN or Deep learning has demonstrated **significant improvements** over the conventional shallow NNs (with one/no hidden layer) in increasing number of **real-world applications** such as image/object recognition.
- **Training** Deep NNs takes much long time due to many layers and often requires GPUs; vanishing gradients.
- Deep learning is making great impacts in general **AI, gaming, robotics/autonomous systems, and many other fields**; and will shrive in the next few years.

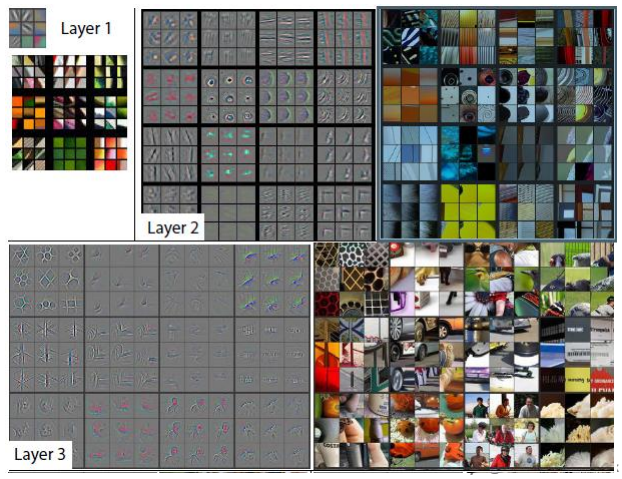
h.yin@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

Deep Representation

Visualizing Features of ConvNet (AlexNet):



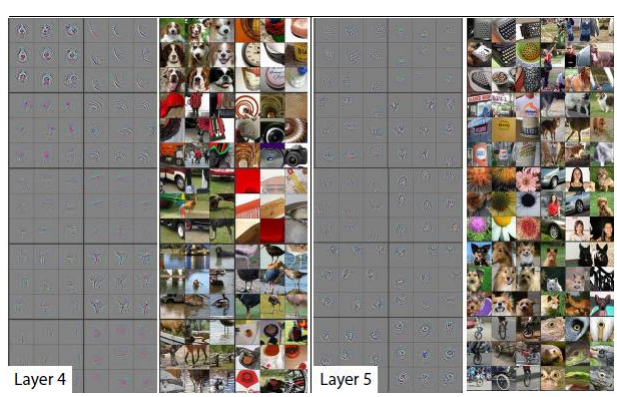
*From **Zeiler & Fergus, ECCV2014***

MANCHESTER 1824

The University of Manchester

Deep Representation

Visualizing Features of ConvNet (AlexNet):



*From **Zeiler & Fergus, ECCV2014***

h.yin@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

Manifold

Mannigfaltigkeit

Topology

Set

Curvature

Neighbourhood

流形

Surface

Topological space

Metric

Topological manifold

chart

Tensor

Metric space

Tangent space

Riemannian metric

Differentiable manifold

Riemannian manifold

Geodesic

Differential geometry

Intrinsic

Invariance

Dimensionality reduction

Retinotopic mapping

Data visualisation

MANCHESTER 1824

The University of Manchester

Manifold

Intuitively, a **manifold** is a generalization of curves and surfaces to higher dimensions. It is locally Euclidean in that every point has a neighborhood, called a chart, homeomorphic to an open subset of $\mathbb{R}^n$. The coordinates on a chart allow one to carry out computations as though in a Euclidean space, so that many concepts from $\mathbb{R}^n$, such as differentiability, point-derivations, tangent spaces, and differential forms, carry over to a manifold.

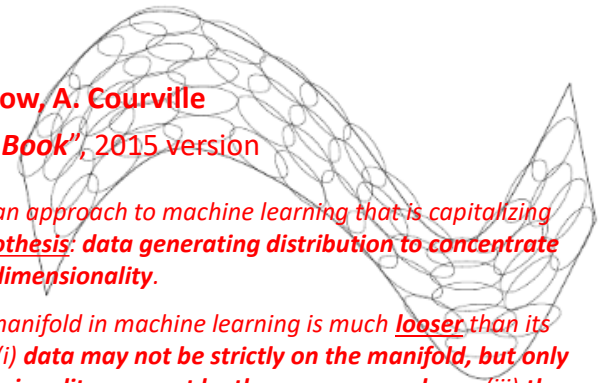
L. Tu "*An Introduction to Manifolds*"

MANCHESTER 1824

The University of Manchester

Manifold

Manifold is a topological space that is locally Euclidean



Y. Bengio, I Goodfellow, A. Courville

The "***Deep Learning Book***", 2015 version

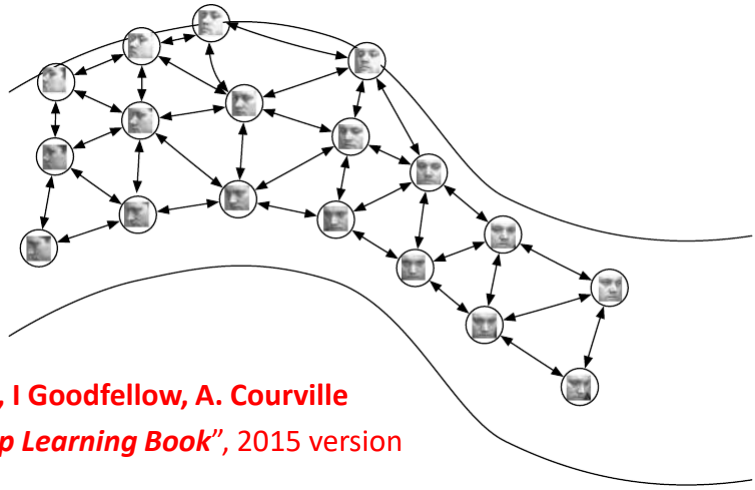
- *Manifold learning is an approach to machine learning that is capitalizing on the* ***manifold hypothesis****:* ***data generating distribution to concentrate near regions of low dimensionality****.*
- *The use of the term manifold in machine learning is much* ***looser*** *than its use in mathematics: (i)* ***data may not be strictly on the manifold, but only near it****; (ii)* ***the dimensionality may not be the same everywhere****; (iii)* ***the notion actually referred to in machine learning naturally extends to discrete spaces****.*

MANCHESTER 1824

The University of Manchester

Manifold

Manifold is a topological space that is locally Euclidean



Y. Bengio, I Goodfellow, A. Courville

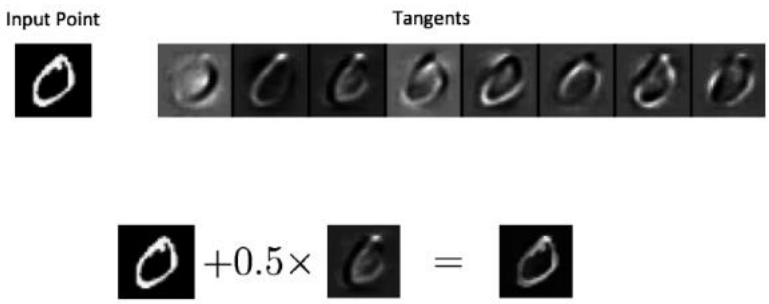
The "***Deep Learning Book***", 2015 version

MANCHESTER 1824

The University of Manchester

Manifold

Manifold is a topological space that is locally Euclidean



Y. Bengio, I Goodfellow, A. Courville

The "***Deep Learning Book***", 2015 version

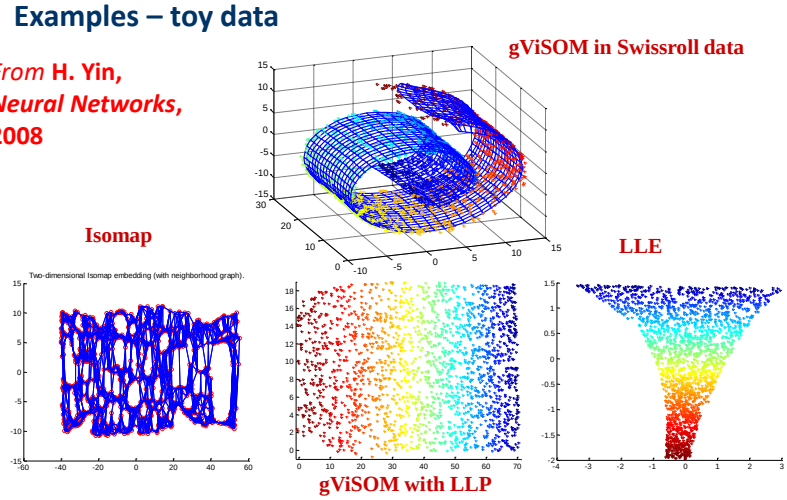
MANCHESTER 1824

The University of Manchester

Learning Data Manifold

- **Examples – toy data**

From **H. Yin, *Neural Networks*, 2008**

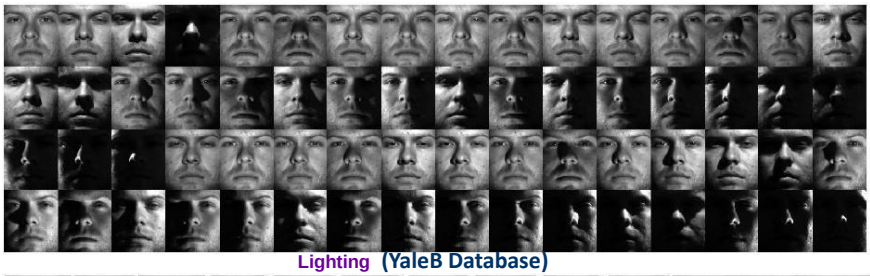


MANCHESTER 1824

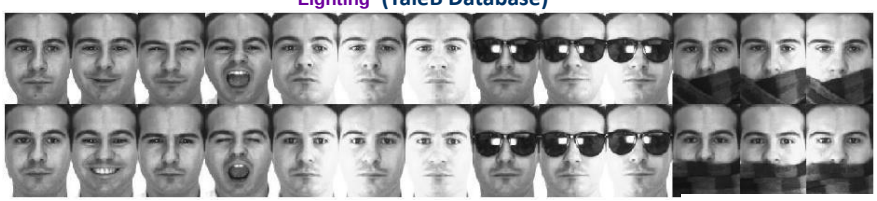
The University of Manchester

Manifold & Data Variations

- **Examples – *images* (Huang & Yin, *Img. & Vis. Comp.* 2012)**



Lighting (YaleB Database)



Expression Lighting Occlusion&Lighting Occlusion&Lighting

(AR Database)

h.yin@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

Autoassociative NNs/Autoencoder

- **Autoassociative Neural Networks (Kramer, *AIChE*, 1991)**

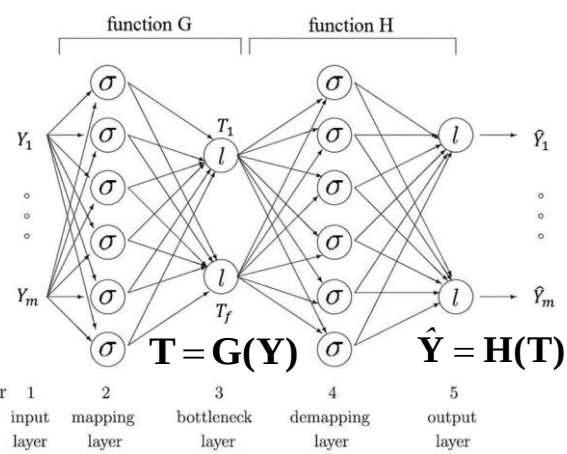
PCA

$$\mathbf{T} = \mathbf{V}^T \mathbf{Y}$$

Y: data matrix
V: eigenvector matrix
T: principal components

$$\hat{\mathbf{Y}} = \mathbf{V}\mathbf{T}$$

$\hat{Y}$: reconstruction



$$min\, E(\mathbf{Y} - \hat{\mathbf{Y}})$$

Via
back-propagation
self-supervised learning

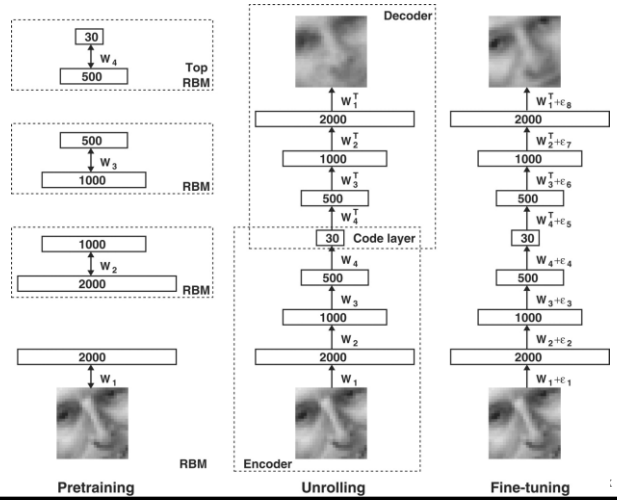
Diagram *from*, Z. Sadough, et al, *J. Eng. Gas Turbines Power* , 2014

MANCHESTER 1824

The University of Manchester

Deep Autoencoder (DAE)

- **Deep Autoencoder (Hinton & Salakhutdinov, *Science*, 2006)**



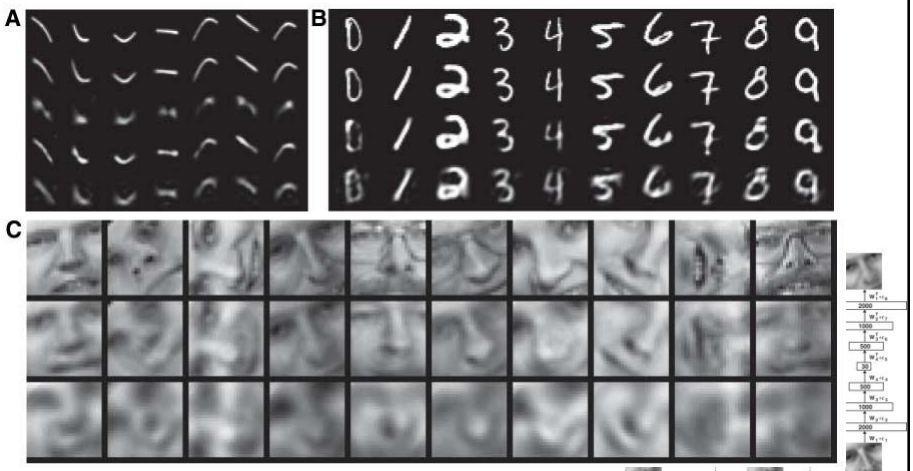
Manifold, Deep Learning, Big Data

MANCHESTER 1824

The University of Manchester

Deep Autoencoder (DAE)

- **Deep Autoencoder (Hinton & Salakhutdinov, *Science*, 2006)**



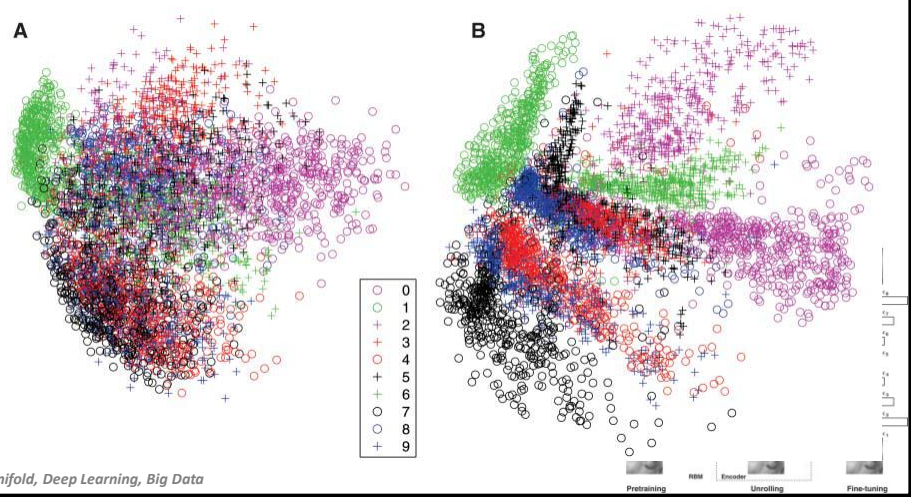
Manifold, Deep Learning, Big Data

MANCHESTER 1824

The University of Manchester

Deep Autoencoder (DAE)

- **Deep Autoencoder (Hinton & Salakhutdinov, *Science*, 2006)**



MANCHESTER 1824

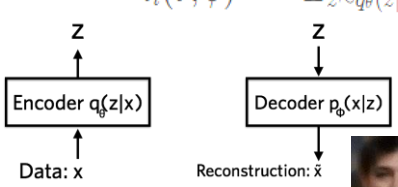
The University of Manchester

Variational Autoencoder (VAE)

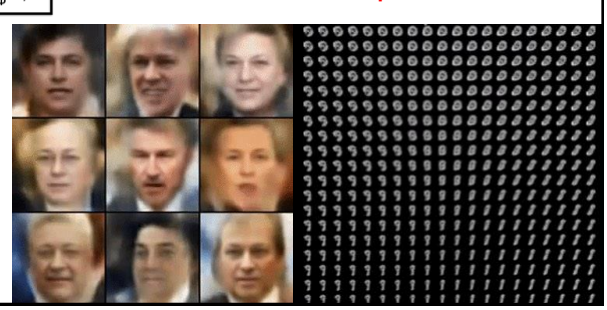
- **VAEs (Kingma & Welling, 2014, 2015)**

from the tutorial by Jaan Altosaar: **http://jaan.io**

$$l_i(\theta,\phi) = -E_{z\sim q_\theta(z|x_i)}[\log p_\phi(x_i|z)] + KL(q_\theta(z|x_i)||p(z))$$



← low-dimensional/latent space is stochastic



MANCHESTER 1824

The University of Manchester

Deep Neural Networks

- **ConvNets + VAEs (Brock, et al, *arXiv*, 2016)**

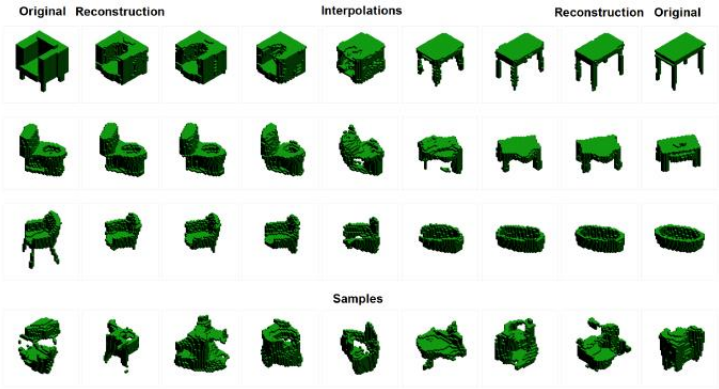
ConvNets as VAEs
for voxel/3D object modelling with RGB-D data

MANCHESTER 1824

The University of Manchester

Deep Neural Networks

- **ConvNets + VAEs (Brock, et al, *arXiv*, 2016)**



… the samples consistently **bear a semblance of structure**, with **few to no free floating voxels**, suggesting that the decoder network has learned to **maintain output voxel connectivity** regardless of the latent configuration. The major limitation of the VAE is that its generated **samples do not, however, resemble real objects** ….

MANCHESTER 1824

The University of Manchester

Deep Representation

- **Unsupervised Learning DNN Features**

*from **Hankins, Peng & Yin, WCCI2018***

TABLE II
ERROR RATE ON MNIST

Method	Error rate (%)
CDBN [26]	0.82
CSOM (linear SVM) [16]	0.82
ConvNet [27]	0.53
ScatNet-2 (RBF SVM) [10]	0.43
MRF-CNN [19]	0.38
PCANet-2 [1]	0.66
PCANet-2 (ours)	0.77
DCTNet	0.74
DCTNet (TR Norm)	0.76
$\text{DCTNet}_{overlap}$	0.68
$\text{DCTNet}_{overlap}$ (TR Norm)	0.68
SOMNet	0.86 ± 0.05
SOMNet (fine tune)	0.83 ± 0.07
SOMNet_{16-32}	0.65 ± 0.02
SOMNet_{16-32} (fine tune)	- ± -
SOMNet_{rep}	1.01
MRF-SOMNet	1.24

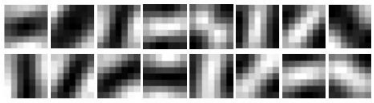


Fig. 3. Learned SOMNet filters on MNIST. Top row: the first layer filter bank. Bottom row: the second layer filter bank.

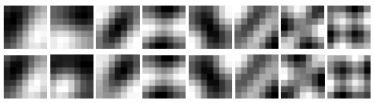


Fig. 4. Learned PCANet filters on MNIST. Top row: the first layer filter bank. Bottom row: the second layer filter bank.

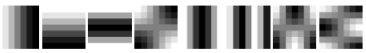


Fig. 5. Generated DCTNet filters for MNIST. Replicated across both filter banks.

Fig. 6. Clustered MRF filters for MNIST. Replicated across both filter banks.

MANCHESTER 1824

The University of Manchester

Deep Representation

- **Unsupervised Learning DNN Features**

*from **Hankins, Peng & Yin, WCCI2018***

TABLE III
ACCURACY ON CIFAR-10

Method	Accuracy (%)
K-means (Triangle, 4000 features) [31]	79.60
Stochastic Pooling ConvNet [32]	84.87
NIN + Dropout [33]	89.59
MRF-CNN [19]	91.25
PCANet-2 [1]	77.14
SOMNet (fine tune)	71.81 ± 0.06
MRF-SOMNet (fine tune)	71.1

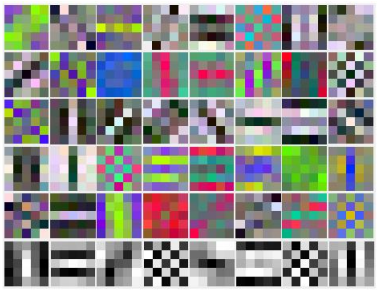


Fig. 7. Learned SOMNet filters on CIFAR-10. Top five rows: the first layer filter bank. Bottom row: the second layer filter bank.

h.yin@manchester.ac.uk

MANCHESTER 1824

The University of Manchester

Deep Representation

- **Pre-generated DNN Early Features**

*from **Peng & Yin, IDEAL2017***

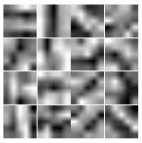


Table 1. Classification results on the MNIST dataset.

Methods	Error Rate (%)
CNN [22]	0.53
Stochastic Pooling [24]	0.47
Maxout [35]	0.45
DropConnect [36]	0.57
Network in Network [23]	0.47
ScatNet [26]	0.43
PCANet [27]	0.62
MRF-CNN	**0.38**±0.01

Table 3. Classification results on the CIFAR-10 dataset.

Methods	Error Rate (%)
Stochastic Pooling [24]	15.13
Maxout [35]	11.68
DropConnect [36]	9.32
Network in Network [23]	10.41
All-CNN [25]	**9.08**
MRF-CNN	9.87±0.16

MANCHESTER 1824

The University of Manchester

Some Recent Studies

- **ConvNet with BGP on LFW (Huang & Yin, *Pattern Recognition*, 2017)**

Table 7
Comparisons with state-of-the-art deep learning based approaches on the LFW. Only single model performances are compared.

State-Of-The-Art Results				Our Results		
DeepFace [77]	Wen et. al [78].	DeepID2+ [79]	VGGface [80]	RGB	BGP	RGB+BGP
97.35%	97.37%	98.70%	98.95%	98.43%	98.82%	99.32%

- **Video synthesis (demo)**

MANCHESTER 1824

The University of Manchester

Team

Previous PhDs and Post-Doc's/Research Associates

- Dr Shireen Zaki (Astro Malaysia)
- Dr Yicun Ouyang (Shenzhen)
- Dr. Aftab Khan (NUST)
- Dr. James Burnstone (eLucid mHealth)
- Dr. Weilin Huang (Oxford)
- Dr. Zareen Mehboob (Surrey)
- Dr. He Ni (Zhejiang Univ)
- Dr. Israr Hussain (Gov.)
- Dr. Lei A. Clifton (Oxford)
- Dr. Carla Möller-Levet (Surrey)
- MPhil Swapna Sarvesvaran
- Dr. Richard Freeman (Michael Page)
- Dr. James Burstone (CEO, eLucid)
- Dr. King-wai Lau (UCL)
- Dr. Qingfu Zhang (Essex)
- Dr. Michel Haritopoulos (France)
- Ann Gledson
- Ben Russell
- Dr. Bicheng Li
- Dr. Bruno Baruque (Burgos)
- Dr. Jose A. Costa (Natal)
- Dr. Lianxiang Zhu
- Dr. Xiaohong Chen (NUAA,CSC)

Current PhDs & post-doc

- **Ananya Gupta, Deep Learning for Object &Volumetric Estimate**
- **Yao Peng, Deep Learning in Face Expression Recognition**
- **Ali Alsuwaidi, Hyperspectral Image Classification**
- **Richard Hankins, Action Recognition in Robotics**
- **Jingwen Sun, DNN Structure Optimization**
- **Dr Qing Tian, Multi-view Ordinal CCA (NUIST)**

MANCHESTER 1824

The University of Manchester

Surmmary

- **Data representation (features) are important to any follow-on classification, recognition and modelling tasks**
- **Manifold hypothesis says high dimensional data lies in lower dimensionality or low-dimensional submanifolds**
- **Deep Learning features are extensions of linear manifold/sub-manifold in multiple and hierarchical fashion**
- **Understanding data representation can help build/design better classifiers or analytic tools**