Title: Machine Learning Lecture - 230320

Speakers: Lauren Hayward

Collection: Machine Learning for Many-Body Physics (2022/2023)

Date: March 20, 2023 - 2:00 PM

URL: https://pirsa.org/23030040

Today: Unsupervised Learning (UL)

↳ examples of UL tasks

↳ relationship to SL

↳ dimensional reduction using principal component analysis (PCA) g (UL)

ing
(PCA)

Unsupervised Learning (UL)

Goal: learn structural properties of an unlabelled dataset

Dataset $\mathcal{D} = \{\vec{x}\}$ where:

- each datapoint has N components $\vec{x} = (x_1, x_2, \dots, x_N)$
- $\mathcal{D}$ contains M datapoints: $\mathcal{D} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M\}$

$x_2, \dots, x_N)$

$, \vec{x}_M \}$

We will study two classes of UL tasks:

① Dimensional reduction (today, next lecture, Tutorial 5, Homework 2)

Idea: reduce the dimensionality N of the dataset by finding the important features or correlations in the data.

If we reduce $N \to N'$ with $N' = 2$ or 3 then we can easily visualize

In some cases, our dataset will naturally form clusters in the lower-dim. rep.

eg) Ising model configs. at various temperatures might form clusters in a 2D representation as:

② Generative modelling (tomorrow, Lecture #12, course in May)

Idea: use the datapoints $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M$ to learn about the underlying prob. dist. $p(\vec{x})$ and generate new samples from it.

Relationship between SL and UL:

UL: $\mathcal{D} = \{\vec{x}\} \longrightarrow$ learn prob. dist. $p(\vec{x})$

SL: $\mathcal{D} = \{(\vec{x}, \vec{y})\} \longrightarrow$ learn conditional dist. $p(\vec{y}|\vec{x})$

tomorrow, Lecture #12,
course in May

$_1, \vec{x}_2, \ldots, \vec{x}_M$ to learn

dist. $p(\vec{x})$ and

m it.

d UL:

prob. dist. $p(\vec{x})$

Conditional dist. $p(\vec{y}|\vec{x})$

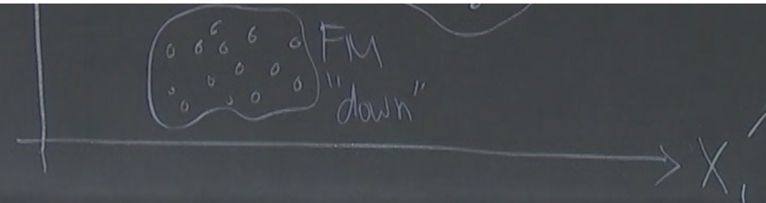
UL and SL are not formally different due to the chain rule of probability:

For $\vec{x} = (x_1, x_2, \ldots, x_N)$

$$p(\vec{x}) = p(x_N | x_1, x_2, \ldots, x_{N-1})\, p(x_1, x_2, \ldots x_{N-1})$$

$$= \ldots$$

$$= \left[\prod_{i=2}^{N} p(x_i | x_1, x_2, \ldots, x_{i-1})\right] p(x_1)$$



$\Rightarrow$ can express a UL problem as $N-1$
SL problems plus the easier problem of learning $p(x_1)$

(eg: for indep. spin configs. with $\mathbb{Z}_2$ symmetry

$$p(x_1) = p(\uparrow) = p(\downarrow) = \frac{1}{2}$$
)

We can also express a SL problem in terms of UL problems: p blem as N-1

asier problem of learning $p(x_1)$

gs. with $\mathbb{Z}_2$ symmetry

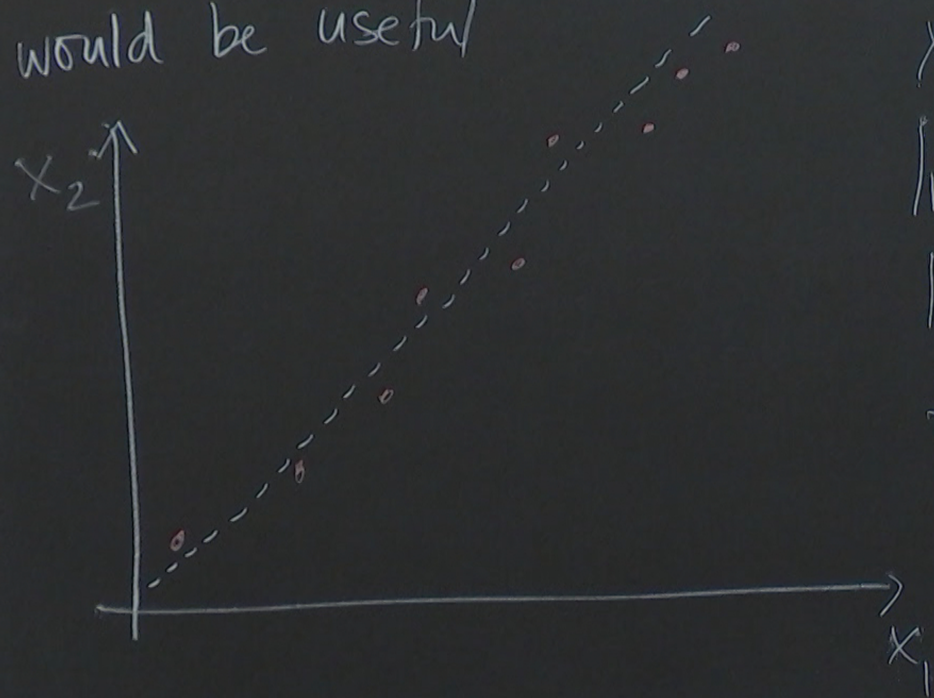
$(\downarrow) = \frac{1}{2}$

roblem in terms of UL problems: $p(\vec{y}|\vec{x}) = \frac{p(\vec{x},\vec{y})}{\sum_{\vec{y}'} p(\vec{x},\vec{y}')}$

Dimensional reduction using Principal Component Analysis (PCA)

Idea: Generate a lower-dim. repr. of the N-dim. datapoints $\vec{x}$ by applying a linear transformation

Let's look at an example for $N=2$ where PCA would be useful



x_1 and x_2 are highly linearly correlated, so knowing x_1 is enough to infer x_2 with high precision

Store the M datapoints $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_M$ in an $M \times N$ matrix:

$$X = \begin{bmatrix} \vec{x}_1 \\ \vec{x}_2 \\ \vdots \\ \vec{x}_M \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & & \\ \vdots & & \ddots & \\ x_{M1} & & & x_{MN} \end{bmatrix}$$

For PCA, the mean of the datapoints must be zero, so first shift the "centre" of each column:

$$\vec{x}_i^{\,c} = \vec{x}_i - \frac{1}{M}\sum_{k=1}^{M}\vec{x}_k$$

$$\left(X_{ij}^{c} = X_{ij} - \frac{1}{M}\sum_{k=1}^{M} X_{kj} \quad \forall j \right)$$

$$1 \le i \le M, \quad 1 \le j \le N$$

Now consider an $N \times N$ matrix V_X

$$V_X = \frac{1}{M-1} (X^c)^T X^c$$

↳ diag. components store the variance of each component of $\vec{X}$

↳ off-diag elements store covariances between components

$1 \leq i \leq M, \quad 1 \leq j \leq N$

Goal of PCA: Find a new repr. $X' = X^c P$ such that $V_{X'}$ is diagonal

The matrix P is $N \times N$ and the "principal components" of X

Now:

$$V_{X'} = \frac{1}{M-1} X'^T X' = \frac{1}{M-1} (X^c P)^T X^c P = \frac{1}{M-1} P^T (X_c)^T X_c P = P^T V_X P$$

When $V_{X'}$ is diagonal:

$$P^T V_X P = D$$

Since $V_X = \frac{1}{M-1}(X^c)^T X^c$ is symmetric, its eigenvecs. can be chosen to be orthogonal so that $P^{-1} = P^T$

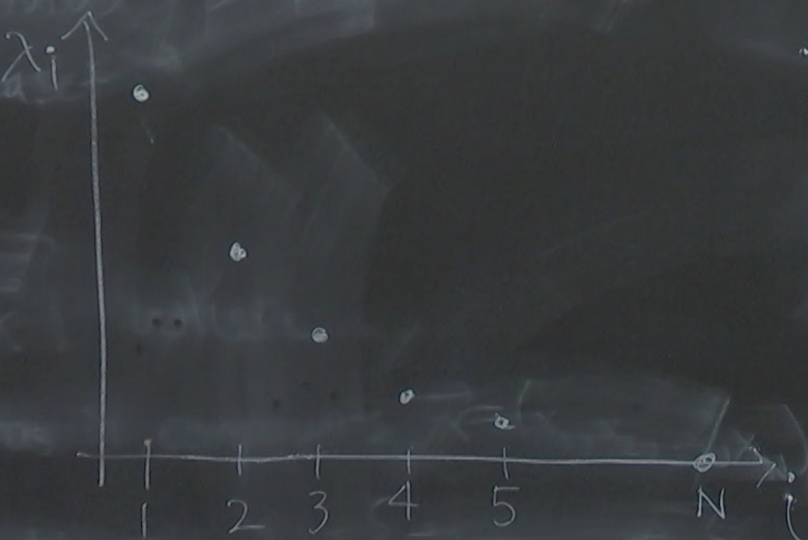
So PCA amounts to solving the eigenvalue problem:

$$V_X = P D P^{-1}$$

where $D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{pmatrix}$ and $\lambda_j = \dfrac{1}{M-1} \displaystyle\sum_{i=1}^{M} X'^{\,2}_{ij} \geq 0$

Sort so $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$

Now the first principal component X_1' corresponds to the direction in space with the highest variance λ_1



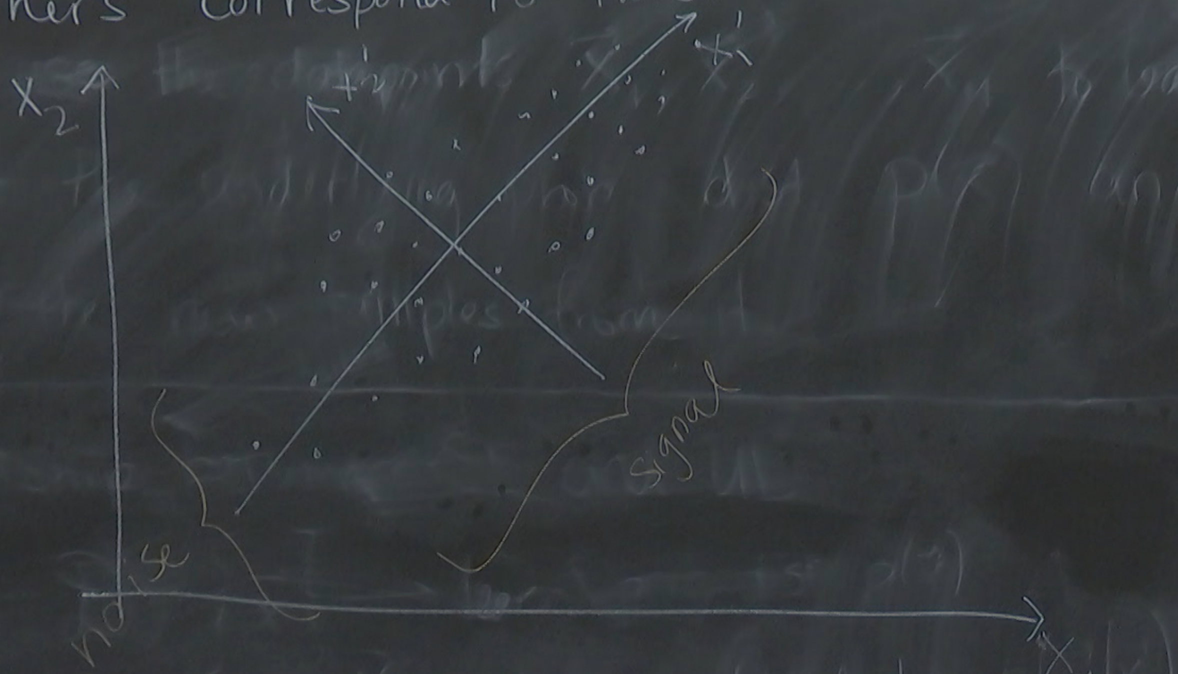
If the λ_i "quickly" become small, we expect a lower-dim. repr. defined by the first few principal components to be a valid and efficient representation

Consider the "explained variance ratio" r_i :

$$r_i = \frac{\lambda_i}{\sum_j \lambda_j}$$

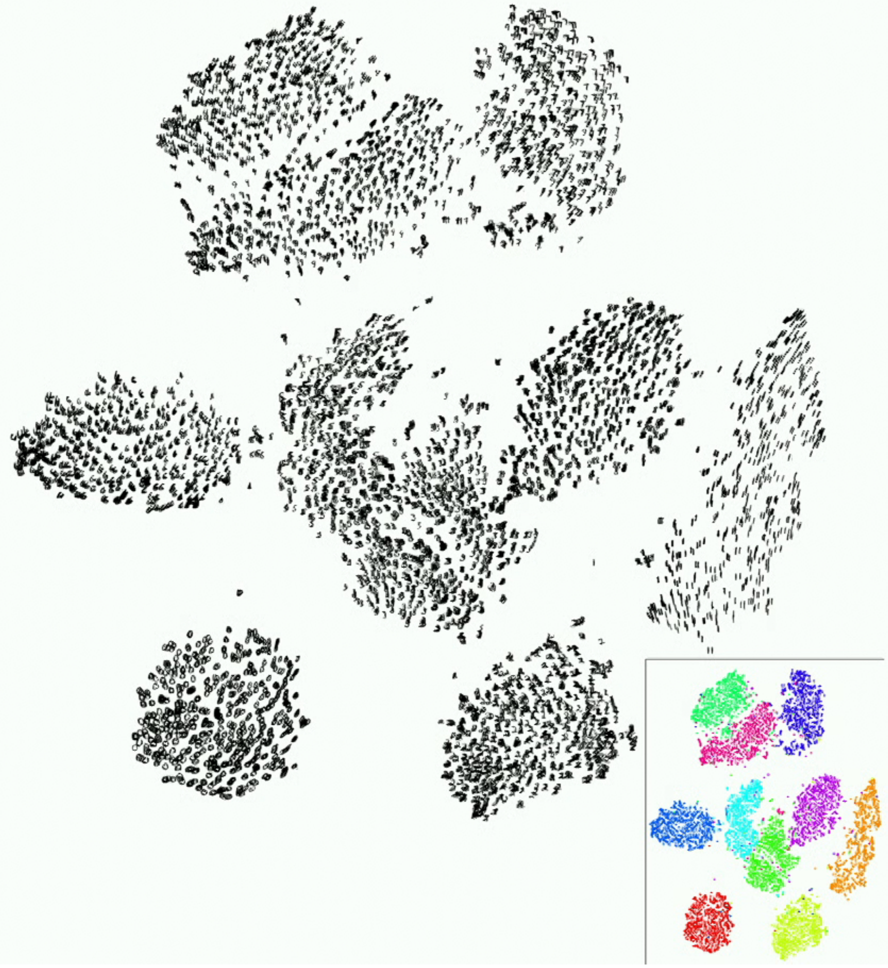
measure how much of the variance in the data can be explained by principal component x_i'

Directions with large variance correspond to "signal", while others correspond to noise:



Dimensional reduction in 2D

van der Maaten, https://lvdmaaten.github.io/tsne



PERIMETER SCHOLARS INTERNATIONAL

Dimensional reduction in 2D

Mehta et al, arXiv:1803.08823

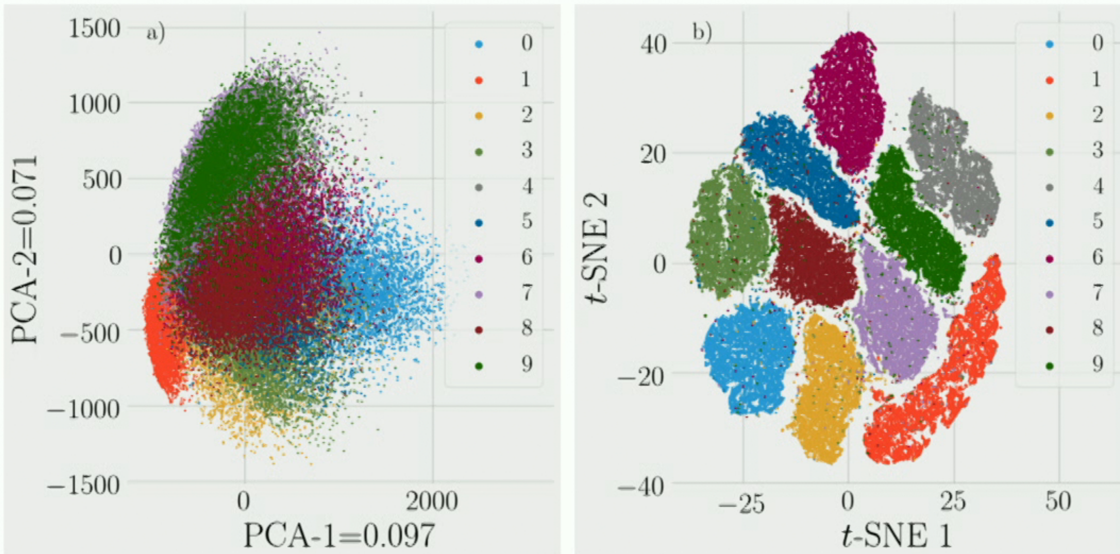
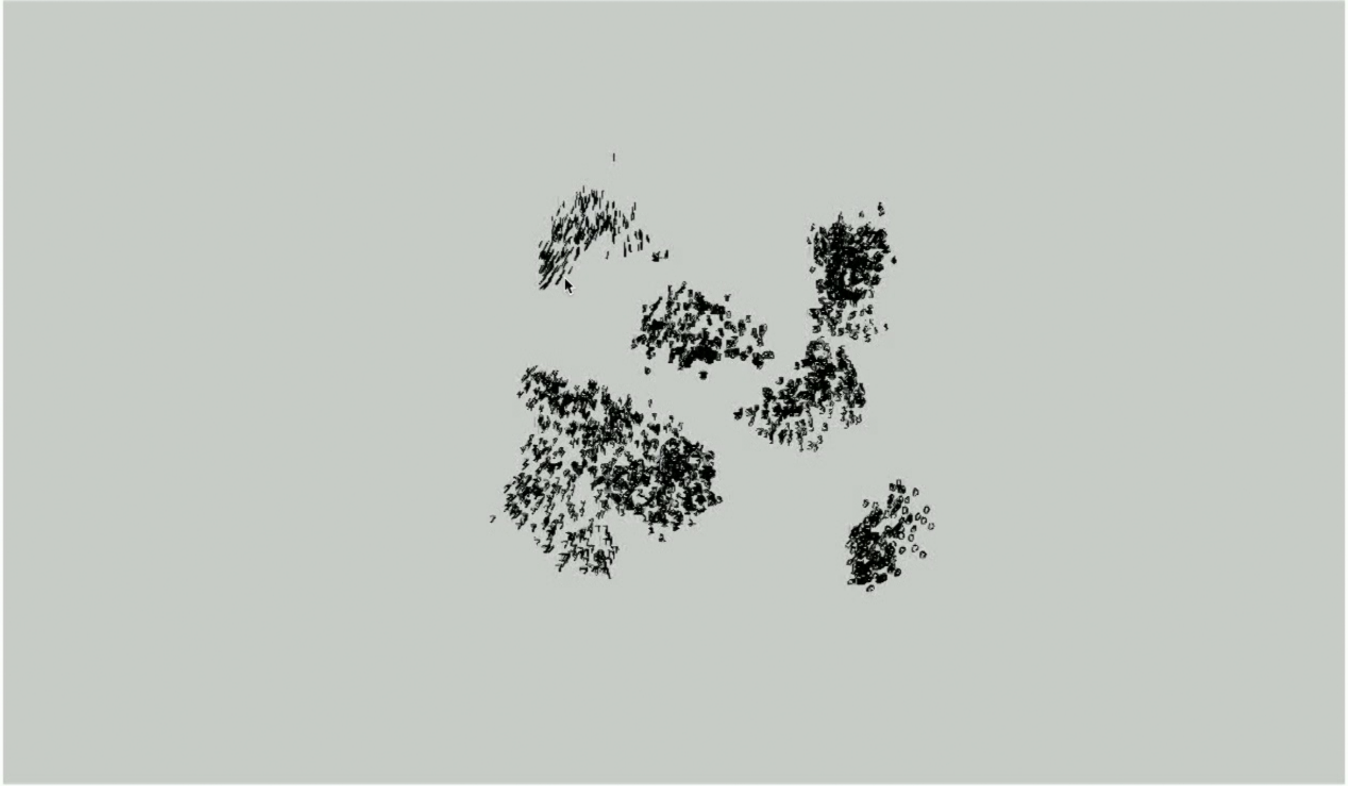


FIG. 54 Visualization of the MNIST handwritten digits training dataset (here $N = 60000$). (a) First two principal components. (b) t-SNE

PERIMETER SCHOLARS INTERNATIONAL

Dimensional reduction in 3D

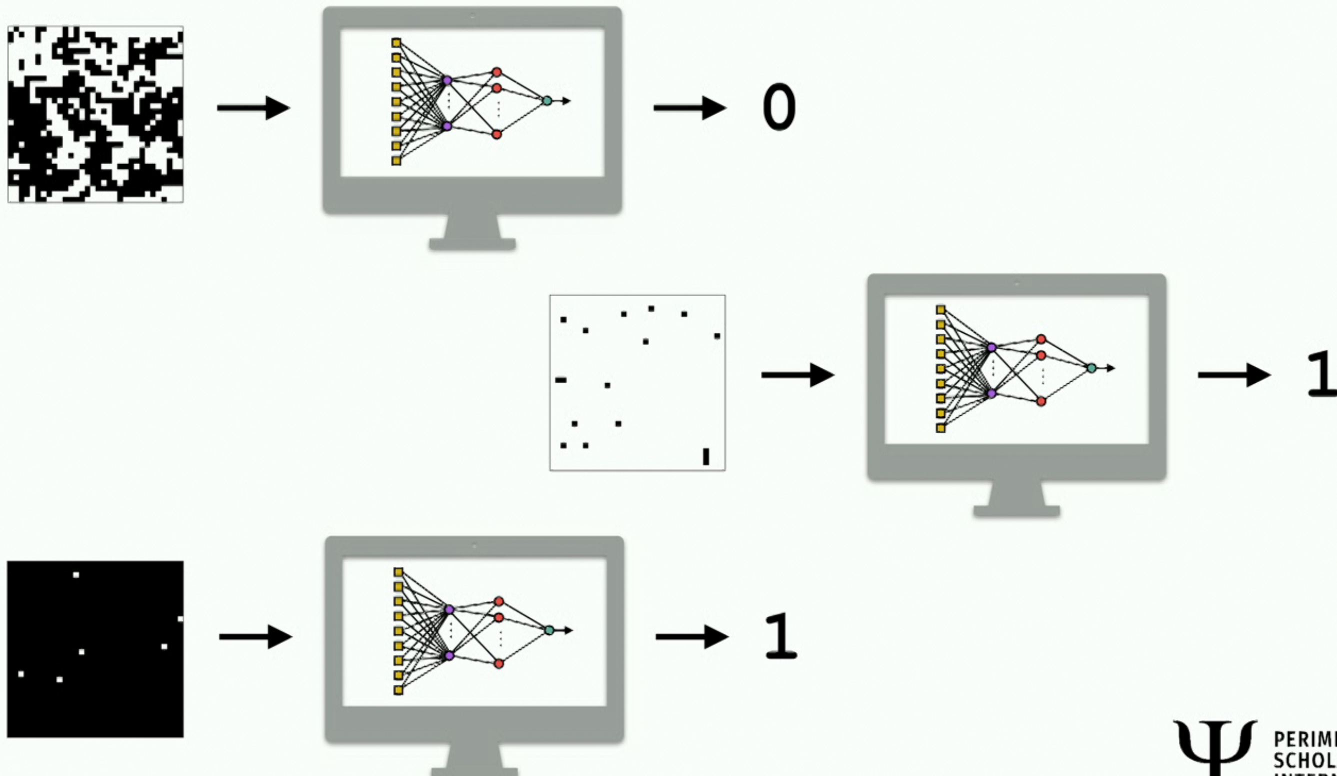
van der Maaten, https://lvdmaaten.github.io/tsne



PERIMETER SCHOLARS INTERNATIONAL

Classifying phases of matter with supervised learning

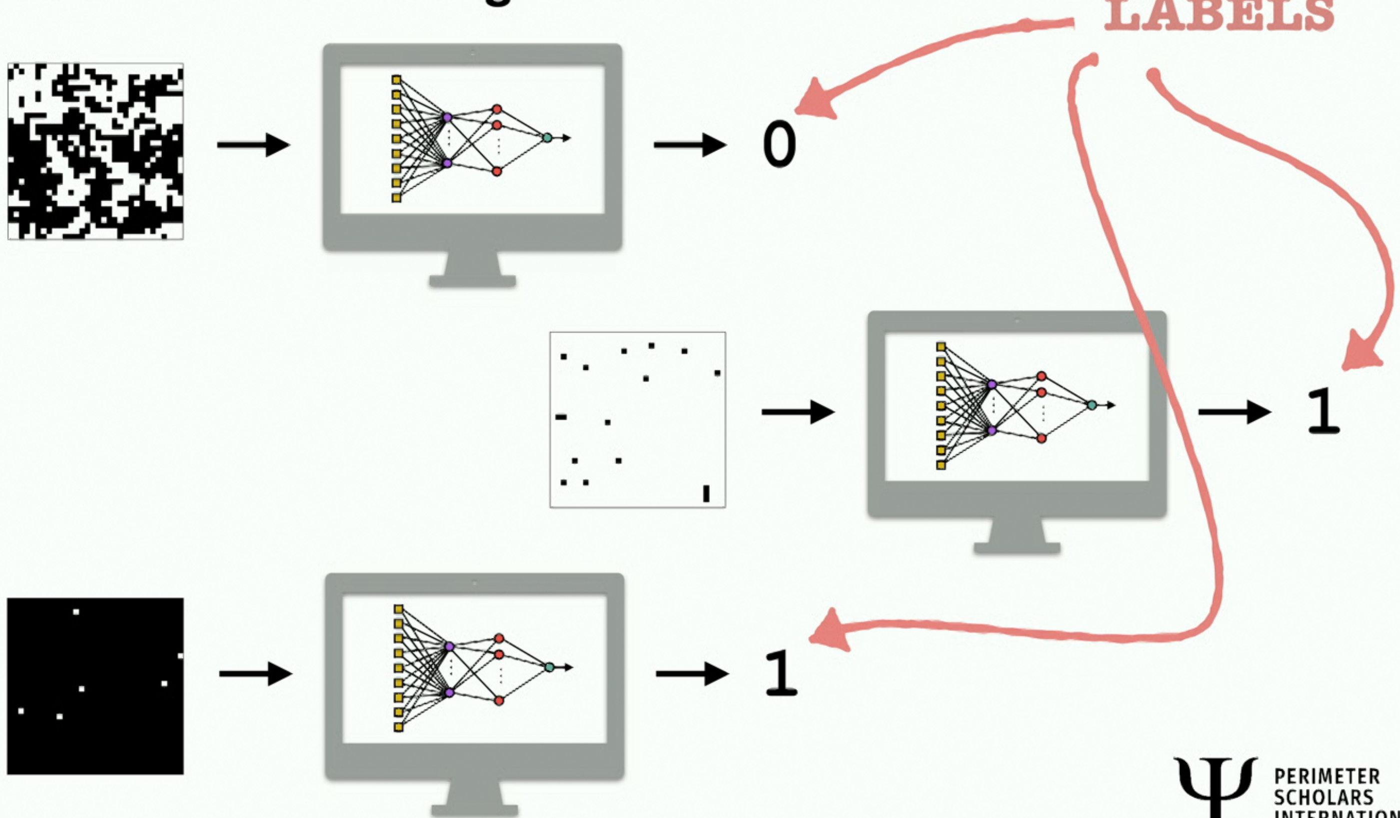
Two-dimensional Ising model:



PERIMETER SCHOLARS INTERNATIONAL

Classifying phases of matter with supervised learning

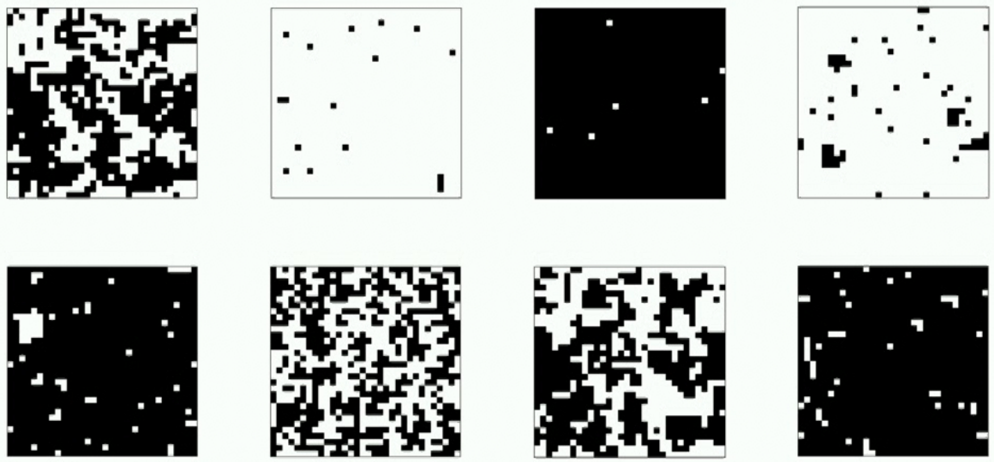
Two-dimensional Ising model:



PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter without labels

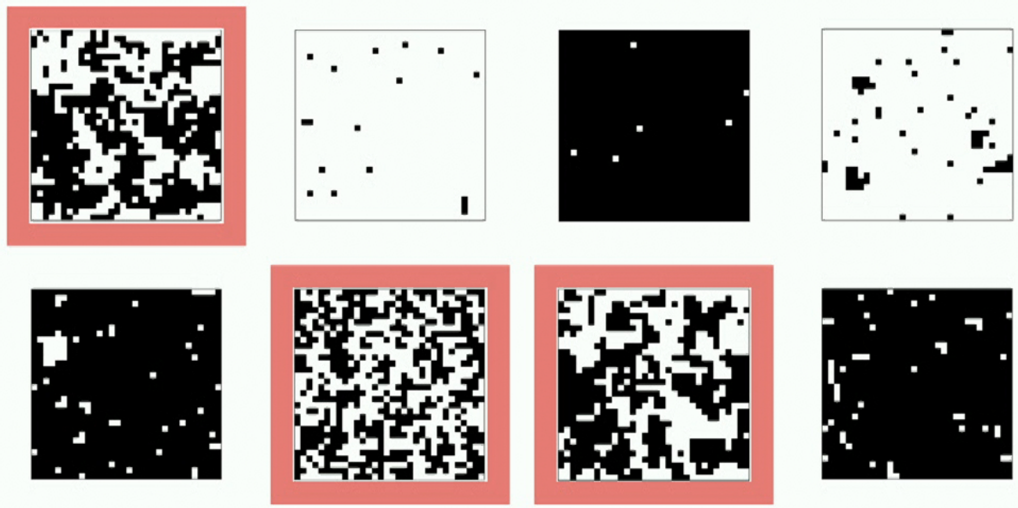
Two-dimensional Ising model:



PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter without labels

Two-dimensional Ising model:

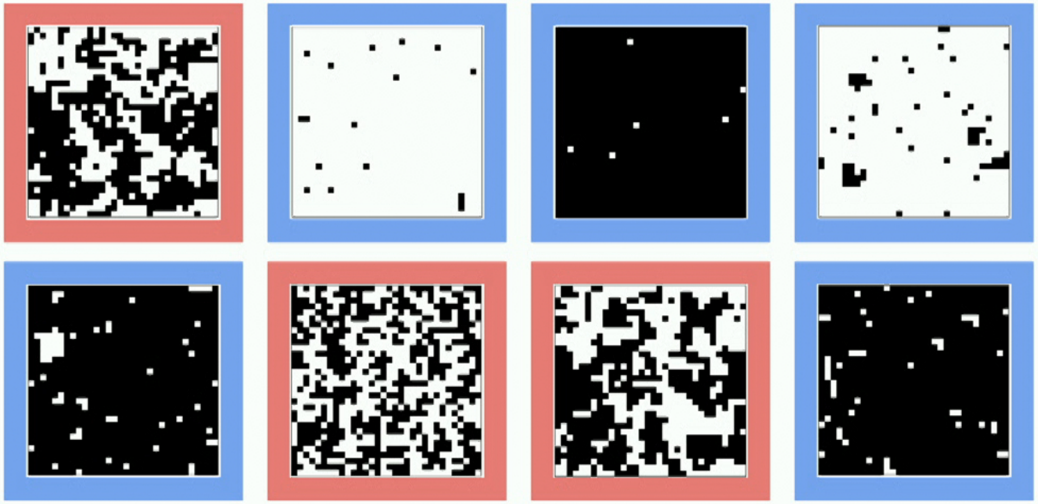


Group #1

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter without labels

Two-dimensional Ising model:

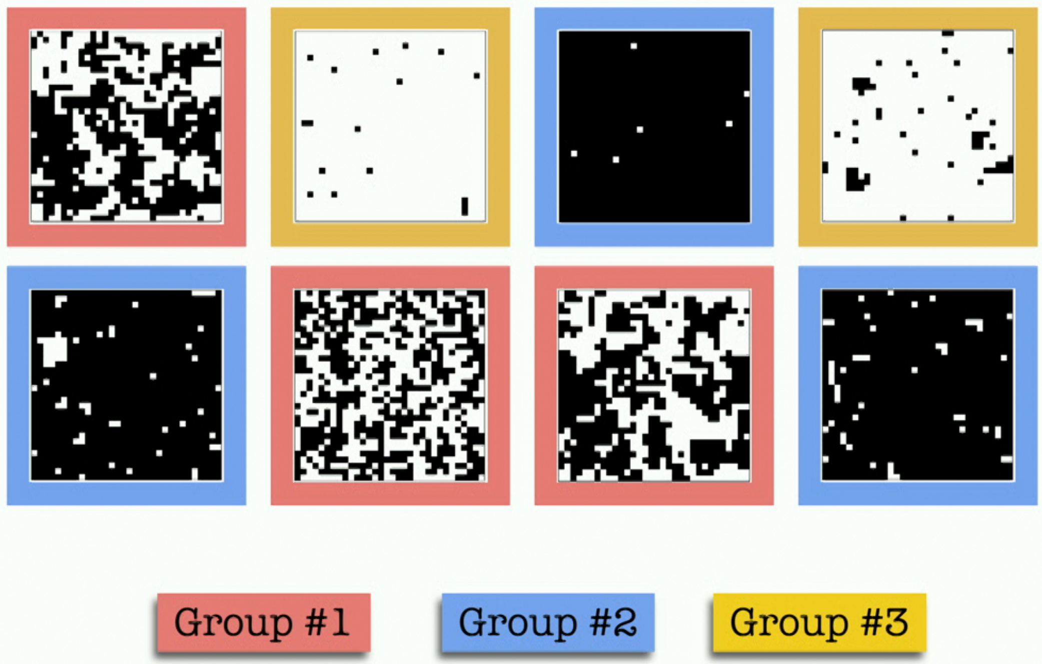


Group #1 Group #2

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter without labels

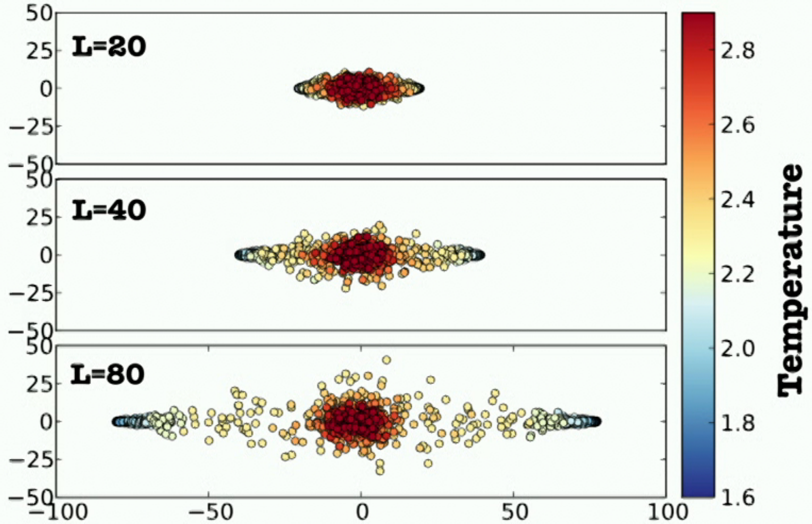
Two-dimensional Ising model:



PERIMETER
SCHOLARS
INTERNATIONAL

Identifying phases of matter with PCA

Two-dimensional Ising model:

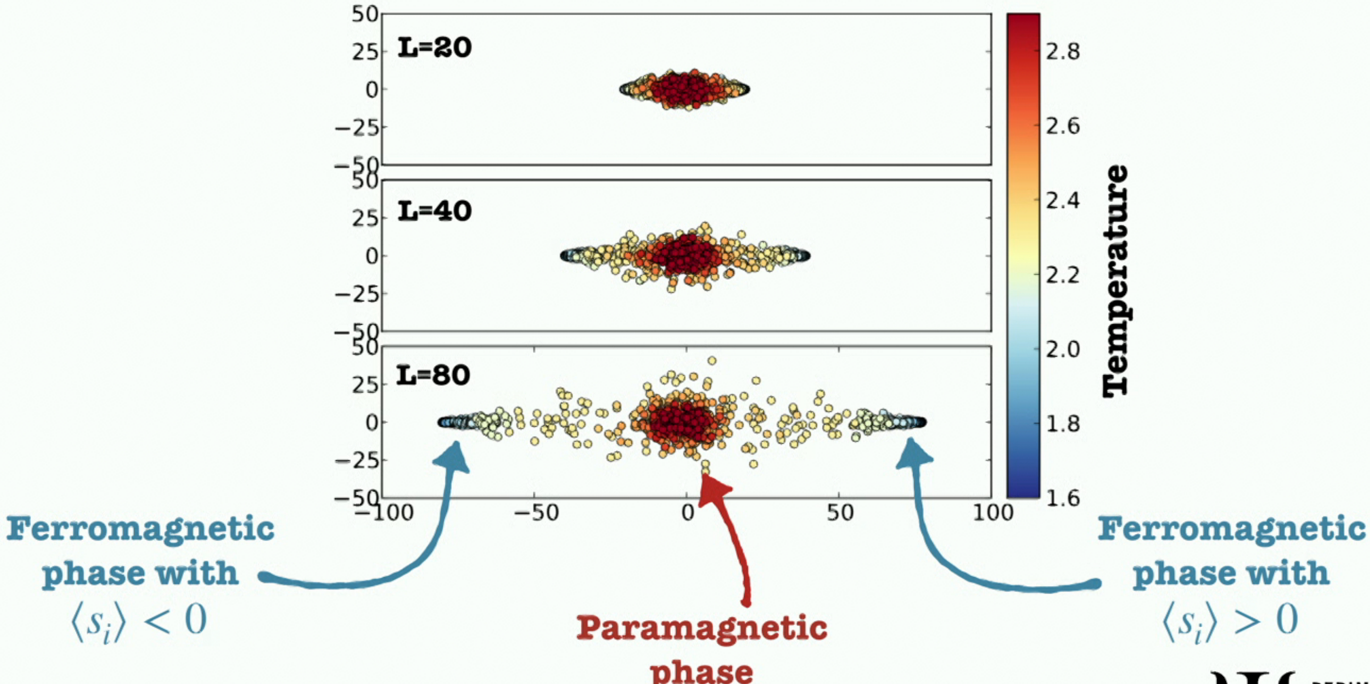


L. Wang, Phys. Rev. B **94**, 195105 (2016)

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter with PCA

Two-dimensional Ising model:

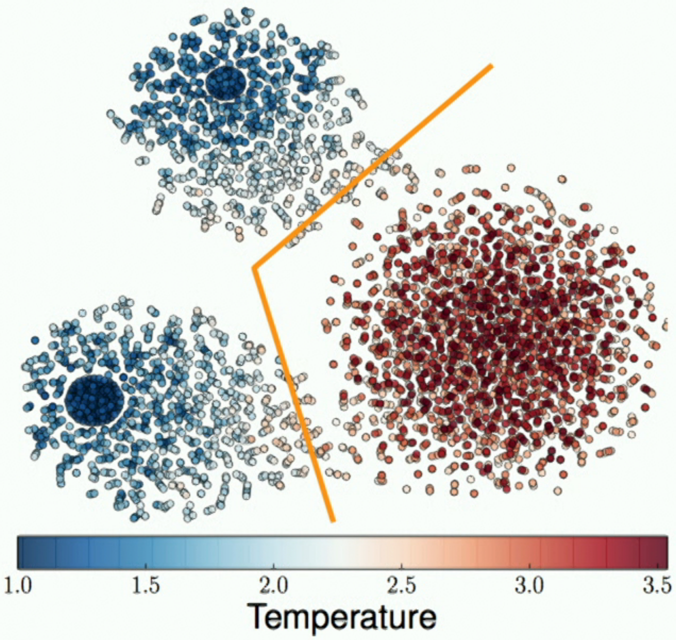


L. Wang, Phys. Rev. B **94**, 195105 (2016)

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter with t-SNE

Two-dimensional Ising model:

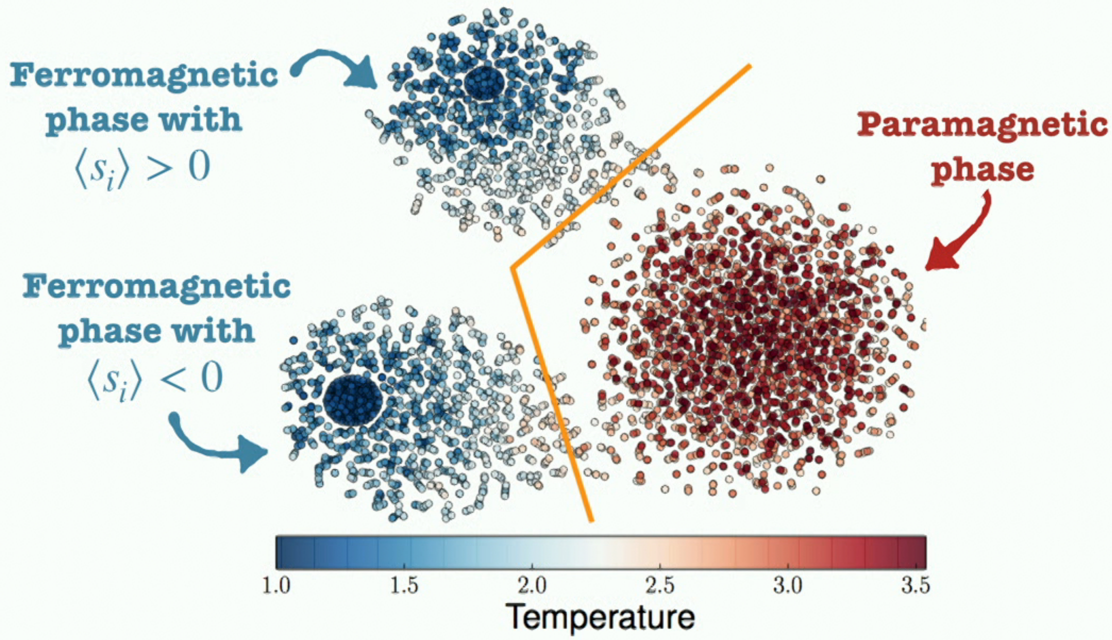


Carrasquilla and Melko, Nature Physics **13**, 431 (2017)

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter with t-SNE

Two-dimensional Ising model:



Carrasquilla and Melko, Nature Physics **13**, 431 (2017)

PERIMETER SCHOLARS INTERNATIONAL