Title: Machine Learning (2021/2022)

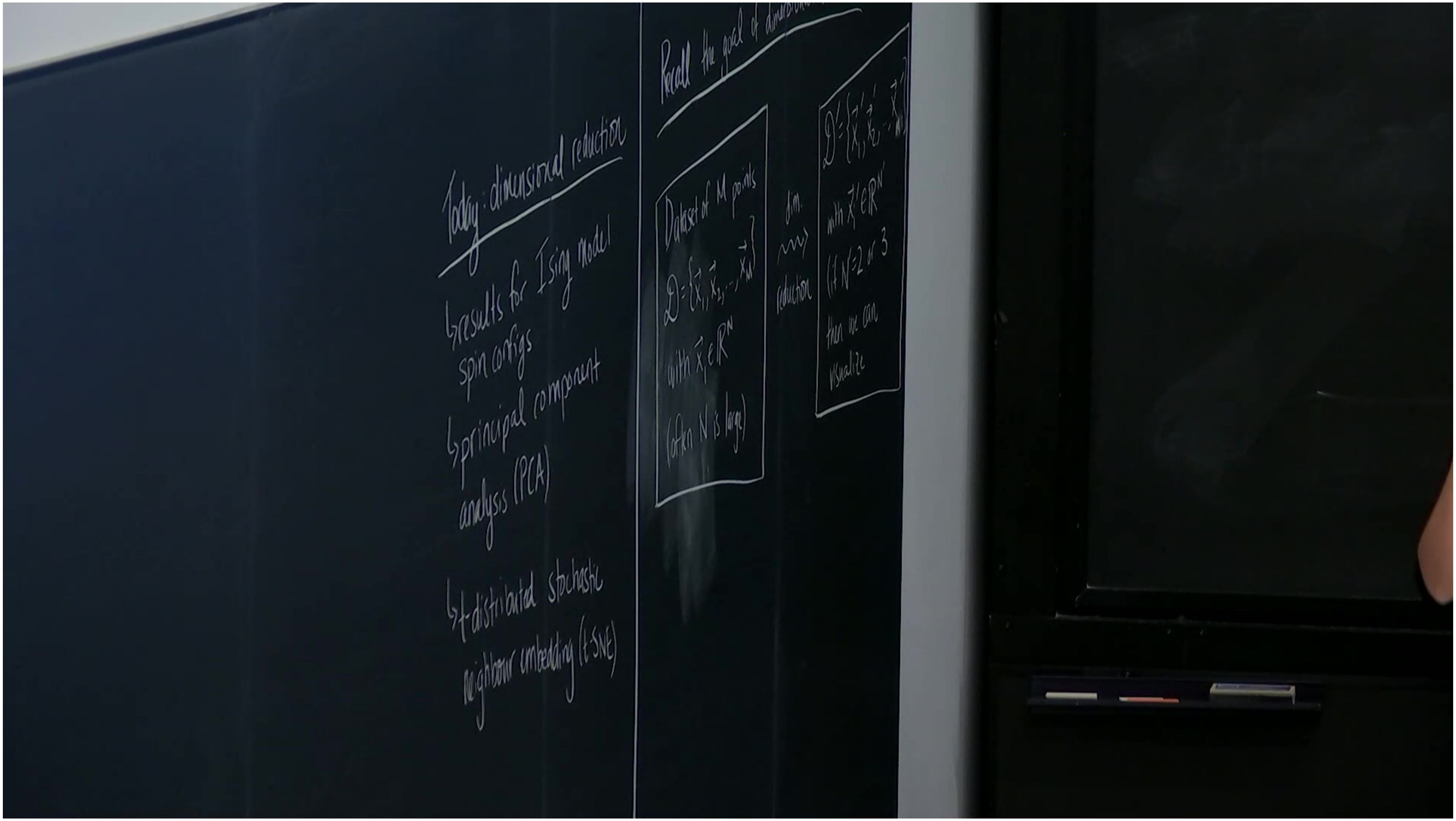
Speakers: Lauren Hayward

Collection: Machine Learning (2021/2022)

Date: April 25, 2022 - 11:30 AM

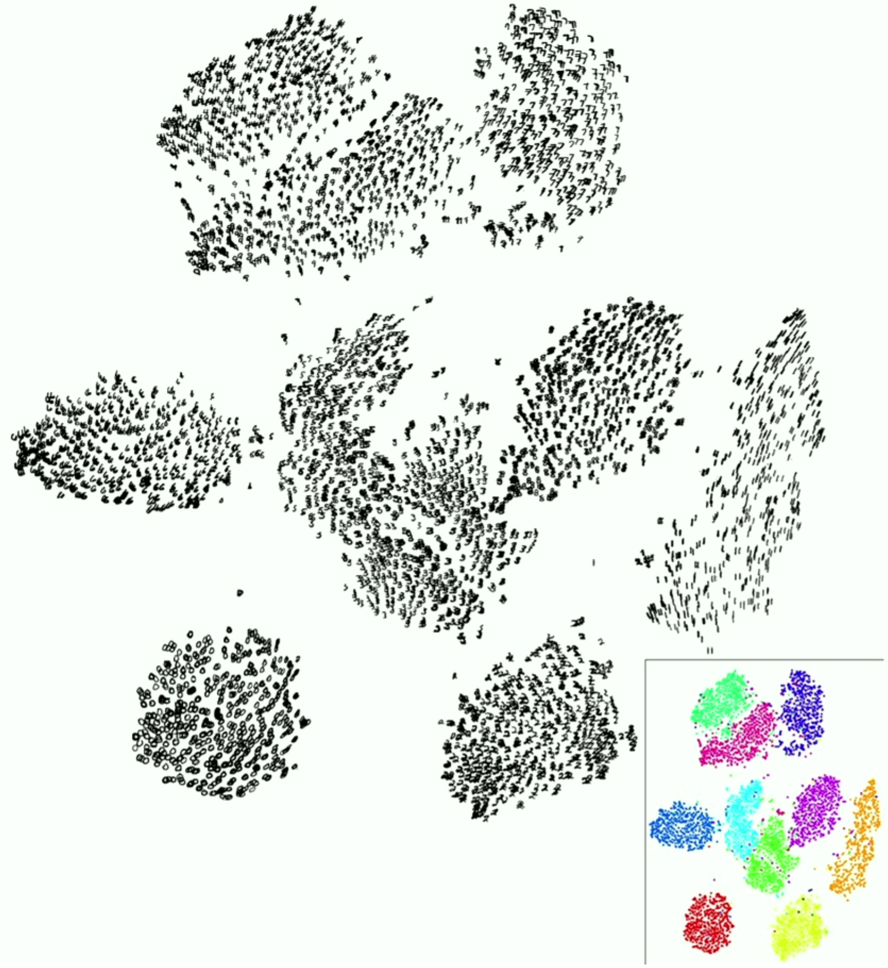
URL: https://pirsa.org/22040073

Abstract: This course is designed to introduce modern machine learning techniques for studying classical and quantum many-body problems encountered in condensed matter, quantum information, and related fields of physics. Lectures will focus on introducing machine learning algorithms and discussing how they can be applied to solve problem in statistical physics. Tutorials and homework assignments will concentrate on developing programming skills to study the problems presented in lecture.


Today: dimensional reduction
↳ results for Ising model spin configs
↳ principal component analysis (PCA)
↳ t-distributed stochastic neighbour embedding (t-SNE)
Recall the goal of
Dataset of M points
$\mathcal{D} = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M\}$
with $\vec{x}_i \in \mathbb{R}^N$
(often N is large)
dim. reduction
$\mathcal{D}' = \{\vec{x}_1', \vec{x}_2', \dots, \vec{x}_M'\}$
with $\vec{x}_i' \in \mathbb{R}^{N'}$
(if N'=2 or 3 then we can visualize

Dimensional reduction in 2D

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



Dimensional reduction in 2D

Mehta et al, arXiv:1803.08823

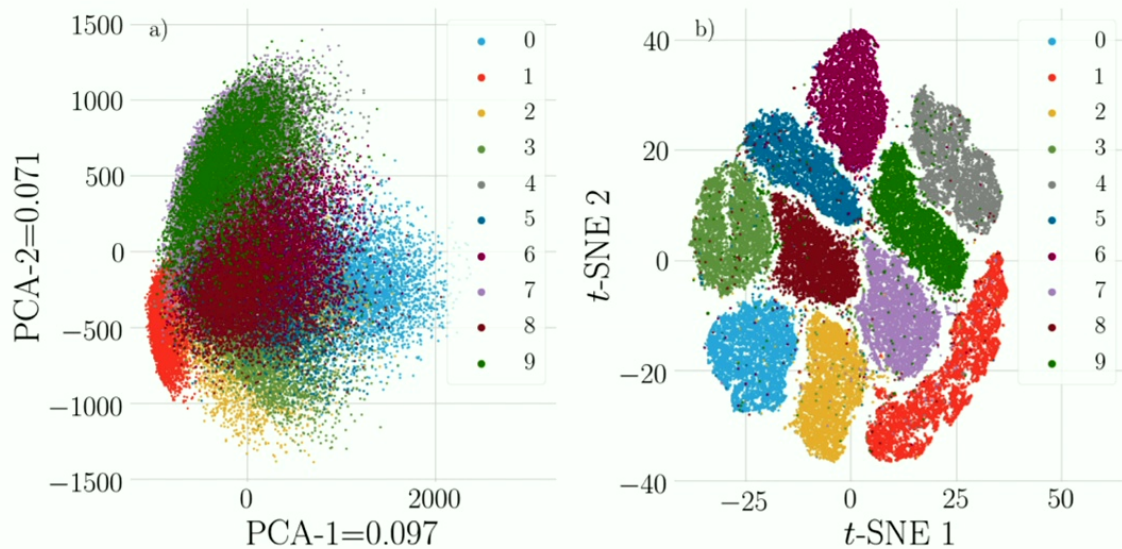
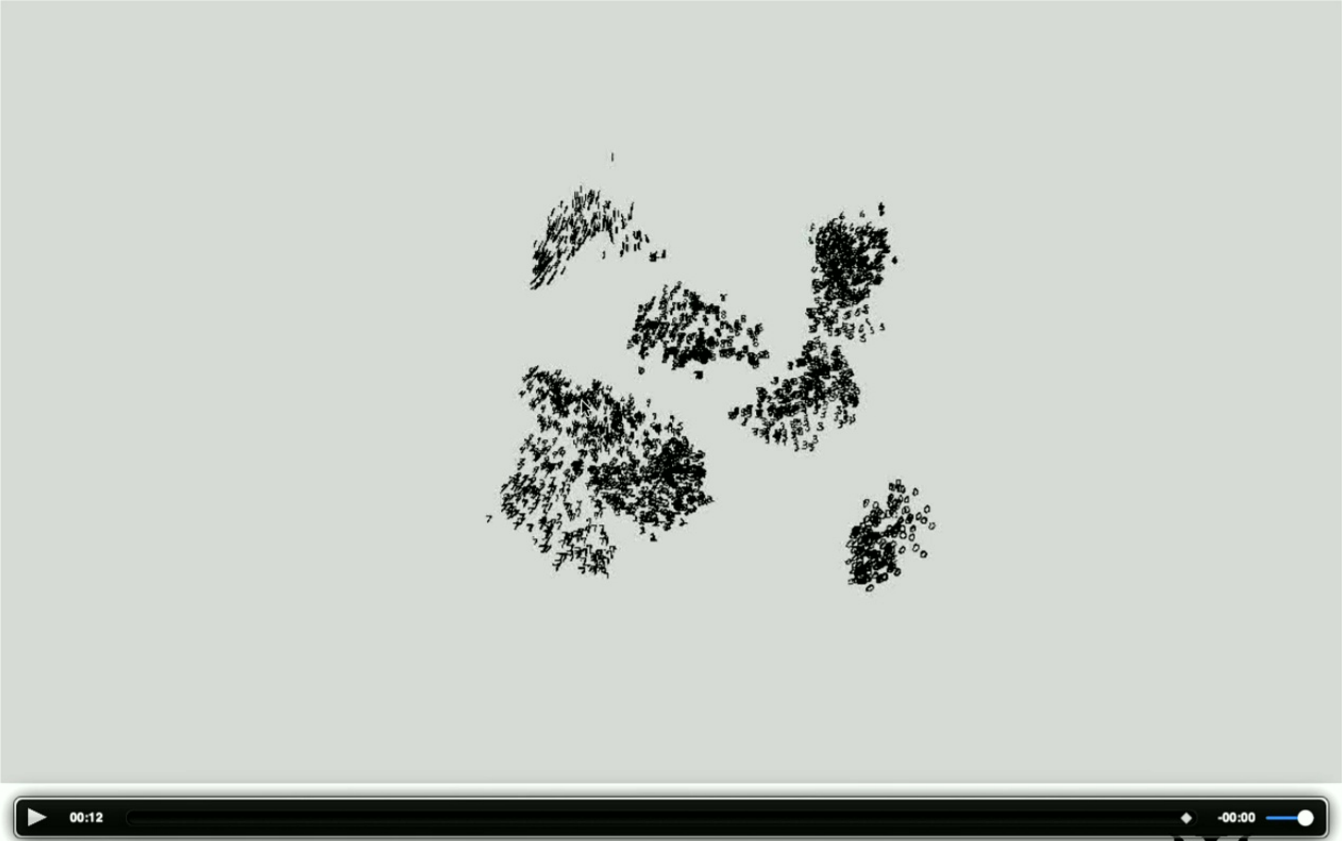


FIG. 54 Visualization of the MNIST handwritten digits training dataset (here $N = 60000$). (a) First two principal components. (b) t-SNE

PERIMETER SCHOLARS INTERNATIONAL

Dimensional reduction in 3D

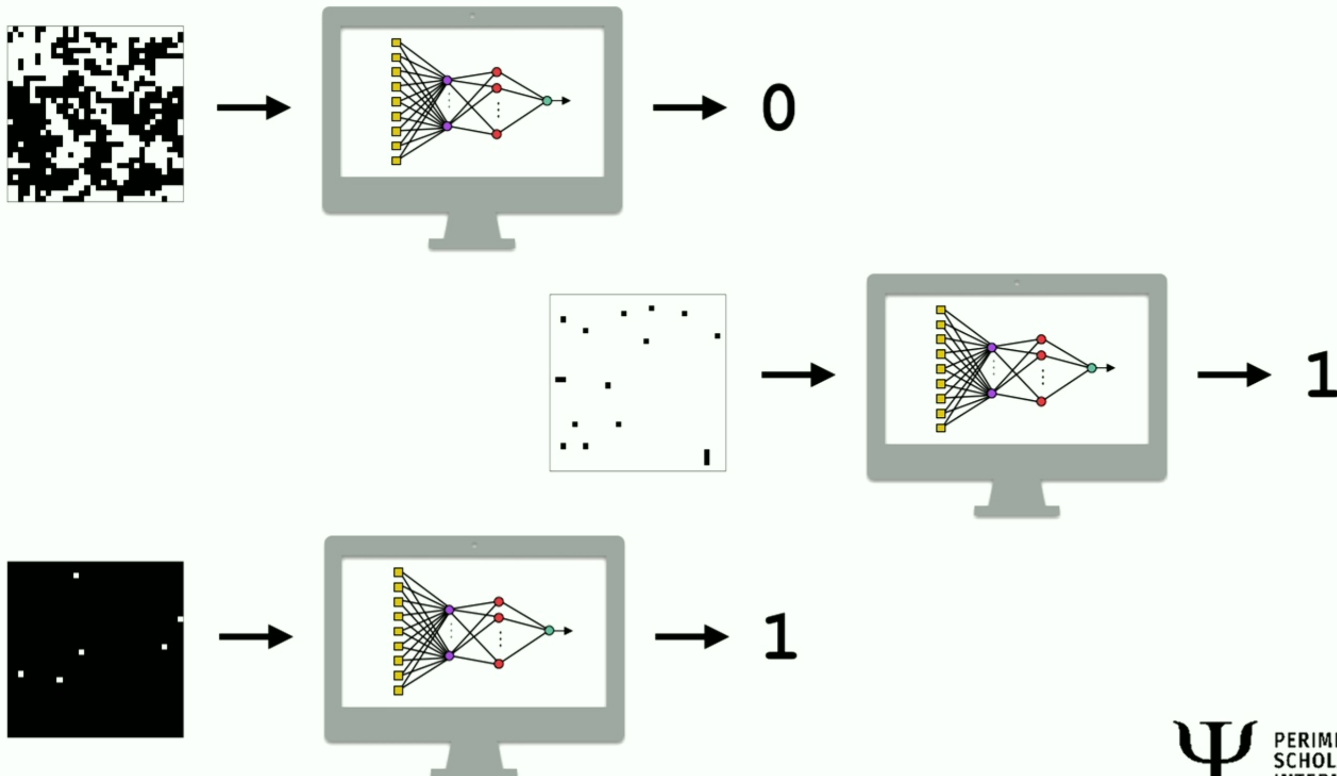
van der Maaten, https://lvdmaaten.github.io/tsne



PERIMETER
SCHOLARS
INTERNATIONAL

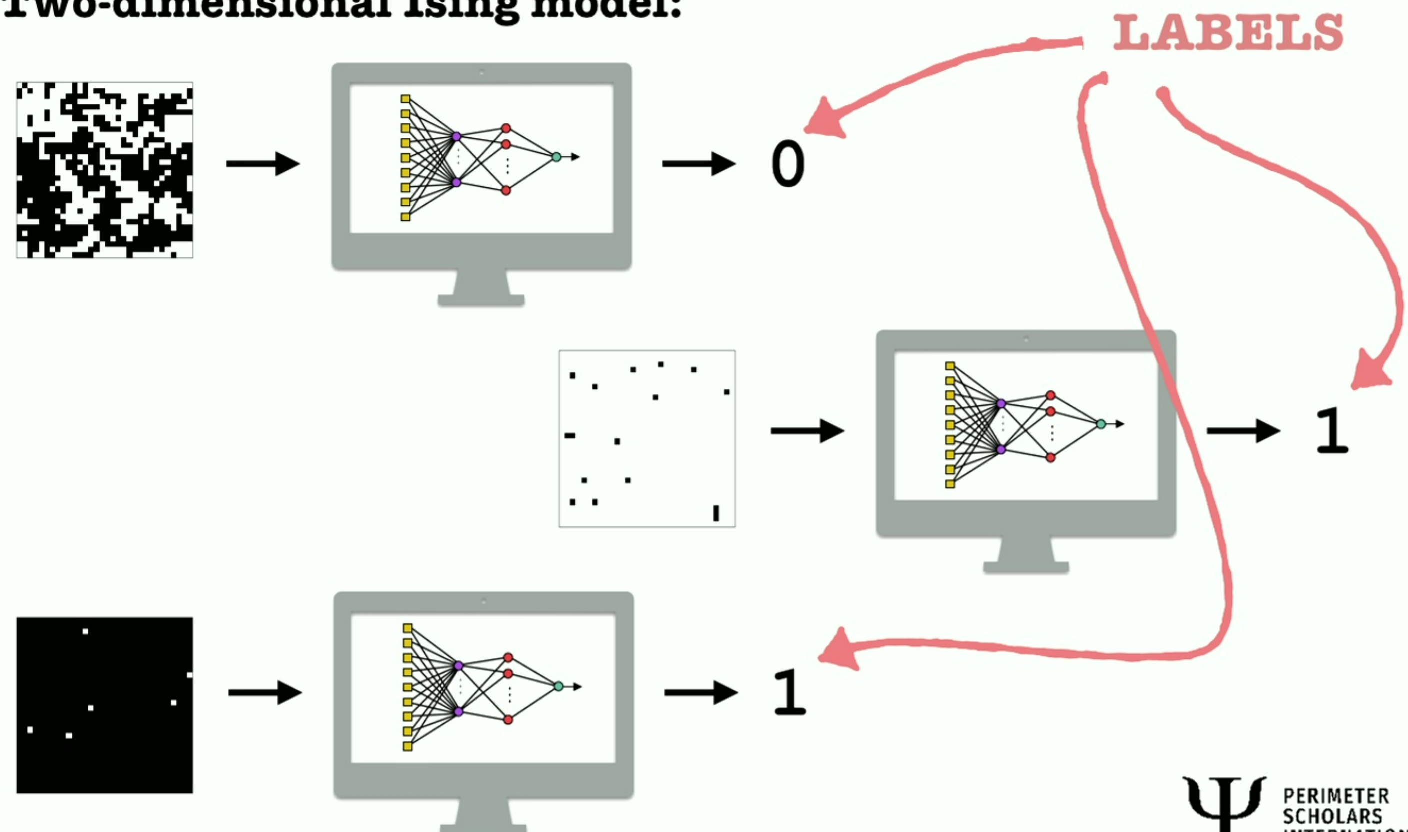
Classifying phases of matter with supervised learning

Two-dimensional Ising model:



Classifying phases of matter with supervised learning

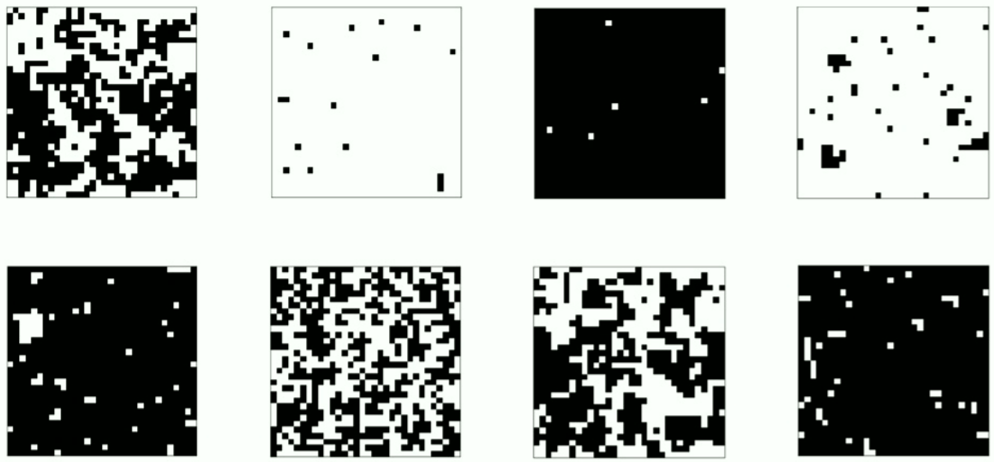
Two-dimensional Ising model:



PERIMETER SCHOLARS INTERNATIONAL

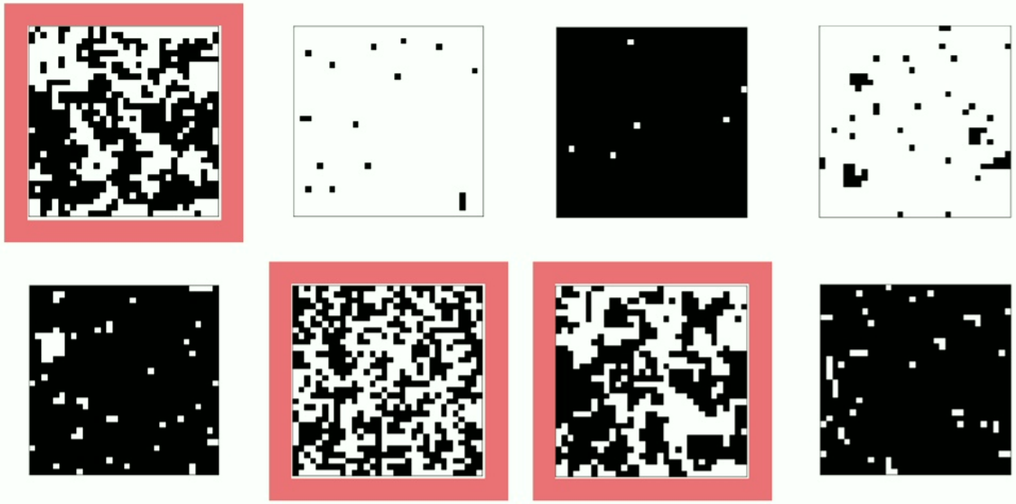
Identifying phases of matter without labels

Two-dimensional Ising model:



Identifying phases of matter without labels

Two-dimensional Ising model:

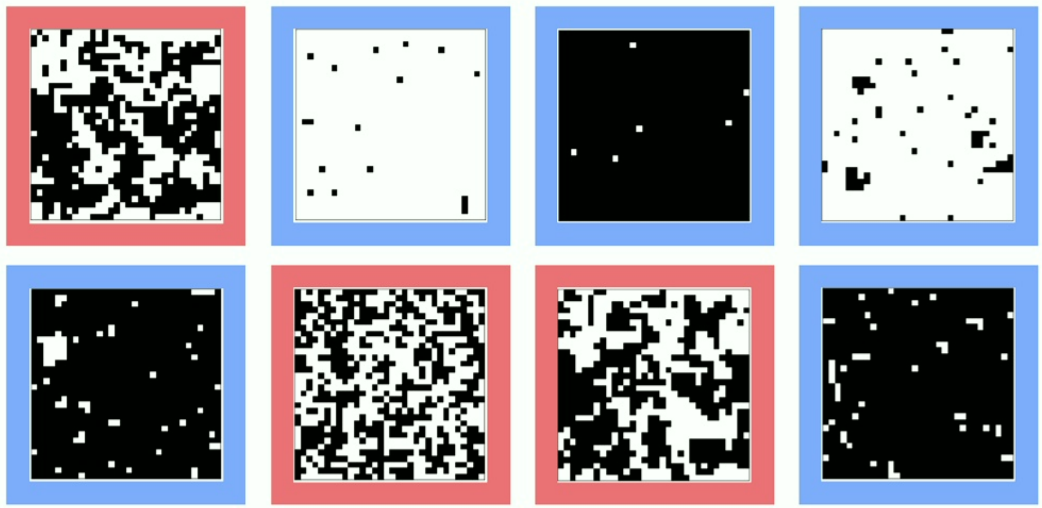


Group #1

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter without labels

Two-dimensional Ising model:

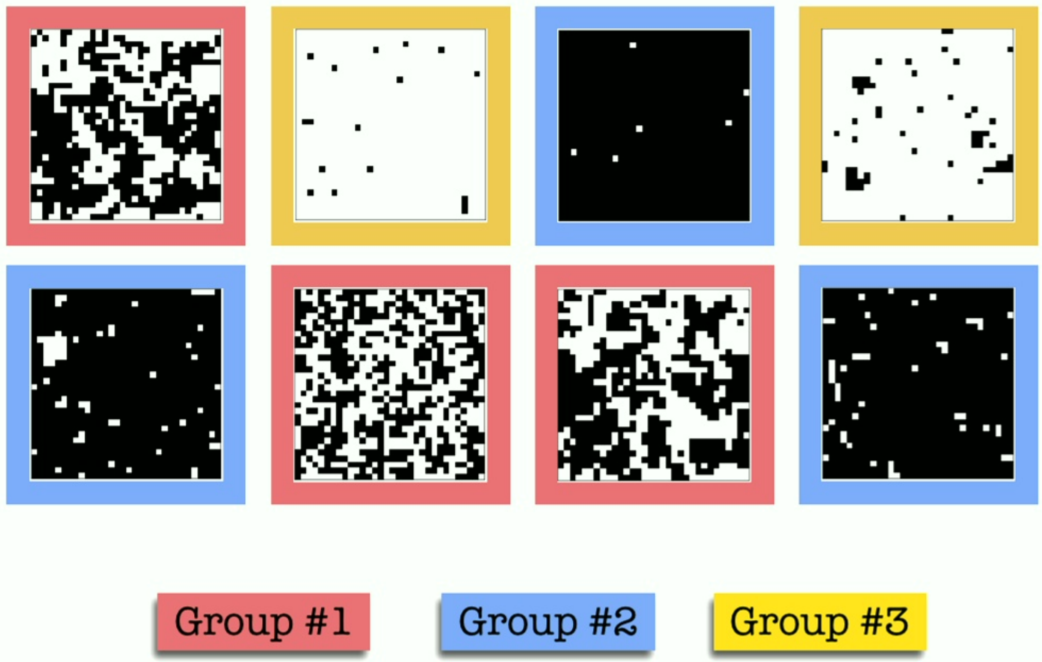


Group #1 Group #2

PERIMETER
SCHOLARS
INTERNATIONAL

Identifying phases of matter without labels

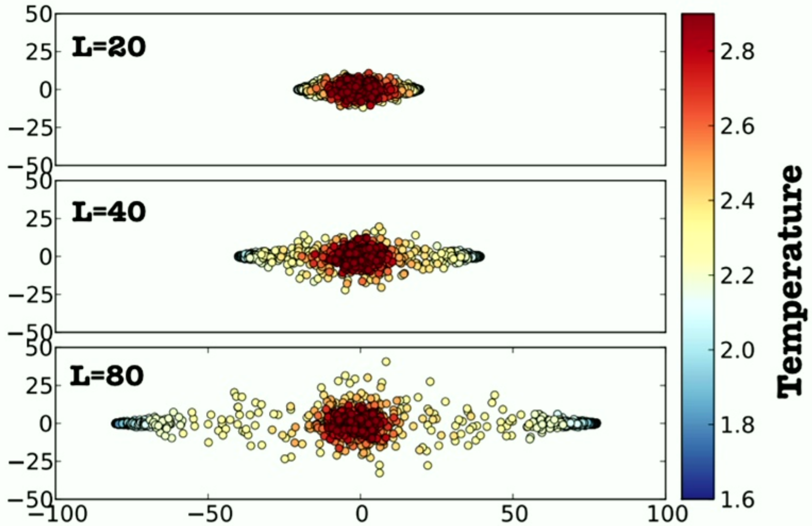
Two-dimensional Ising model:



PERIMETER
SCHOLARS
INTERNATIONAL

Identifying phases of matter with PCA

Two-dimensional Ising model:

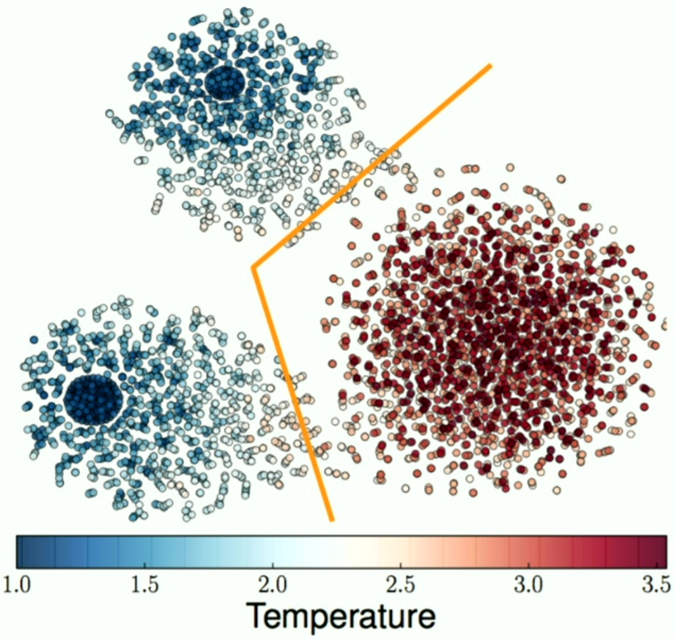


L. Wang, Phys. Rev. B **94**, 195105 (2016)

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter with t-SNE

Two-dimensional Ising model:

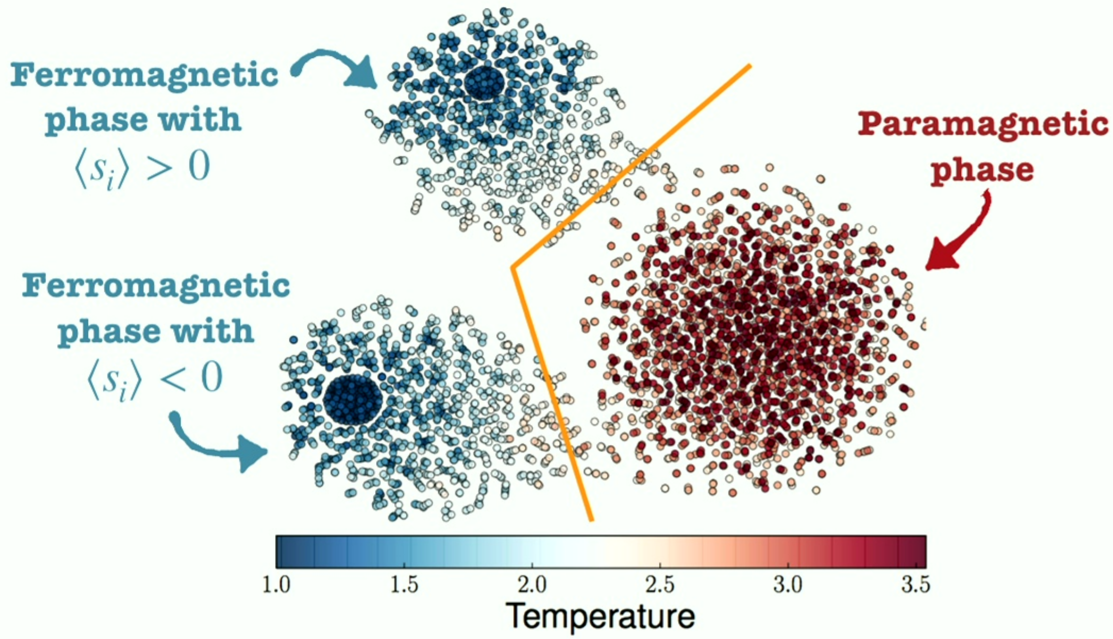


Carrasquilla and Melko, Nature Physics **13**, 431 (2017)

PERIMETER SCHOLARS INTERNATIONAL

Identifying phases of matter with t-SNE

Two-dimensional Ising model:

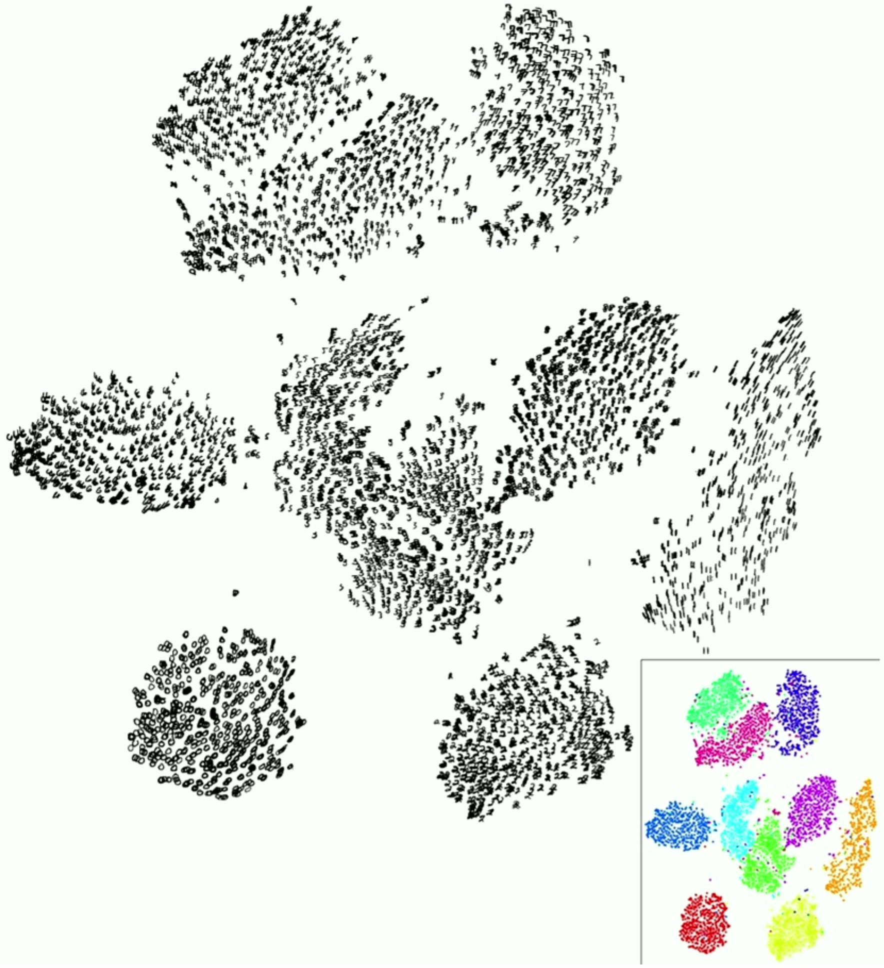


Carrasquilla and Melko, Nature Physics **13**, 431 (2017)

PERIMETER SCHOLARS INTERNATIONAL

Dimensional reduction in 2D

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



PERIMETER SCHOLARS INTERNATIONAL

Dimensional reduction using PCA

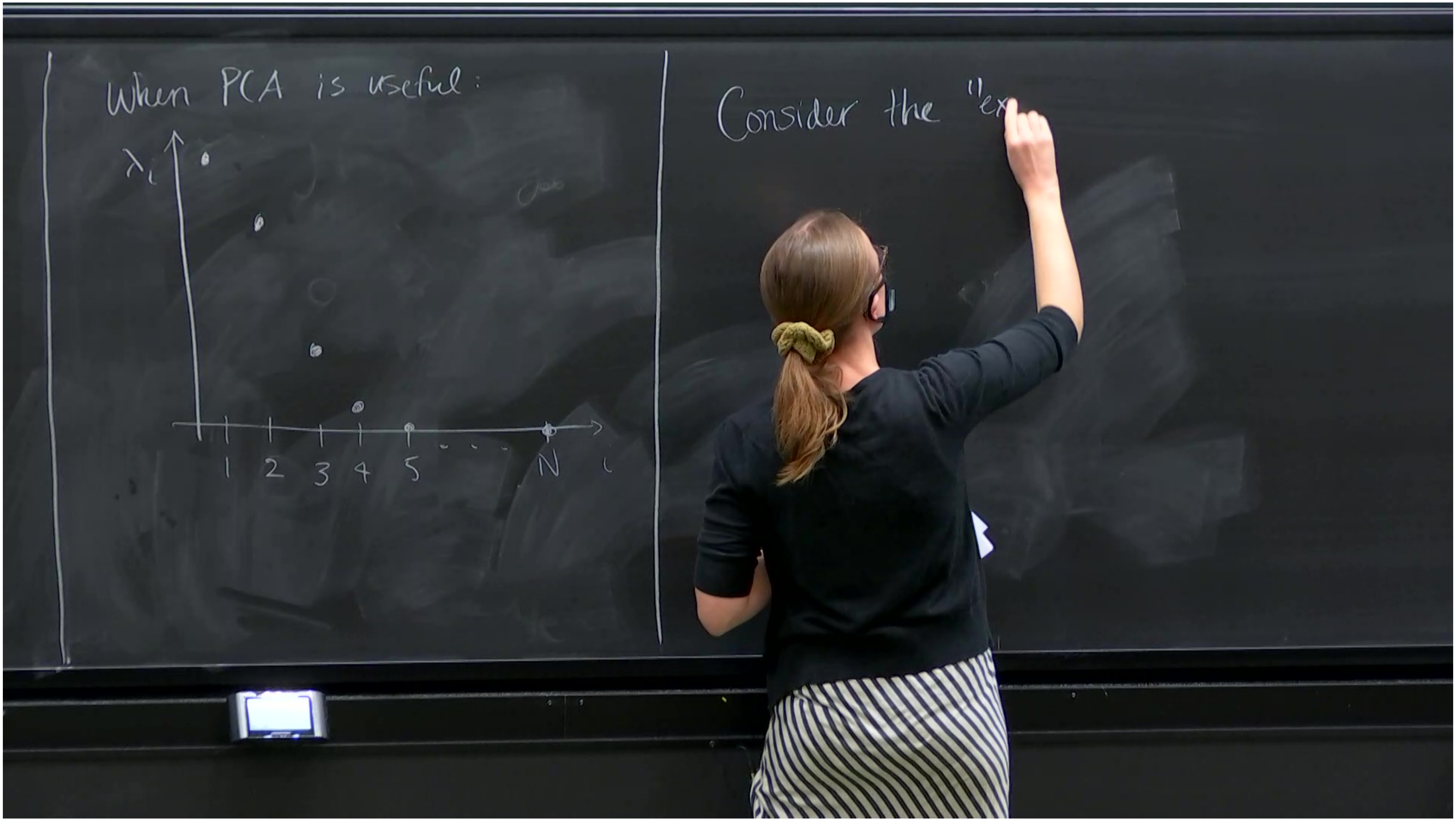
Recall that PCA amounts to solving the eigenvalue problem

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

where:

- $V_x = \frac{1}{M-1} (X^c)^T X^c$
- X^c stores the "centred" dataset
- P is $N \times N$, stores the "principal components"
- $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$
- $\lambda_j = \frac{1}{M-1} \sum_{i=1}^{M} X'^{\,2}_{ij} \geq 0$

When PCA


When PCA is useful:
λ_i
1 2 3 4 5 . . . . N i
Consider the "ex

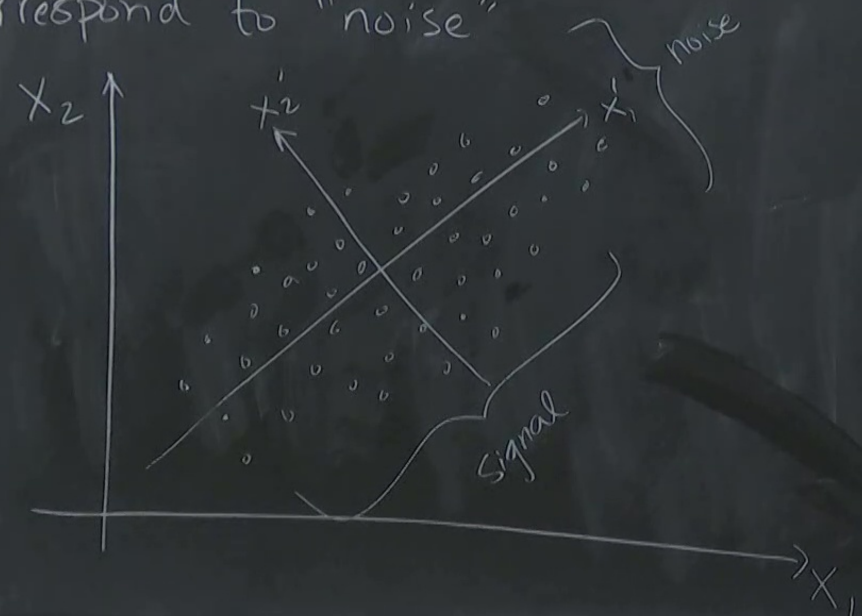
s useful:

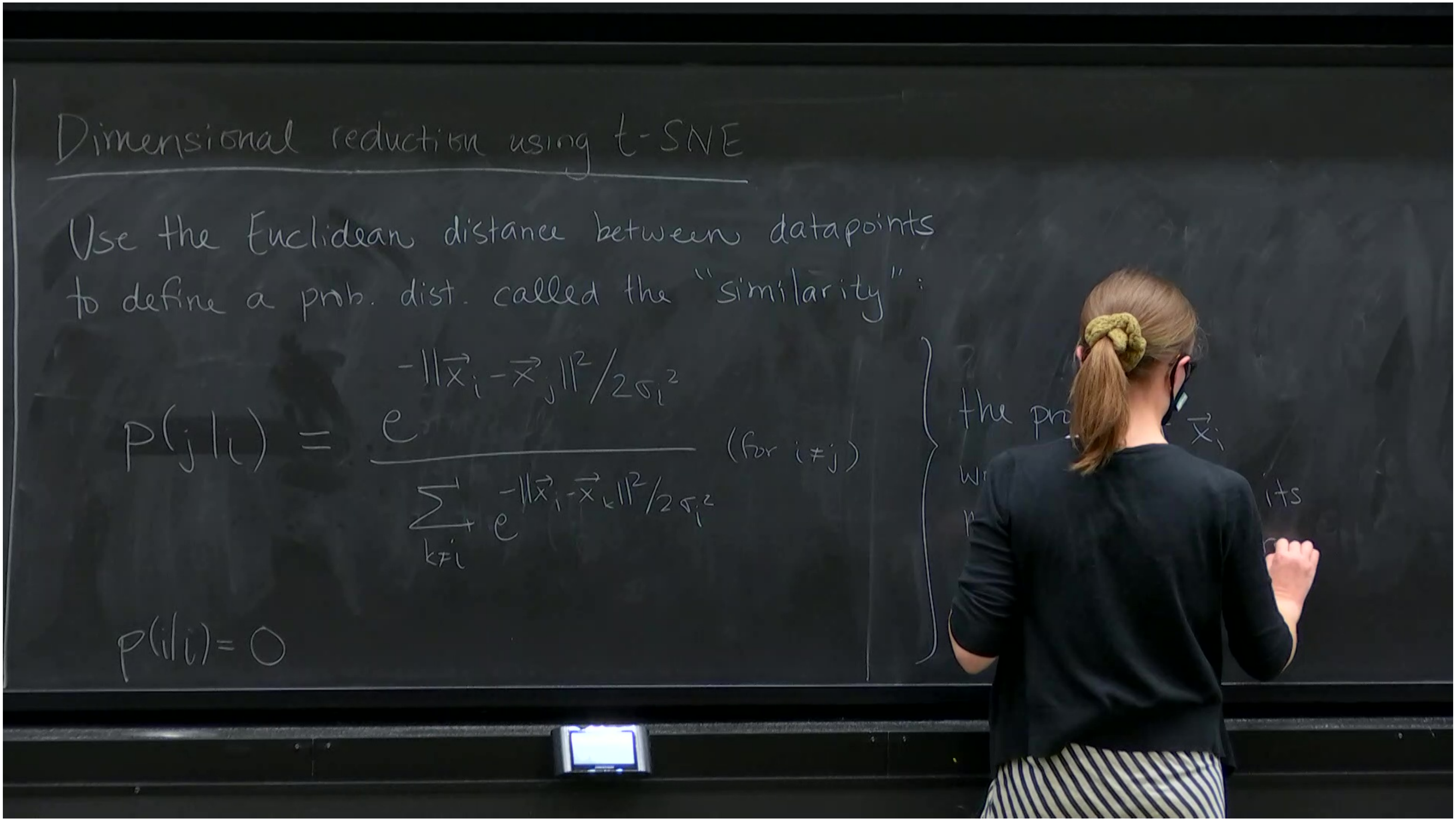
Consider the "explained variance ratio" r_i:

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

measures how much of the variance in the dataset can be explained by the first principal comp. x_i'

In PCA, directions with high variance correspond to "signal", while others correspond to "noise"





Dimensional reduction using t-SNE
Use the Euclidean distance between datapoints
to define a prob. dist. called the "similarity":
$p(j|i) = \dfrac{e^{-\|\vec{x}_i - \vec{x}_j\|^2/2\sigma_i^2}}{\sum_{k\neq i} e^{-\|\vec{x}_i - \vec{x}_k\|^2/2\sigma_i^2}}$ (for $i \neq j$)
$p(i|i) = 0$
the
its

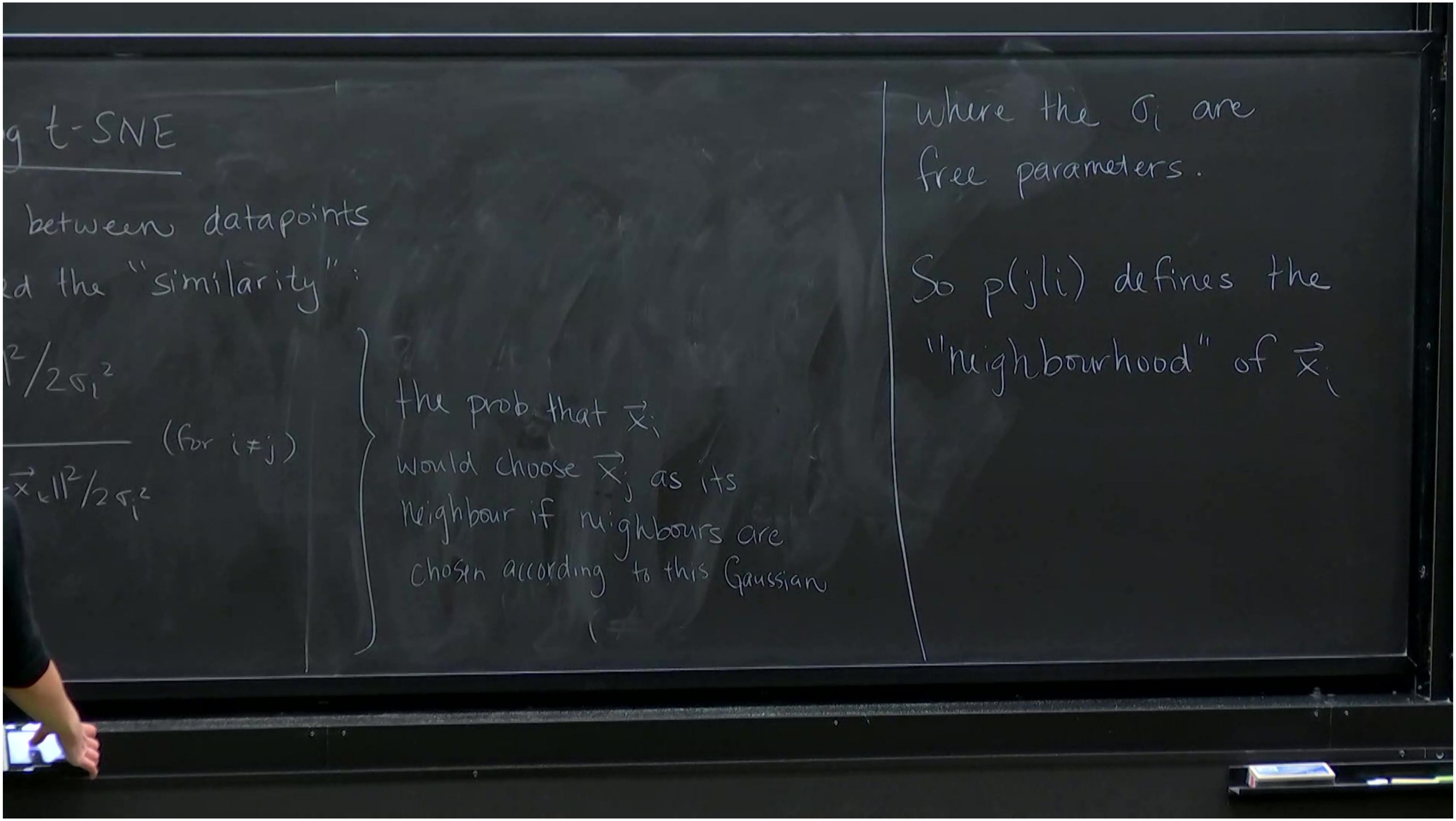
Dimensional reduction using t-SNE

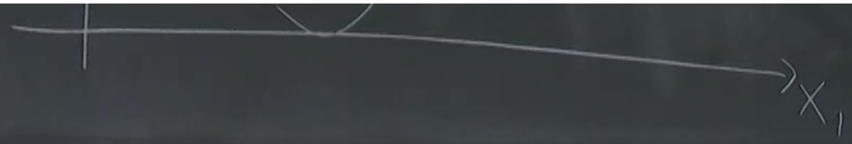
Use the Euclidean distance between datapoints to define a prob. dist. called the "similarity":

$$p(j|i) = \frac{e^{-\|\vec{x}_i - \vec{x}_j\|^2 / 2\sigma_i^2}}{\sum_{k \neq i} e^{-\|\vec{x}_i - \vec{x}_k\|^2 / 2\sigma_i^2}} \quad (\text{for } i \neq j)$$

the prob. that $\vec{x}_i$ would choose $\vec{x}_j$ as its neighbour if neighbours are chosen according to this Gaussian

$$p(i|i) = 0$$


g t-SNE
between datapoints
ed the "similarity":
$\|^2/2\sigma_i^2$
(for $i \neq j$)
$\vec{x}_k\|^2/2\sigma_i^2$
the prob. that $\vec{x}_i$ would choose $\vec{x}_j$ as its neighbour if neighbours are chosen according to this Gaussian
where the σ_i are free parameters.
So $p(j|i)$ defines the "neighbourhood" of $\vec{x}_i$



$$p(i|i) = 0$$

Define the corresponding symmetric dist

$$p(i,j) \equiv \frac{p(j|i) + p(i|j)}{2N}$$

Goal of t-SNE: find a lower-dim representation $\mathcal{D}' = \{\vec{x}\,'\}$ with $\vec{x}\,' \in \mathbb{R}^{N'}$ that preserves the similarity dist.

$p(i|i) = 0$

dist

$\mathcal{D}' = \{\vec{x}'\}$

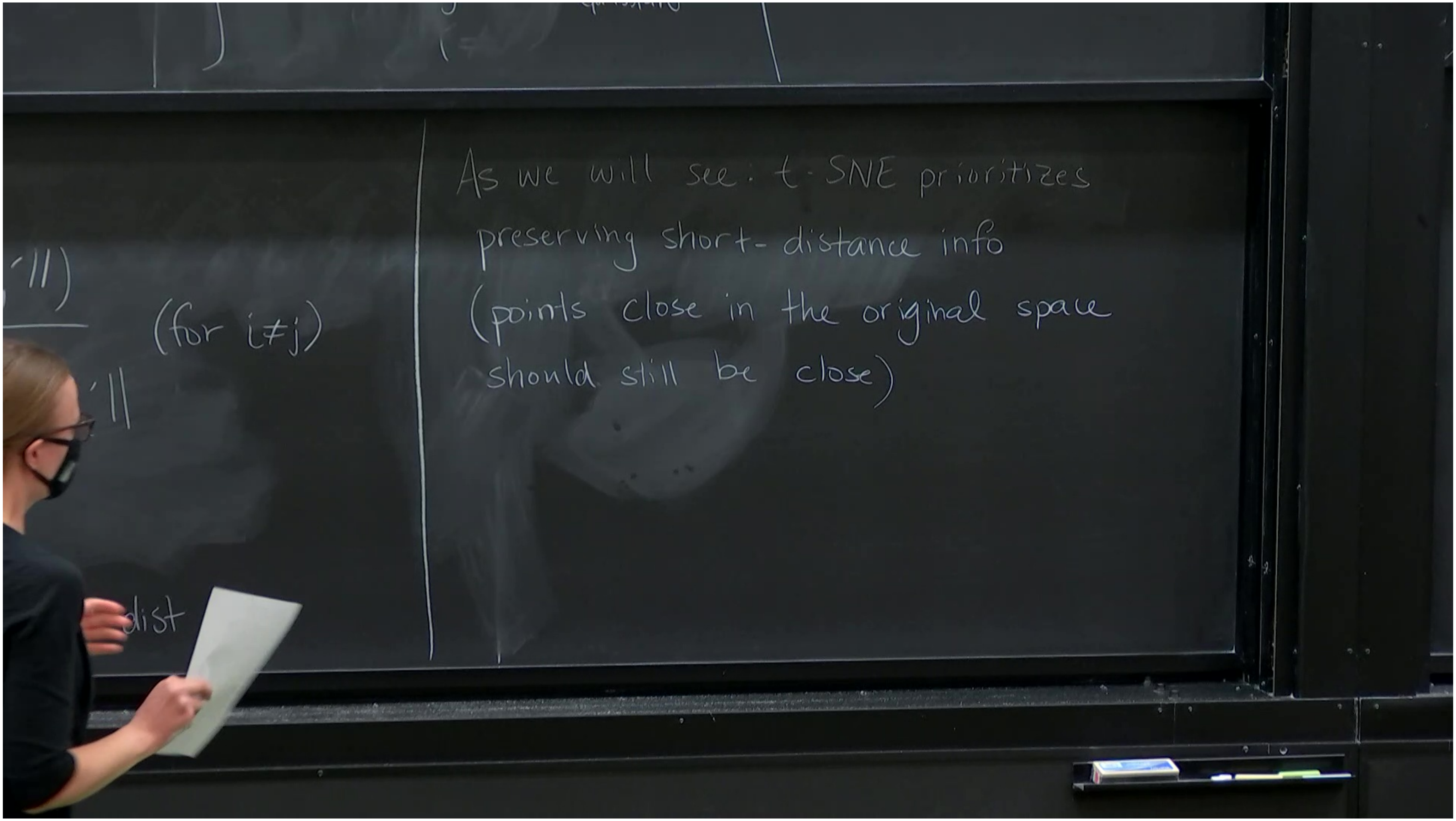
dist.

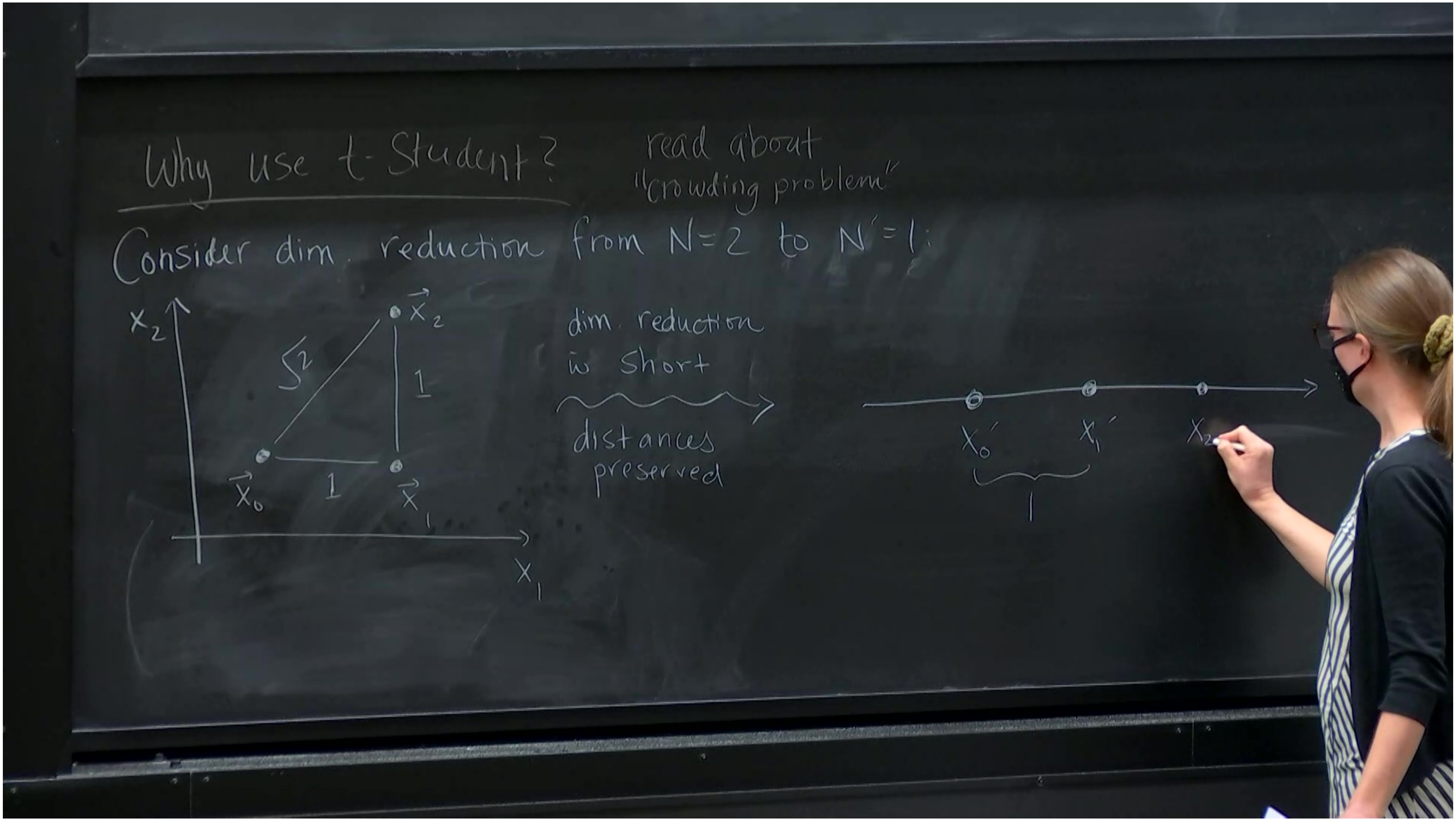
For the lower-dim rep:

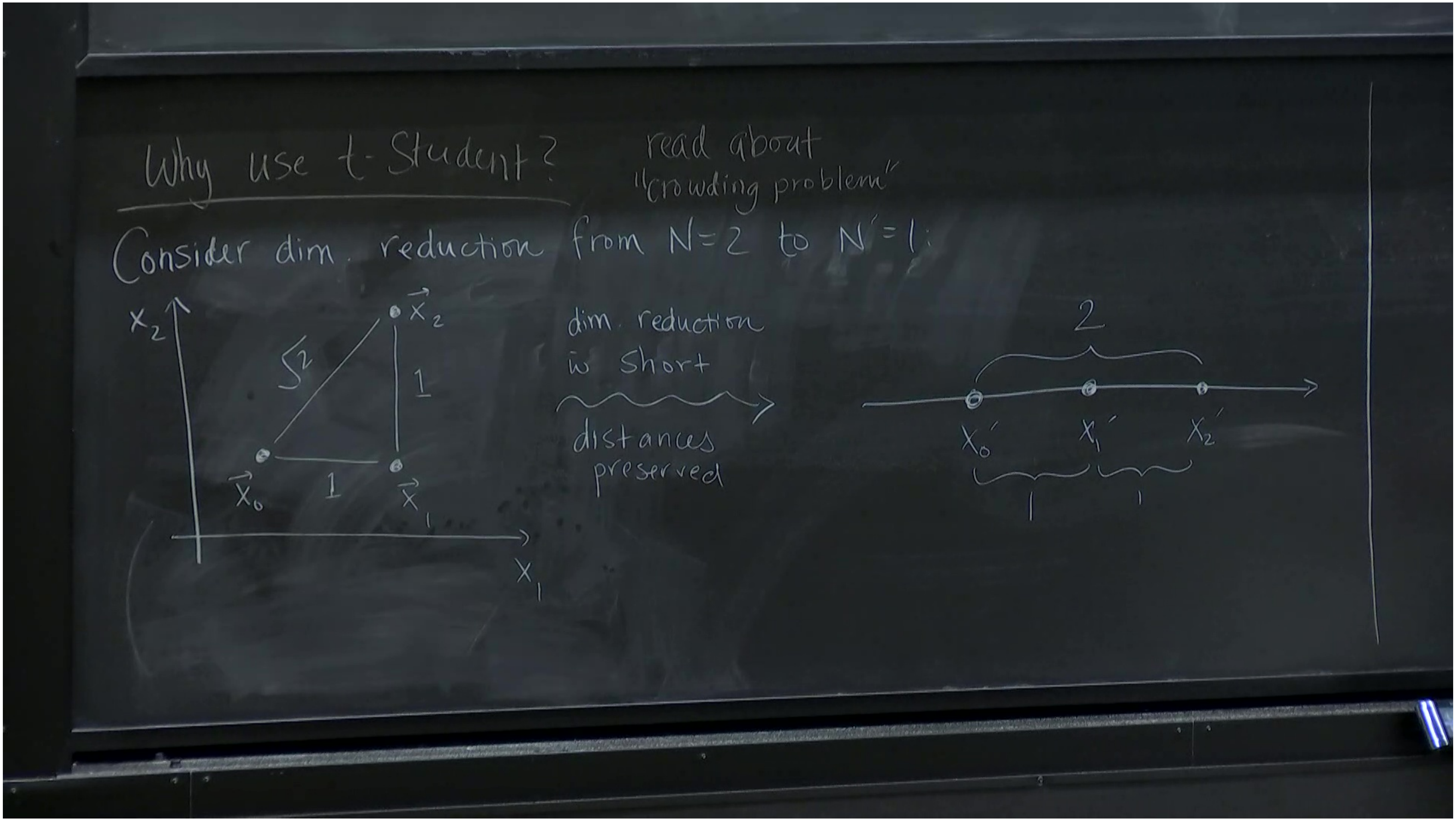
$$q(i,j) = \frac{f(\|\vec{x}_i' - x_j'\|)}{\sum_{k \neq l} f(\|\vec{x}_l' - \vec{x}_k'\|} \quad (\text{for } i \neq j)$$

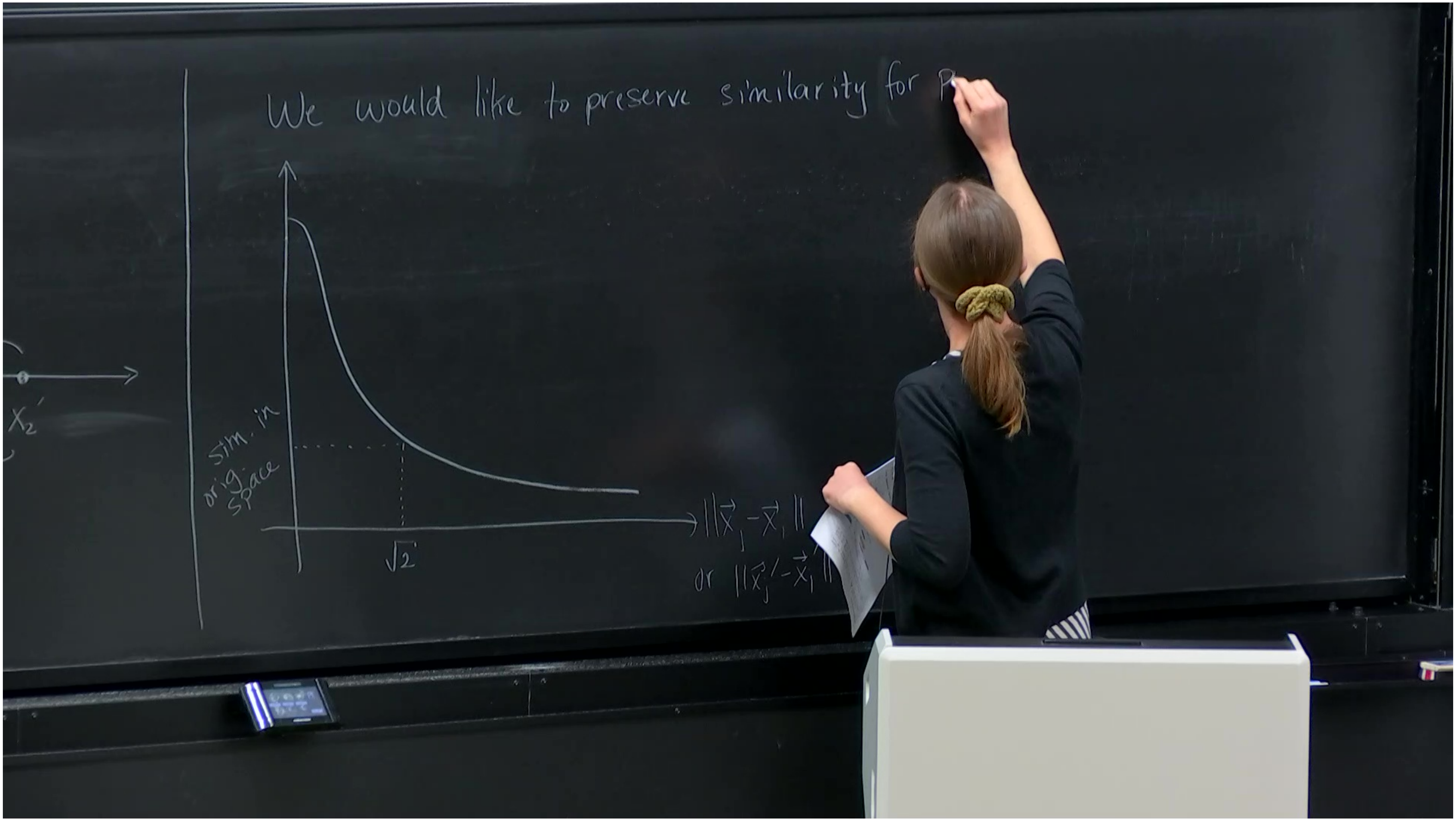
$$q(i,i) = 0$$

where $f(z) = \frac{1}{1+z^2}$ is the t-Student dist

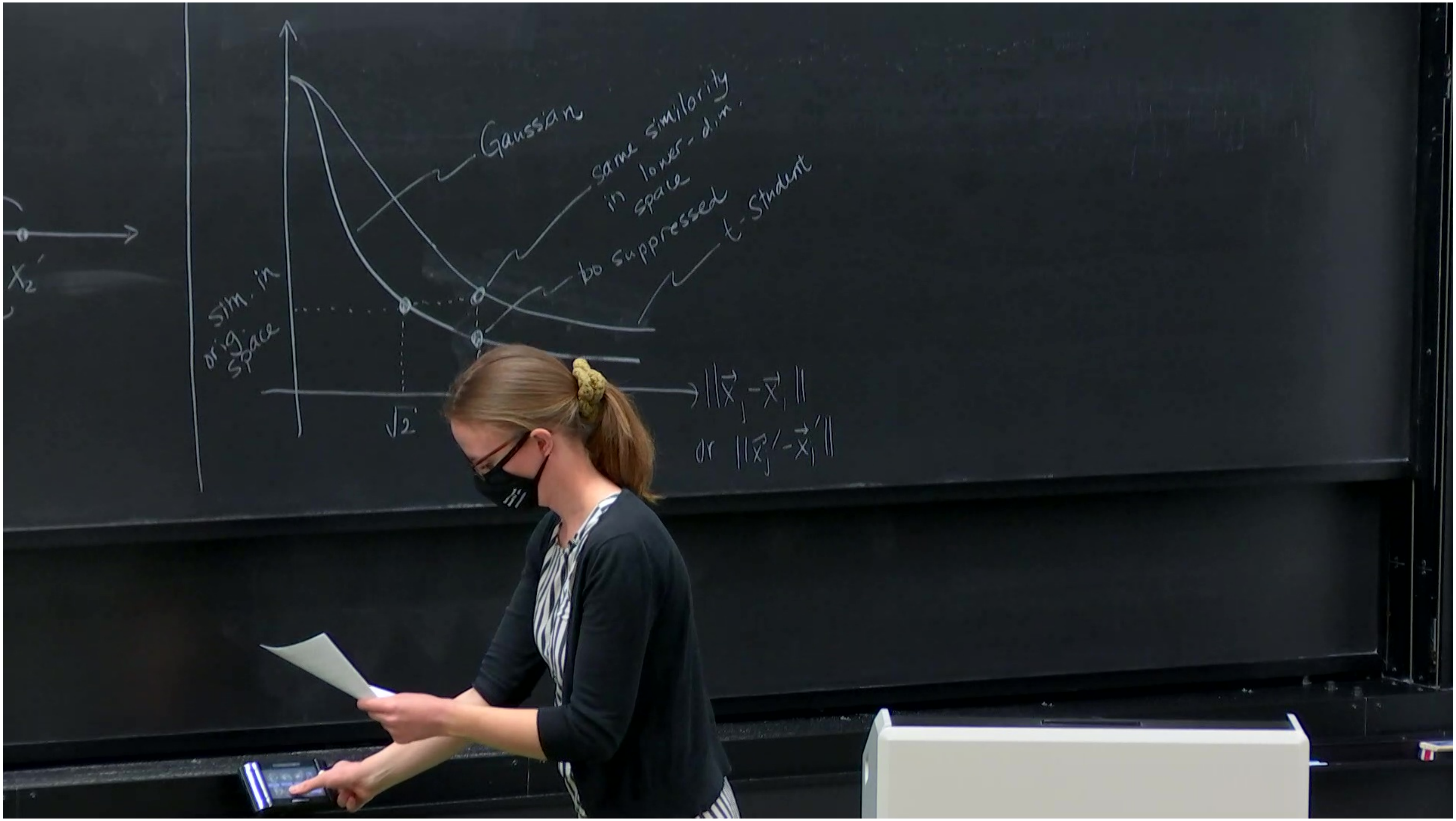

As we will see: t-SNE prioritizes
preserving short-distance info
(points close in the original space
should still be close)
(for i≠j)
dist

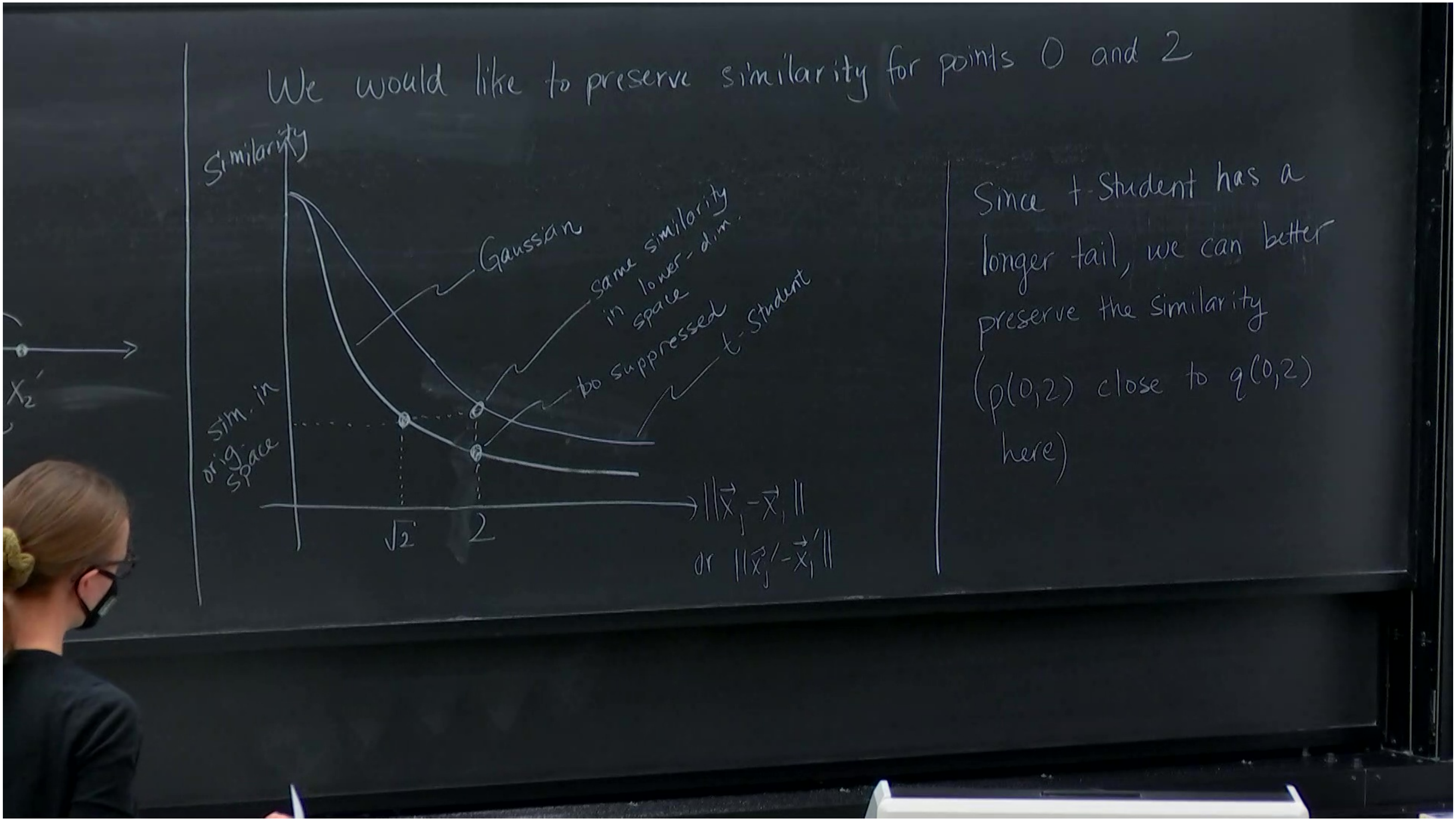

Why use t-Student?
read about
"crowding problem"
Consider dim. reduction from N=2 to N'=1:
x2
x1
X0
X1
X2
√2
1
1
dim. reduction
in short
distances
preserved
X0'
X1'
X2
1


Why use t-Student?
read about
"crowding problem"
Consider dim. reduction from N=2 to N'=1:
x_2
$\vec{x}_2$
$\sqrt{2}$
1
$\vec{x}_0$
1
$\vec{x}_1$
x_1
dim. reduction
w short
distances
preserved
2
x_0'
x_1'
x_2'
1
1



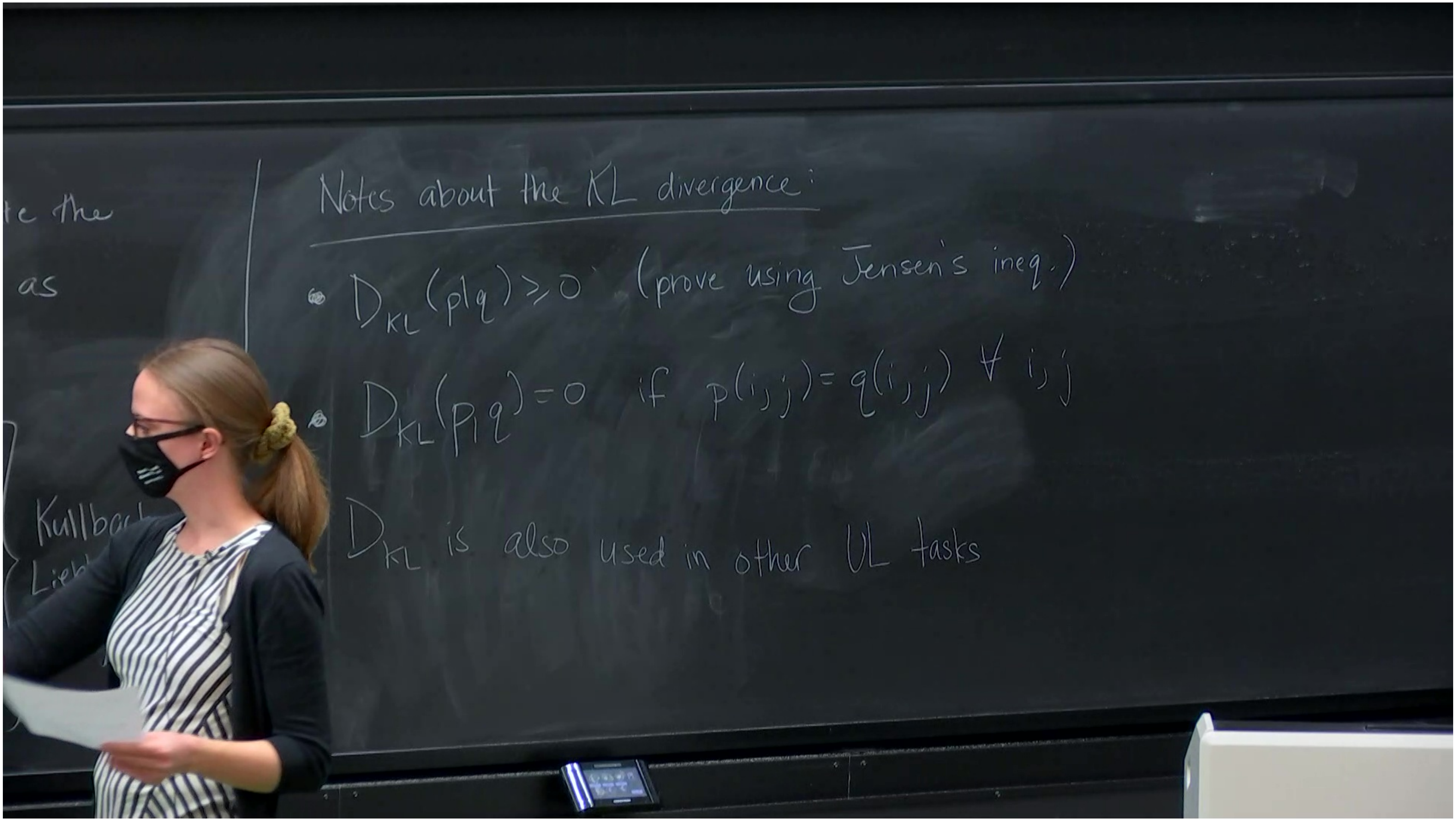
We would like to preserve similarity for P
sim. in orig. space
$\sqrt{2}$
$\|\vec{x}_j - \vec{x}_i\|$
or $\|\vec{x}_j' - \vec{x}_i'\|$


Gaussian
same similarity in lower-dim. space
to suppressed
t-Student
sim. in orig. space
$\sqrt{2}$
$\|\vec{x}_j - \vec{x}_i\|$
or $\|\vec{x}_j' - \vec{x}_i'\|$
x_2'


We would like to preserve similarity for points 0 and 2
Similarity
Gaussian
same similarity in lower-dim space
too suppressed
t-Student
sim. in orig. space
√2
2
||x_j - x_i||
or ||x_j' - x_i'||
Since t-Student has a longer tail, we can better preserve the similarity (p(0,2) close to q(0,2) here)

To find $\mathcal{D}' = \{\vec{x}'\}$, we iteratively update the points $\vec{x}_i'$ using an algorithm such as gradient descent, with cost:

$$\left.\begin{aligned} C(p,q) &= D_{KL}(p\,|\,q) \\ &= \sum_{ij} p(i,j)\left[\log \frac{p(i,j)}{q(i,j)}\right] \end{aligned}\right\} \text{Kullback-Liebler (KL) divergence}$$


te the
as
Notes about the KL divergence:
$D_{KL}(p|q) \geq 0$ (prove using Jensen's ineq.)
$D_{KL}(p|q) = 0$ if $p(i,j) = q(i,j)$ $\forall$ i,j
D_{KL} is also used in other UL tasks
Kullbac
Lieb

Additional notes about t-SNE:

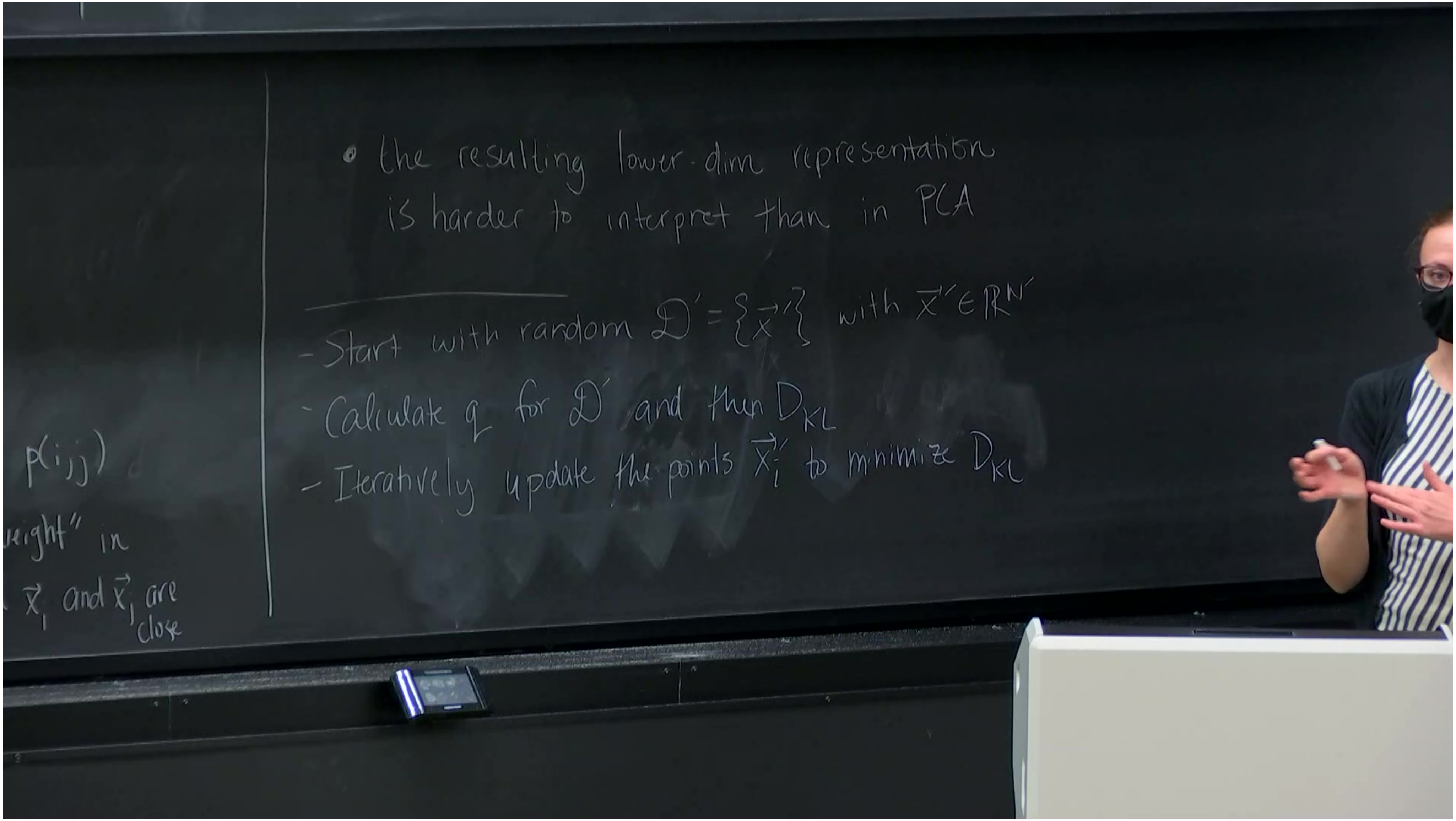
- $p(i,j)$ is fixed and given by the dataset $\mathcal{D}$.
  We adjust the points $\vec{x}_i{}' = (x_{i1}', x_{i2}', \ldots, x_{iN}')$
  such that $q(i,j)$ is as close as possible to $p(i,j)$.

- t-SNE preserves

Additional notes about t-SNE:

- $p(i,j)$ is fixed and given by the dataset $\mathcal{D}$.
  We adjust the points $\vec{x}_i{}' = (x_{i1}', x_{i2}', \dots, x_{iN}')$
  such that $q(i,j)$ is as close as possible to $p(i,j)$.

- t-SNE preserves local structure since $p(i,j)$ acts as a "weight" in the sum $\sum_{i,j}$ to calculate D_{KL}, and $p(i,j)$ is larger when the original $\vec{x}_i$ and $\vec{x}_j$ are close



p(i,j)
weight" in
x_i and x_j are close
• the resulting lower-dim representation is harder to interpret than in PCA
- Start with random $\mathcal{D}' = \{\vec{x}'\}$ with $\vec{x}' \in \mathbb{R}^{N'}$
- Calculate q for $\mathcal{D}'$ and then D_{KL}
- Iteratively update the points $\vec{x}_i'$ to minimize D_{KL}