Introduction to Machine Learning Methods in Condensed Matter Physics

LECTURE 10, FALL 2021

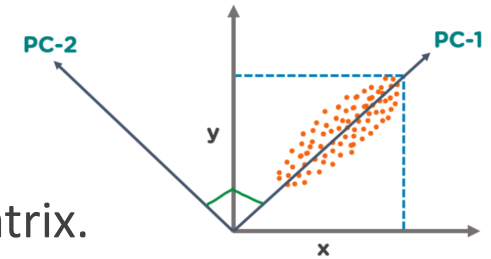
Yi Zhang (张亿)

International Center for Quantum Materials, School of Physics
Peking University, Beijing, 100871, China

Email: frankzhangyi@pku.edu.cn

Dimension reduction for unsupervised learning

- Last class: PCA via eigenvalue problem (orthogonal transformation):



$$C = \frac{X^T X}{n-1} = \frac{\Sigma_{i=1}^{n} x_i x_j^T}{n-1}$$

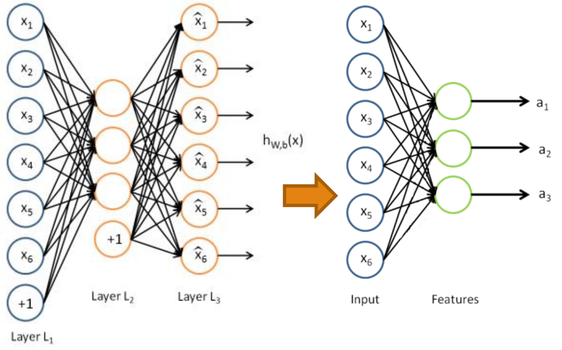
It is often inefficient to sum over the large dataset for the covariance matrix.

- PCA through singular value decomposition (SVD):

$$\mathbf{M} = \mathbf{U\Sigma V^T}$$

where both $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal and $\boldsymbol{\Sigma}$ is diagonal.

$$\lambda_i = \sigma_i^2/(n\text{-}1)$$



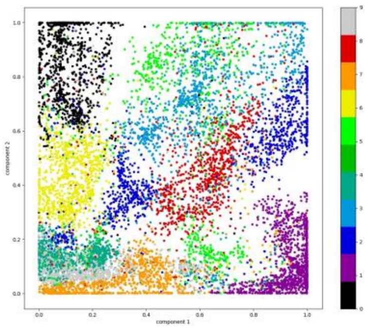
$$C = \frac{X^T X}{n-1} = \frac{V\Sigma^T U^T U\Sigma V^T}{n-1} = V\frac{\Sigma^2}{n-1}V^T$$

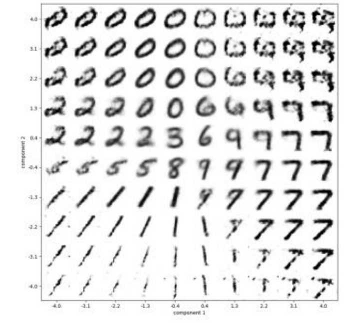
- For nonlinear dimension reduction, we have the autoencoders:

$$a_i^{(2)} = f\left(\sum_{j=1}^{100} W_{ij}^{(1)} x_j + b_i^{(1)}\right)$$

each hidden neuron at the bottleneck composes a dimension.

MNIST data processed by autoencoders:





Self-taught learning

- "Sometimes it's not who has the best algorithm that wins; it's who has the most data."
- We may have a lot of unlabeled data, but only a smaller amount of (expensive) labeled data.

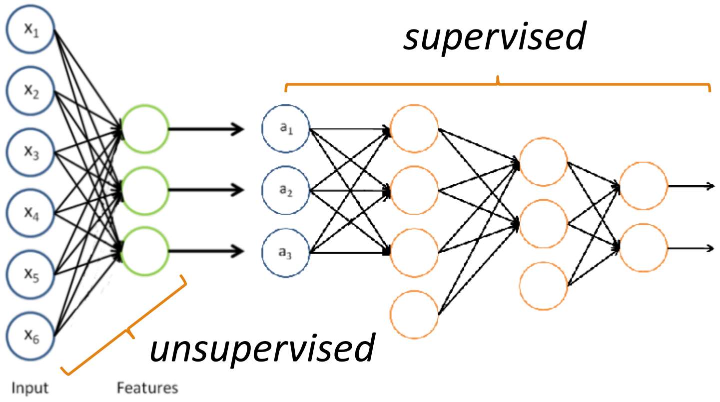
$$\{x_u^{(1)}, x_u^{(2)}, \ldots, x_u^{(m_u)}\} \qquad \{(x_l^{(1)}, y^{(1)}), (x_l^{(2)}, y^{(2)}), \ldots (x_l^{(m_l)}, y^{(m_l)})\}$$

- Combining unsupervised machine learning and supervised machine learning:
  - Step1. learn a good feature representation of the input from the (large set of) unlabeled data as a preprocessing
  - Step2. apply supervised learning on the this feature representation of the (small set of) labelled data for the classification task

$$\{(a_l^{(1)}, y^{(1)}), (a_l^{(2)}, y^{(2)}), \ldots (a_l^{(m_l)}, y^{(m_l)})\}$$ or

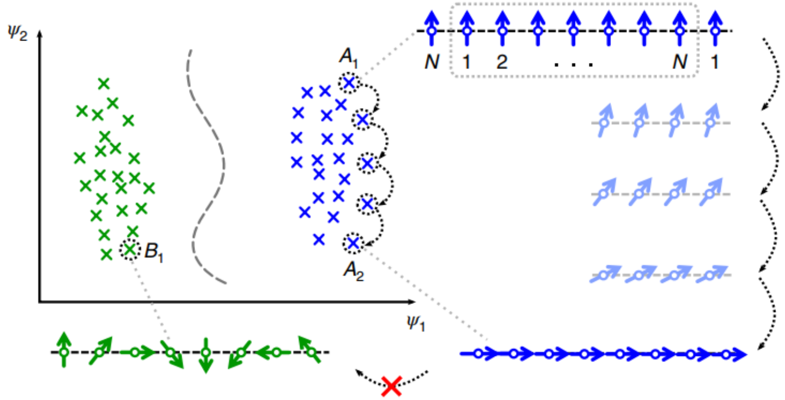
$$\{((x_l^{(1)}, a_l^{(1)}), y^{(1)}), ((x_l^{(2)}, a_l^{(1)}), y^{(2)}), \ldots, ((x_l^{(m_l)}, a_l^{(1)}), y^{(m_l)})\}$$

  - Step3. an unified ANN: test data should subject to the same preprocessing



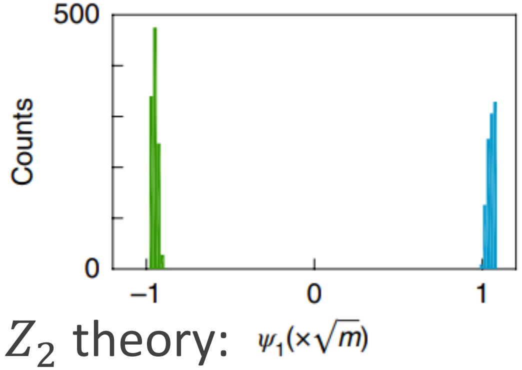
Unsupervised machine learning topological states

- Topological invariants: via dimension reduction and connectivity analysis:
  - Difference between A_1 and A_2 can be larger than B_1 and A_2
  - Global connectivity instead of local similarity is more important



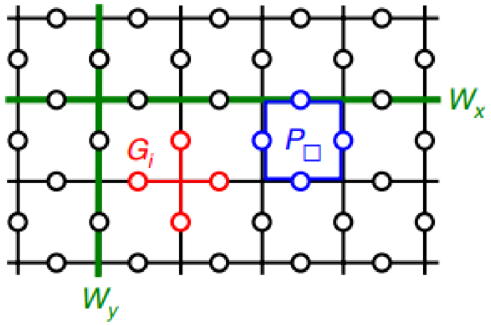
1D chain of N classical XY spins

- Indeed, a good separation without preexisting labels:

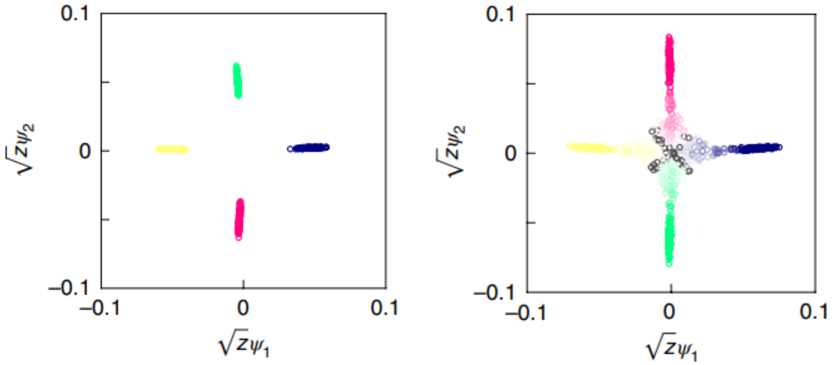


- Also applicable to 2D:

The four topological sectors of Z_2 theory:



$$E[\{\sigma_b\}] = -K\sum_{\square} P_\square, \quad P_\square \equiv \prod_{b\in\square} \sigma_b$$



$T/K = 0.05$ $T/K = 0.18$

Joaquin F. Rodriguez-Nieva, Mathias S. Scheurer, Nature Physics ***15****, 790–795 (2019).*

Sparse modeling

- Other unsupervised machine learning methods: t-SNE and K-means
- Unsupervised machine learning with increased dimensionality: sparse modeling, express data

$$\mathbf{x} = \sum_{i=1}^{k} a_i \phi_i$$

with an **over-complete** set of basis vectors ϕ_i $(k > n)$ and a_i is sparse (mostly zeros).

$$\Rightarrow \text{minimize}_{a_i^{(j)}, \phi_i} \sum_{j=1}^{m} \left|\left| \mathbf{x}^{(j)} - \sum_{i=1}^{k} a_i^{(j)} \phi_i \right|\right|^2 + \lambda \sum_{i=1}^{k} S(a_i^{(j)})$$

optimizable via convex optimization or gradient descent

Sparsity penalty functions: L_0: $S(a_i) = \mathbf{1}(|a_i| > 0)$ L_1: $S(a_i) = |a_i|$ log: $S(a_i) = \log(1 + a_i^2)$

- Probabilistic interpretation: minimize the KL divergence: $\phi^* = \text{argmax}_\phi E\left[\log(P(\mathbf{x} \mid \phi))\right]$

$$D(P^*(\mathbf{x}) || P(\mathbf{x} \mid \phi)) = \int P^*(\mathbf{x}) \log\left(\frac{P^*(\mathbf{x})}{P(\mathbf{x} \mid \phi)}\right) d\mathbf{x}$$

$$P(\mathbf{x} \mid \phi) = \int P(\mathbf{x} \mid \mathbf{a}, \phi) P(\mathbf{a}) d\mathbf{a}$$

$$E(\mathbf{x}, \mathbf{a} \mid \phi) = -\log(P(\mathbf{x} \mid \phi, \mathbf{a}) P(\mathbf{a}))$$

$$P(\mathbf{x} \mid \mathbf{a}, \phi) = \frac{1}{Z} \exp\left(-\frac{(\mathbf{x} - \sum_{i=1}^{k} a_i \phi_i)^2}{2\sigma^2}\right) \quad P(a_i) = \frac{1}{Z} \exp(-\beta S(a_i)) \Rightarrow$$

$$= \sum_{j=1}^{m} \left|\left| \mathbf{x}^{(j)} - \sum_{i=1}^{k} a_i^{(j)} \phi_i \right|\right|^2 + \lambda \sum_{i=1}^{k} S(a_i^{(j)})$$