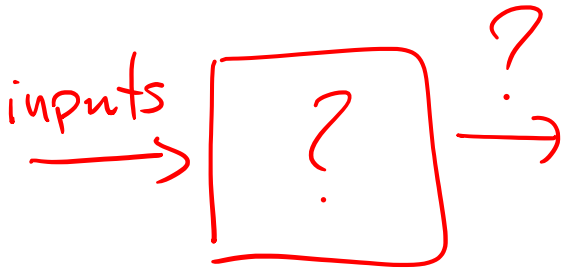
Unsupervised Learning

Heni Ben Amor, Ph.D.
Assistant Professor
Arizona State University

Ira A. Fulton Schools of Engineering
Arizona State University

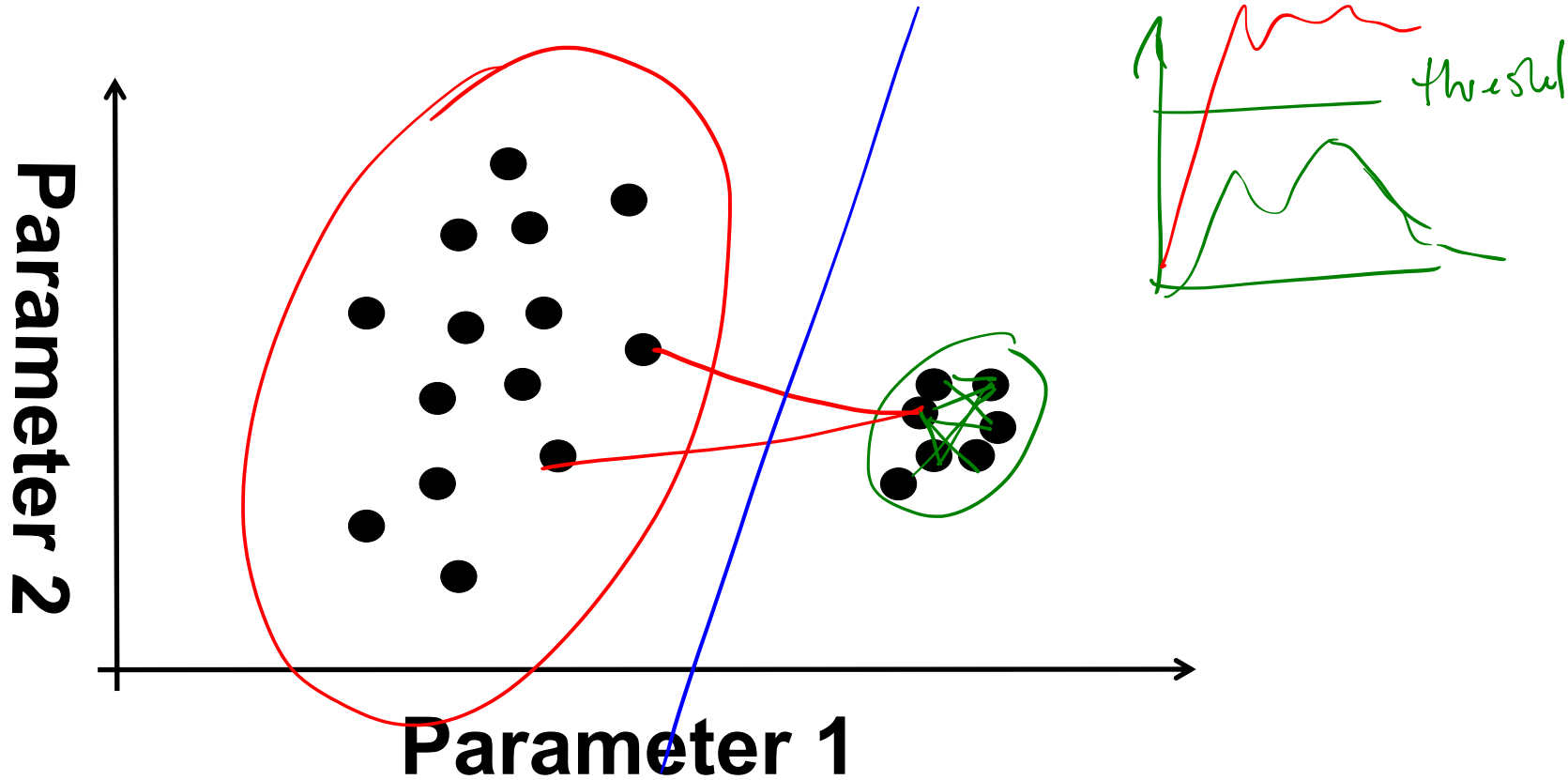
Unsupervised Learning

- What if we are only provided data, no labels?
- Can we still learn something from it?
- Focus on the structure of the data set
- Identifying hidden structures and manifolds
- Distances and spatial relationships
- Distribution of points



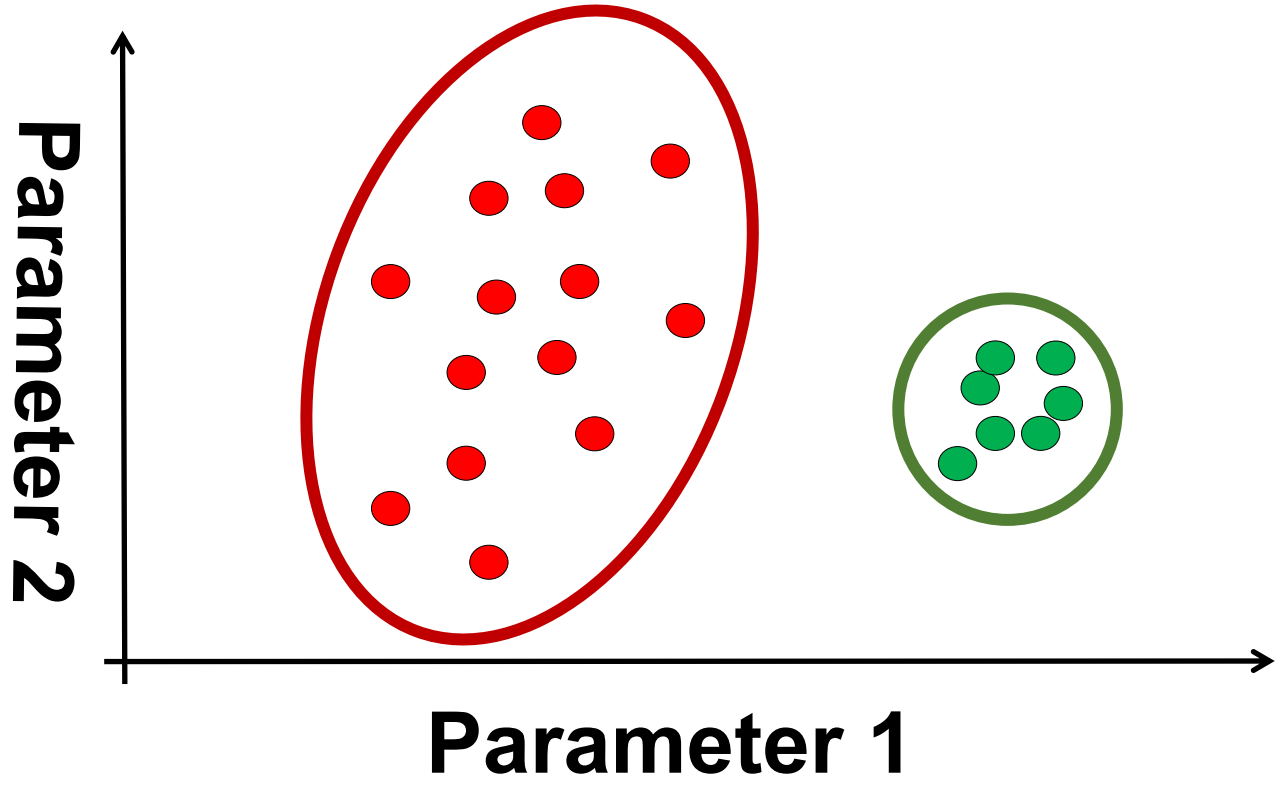
Clustering

- What if we get a data set like this?
- Can we discover some structure?



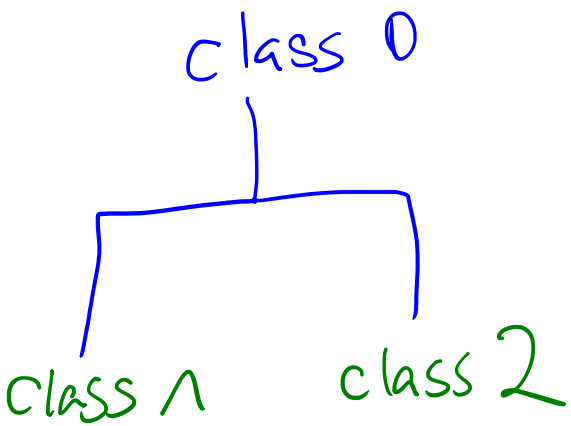
Clustering

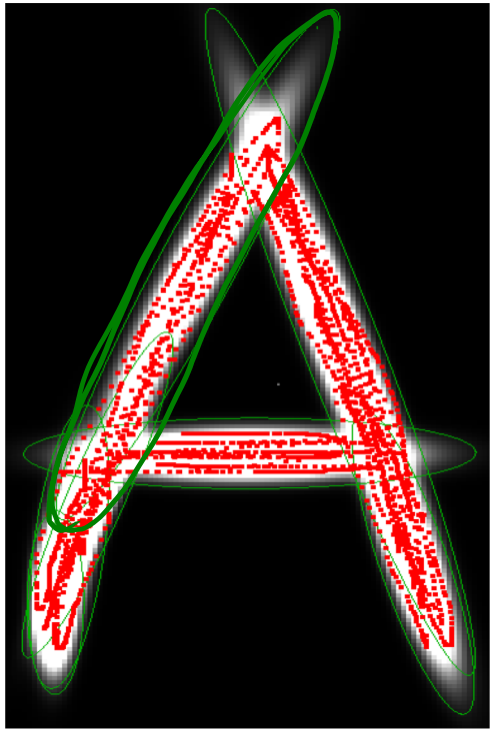
- What if we get a data set like this?
- Can we discover some structure?



Clustering

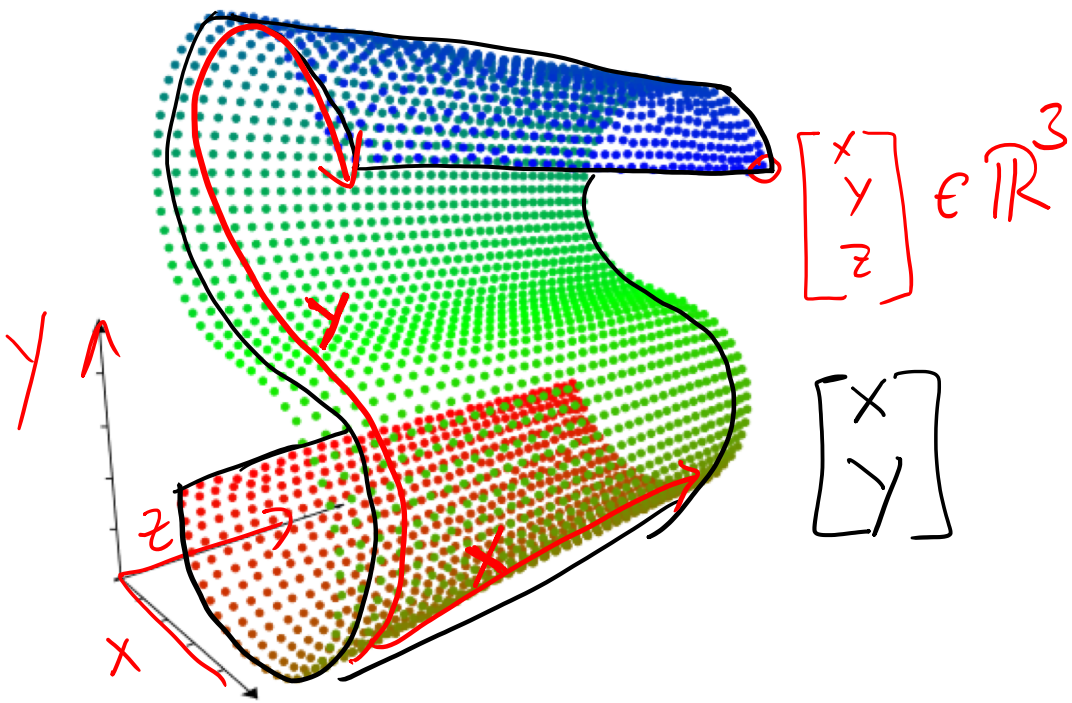
- Grouping individual data points together
- Points with small distances form a group
- Hard clustering or soft clustering possible
- Hierarchical clustering
- Identify hidden classes



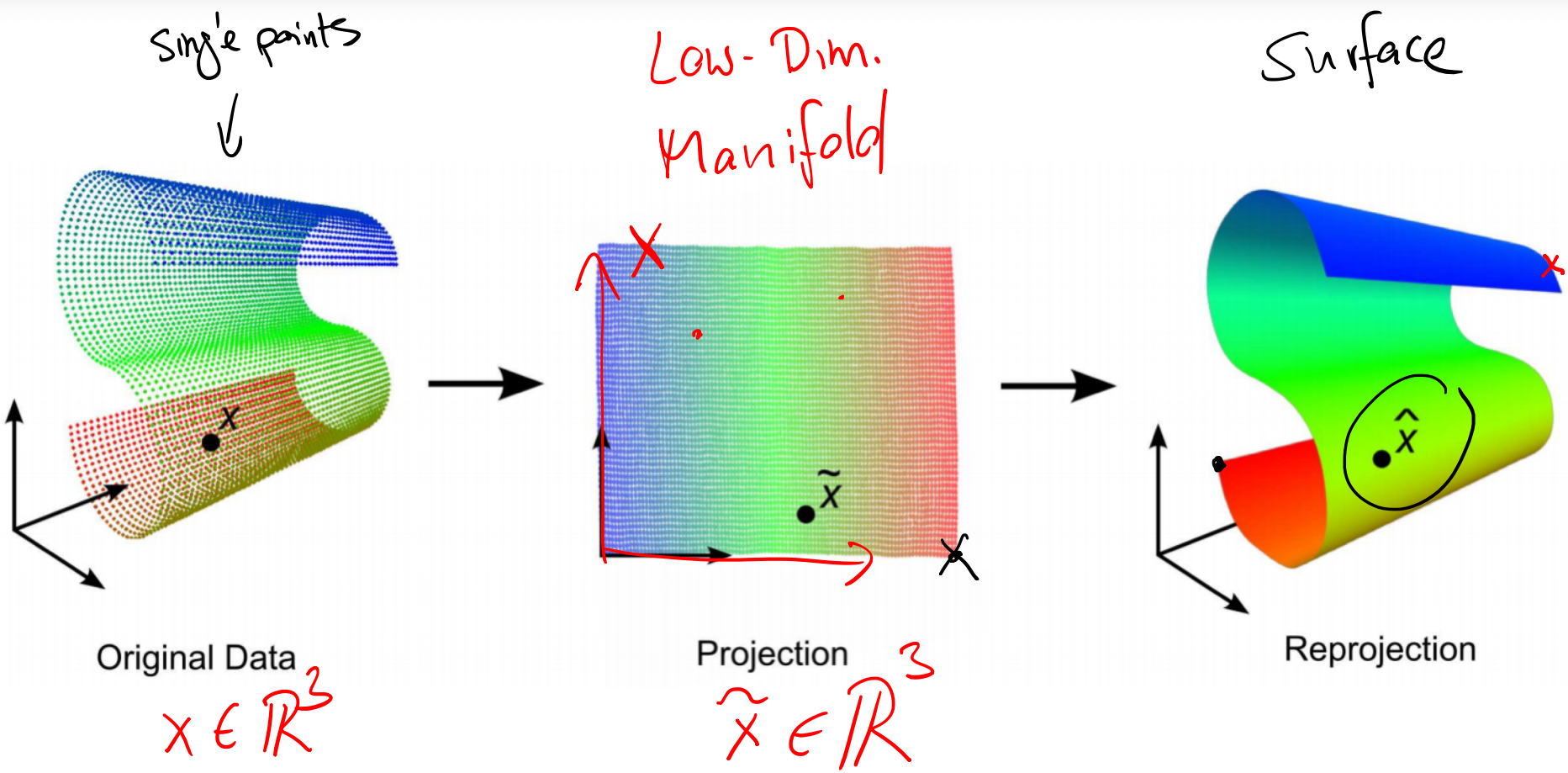


Dimensionality Reduction and Manifold Learning

- Assume data lies on a Manifold
- Manifold has lower dimensionality
- Identify manifold structure and represent data in lower-dimensional space

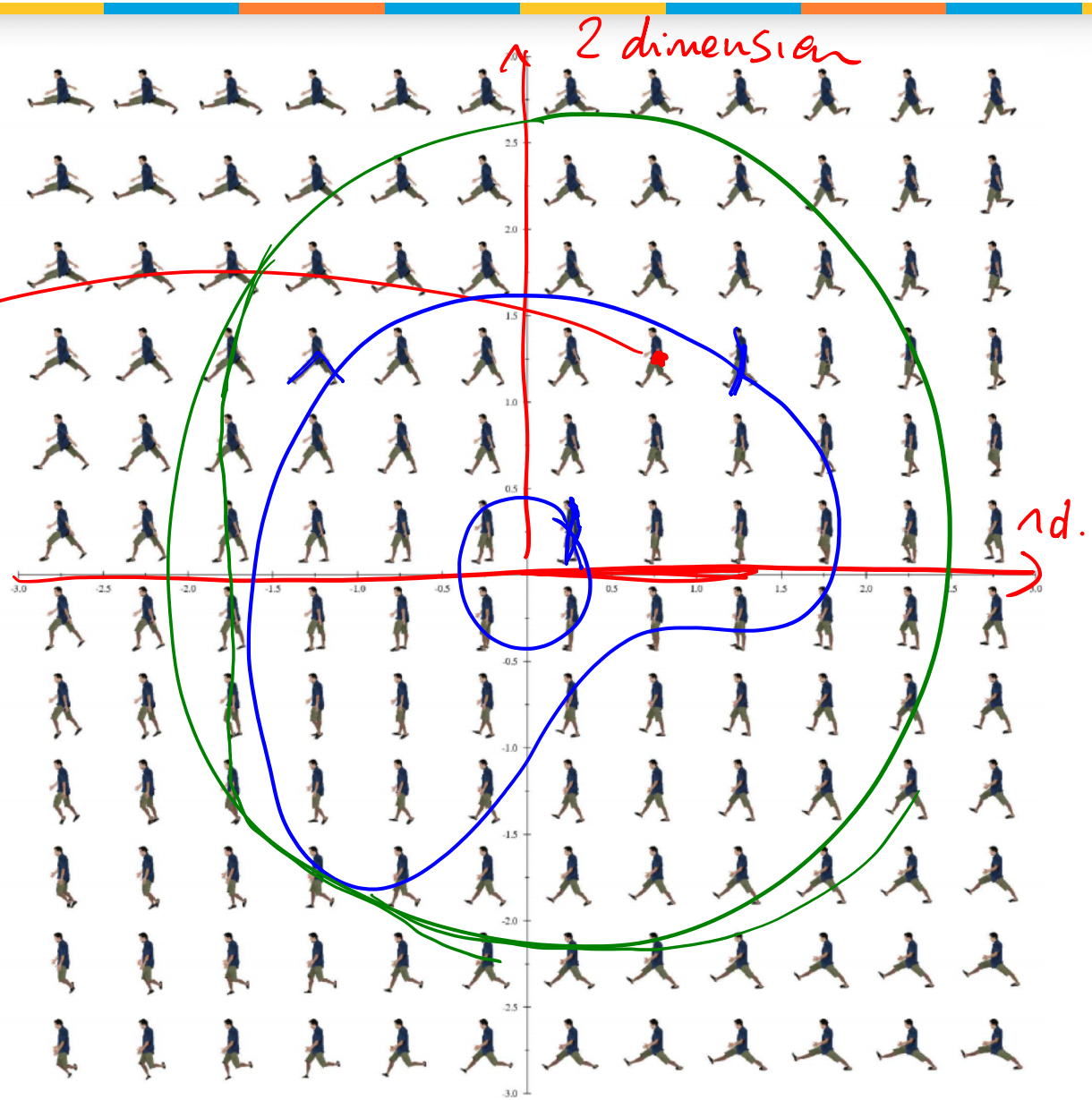


Dimensionality Reduction and Manifold Learning

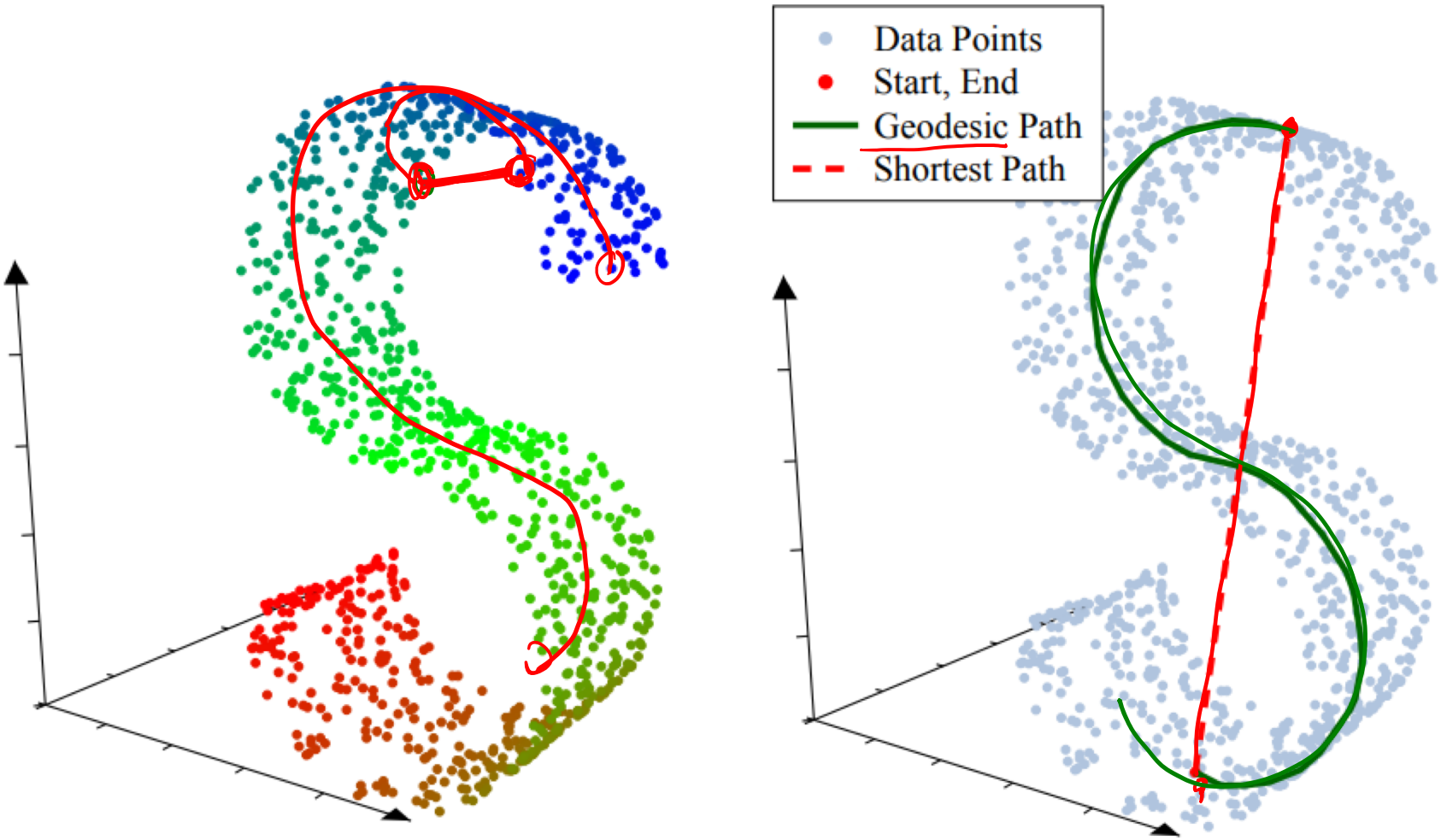


Two-Dimensional Walking Manifold

dim = 30



Distance Functions

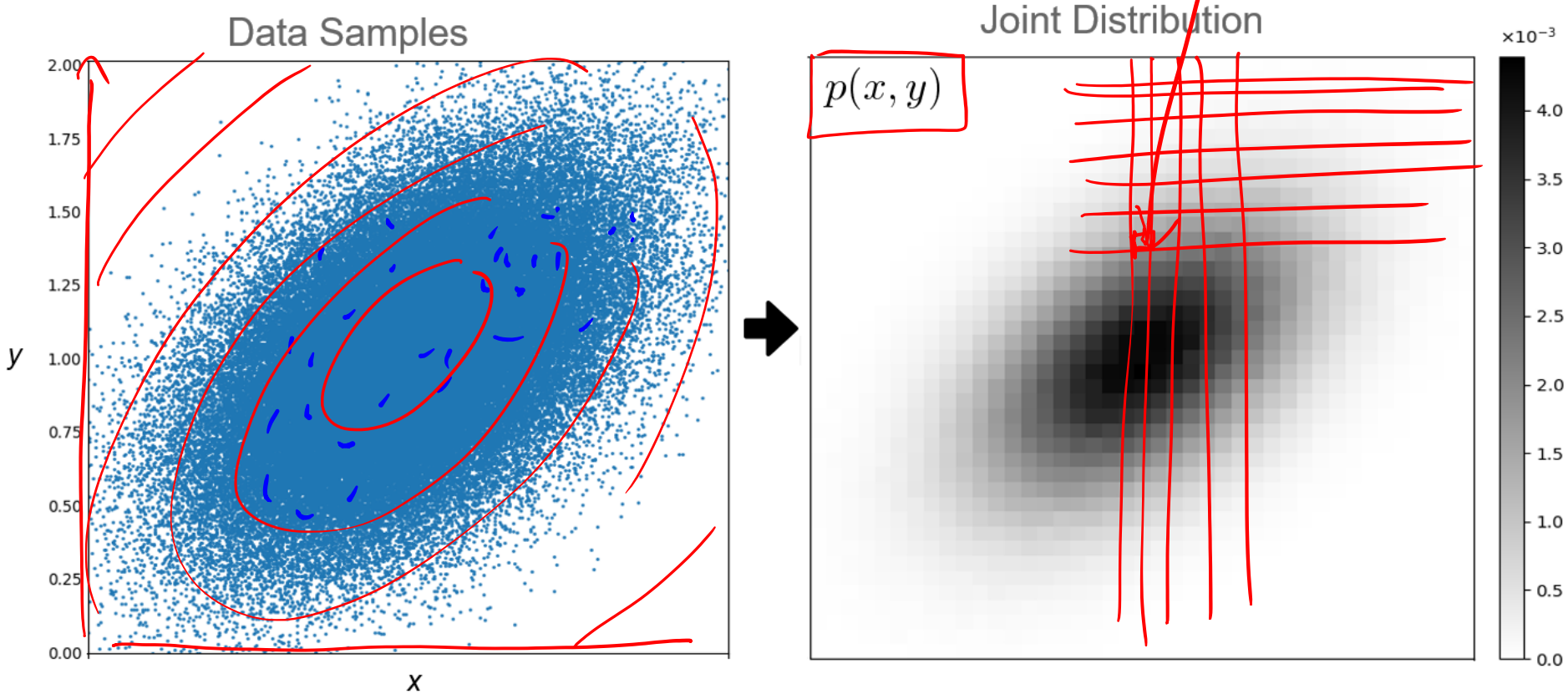


Learning Data Distributions

Example

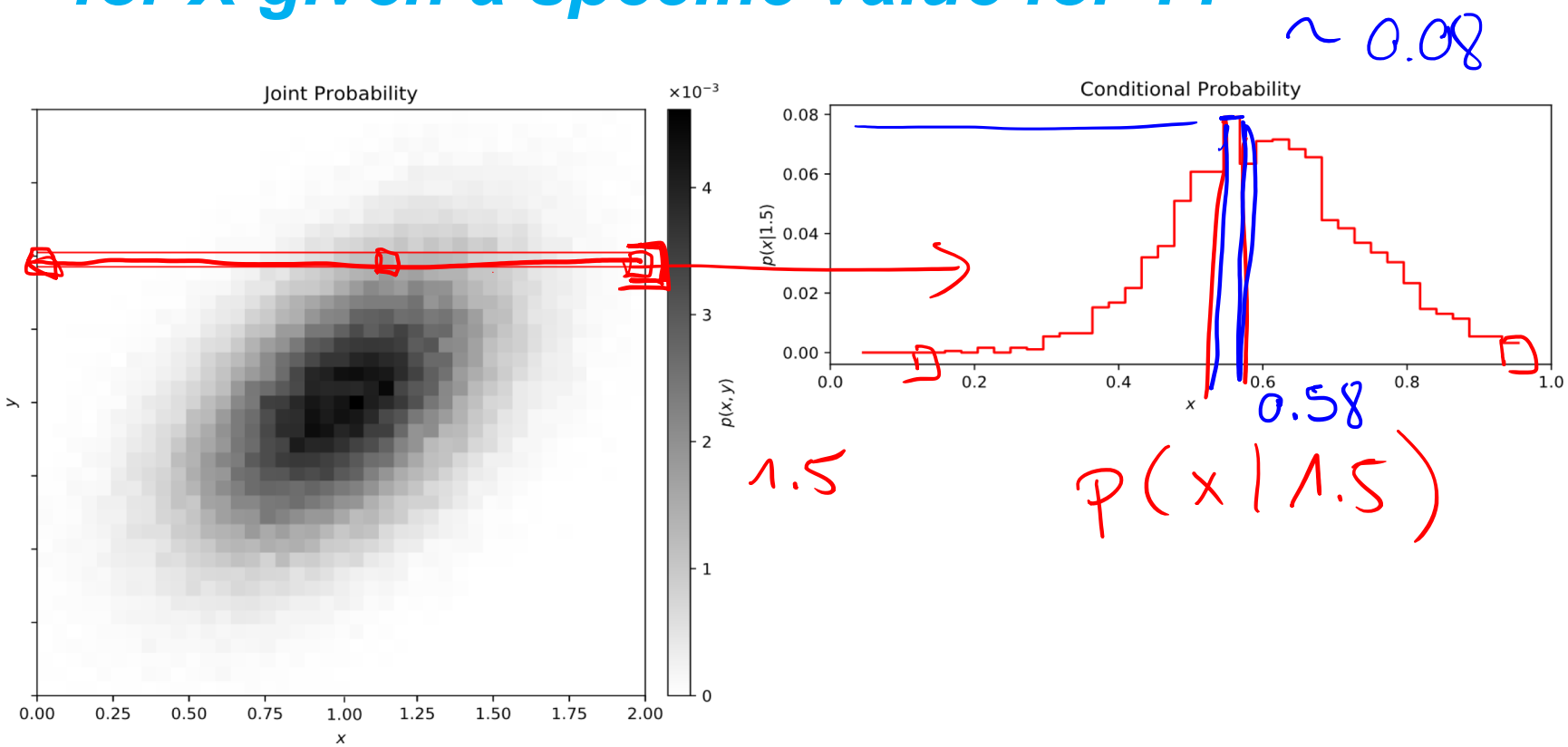
- Turn the data samples into an (approximate) joint distribution by binning and counting relative frequencies

how many points are inside / total # points



Probabilistic Reasoning

- Conditional probability distribution
- Now we can ask: ***"What is the probability for X given a specific value for Y?"***



Summary

- Unsupervised Learning
- Clustering
- Manifold Learning
- Distance Functions
- Data Distributions and Probabilistic Reasoning