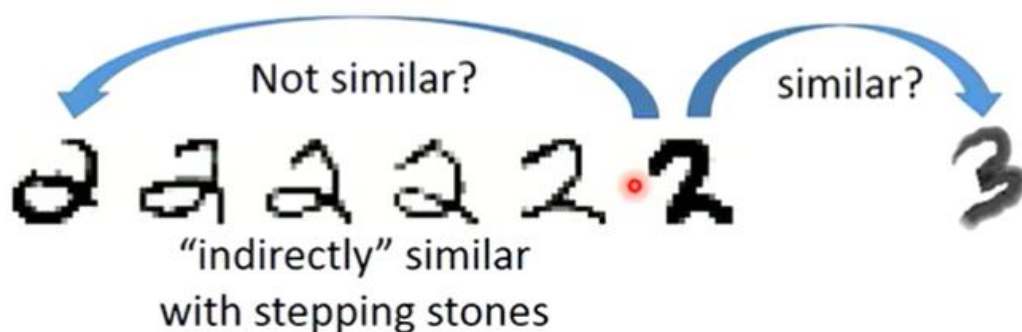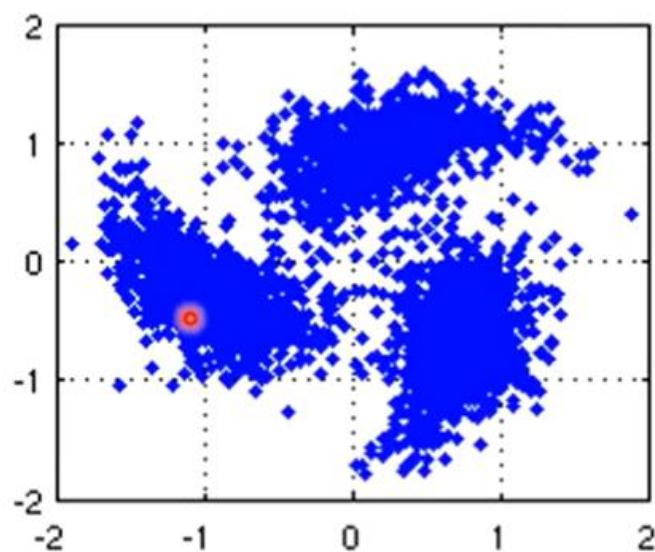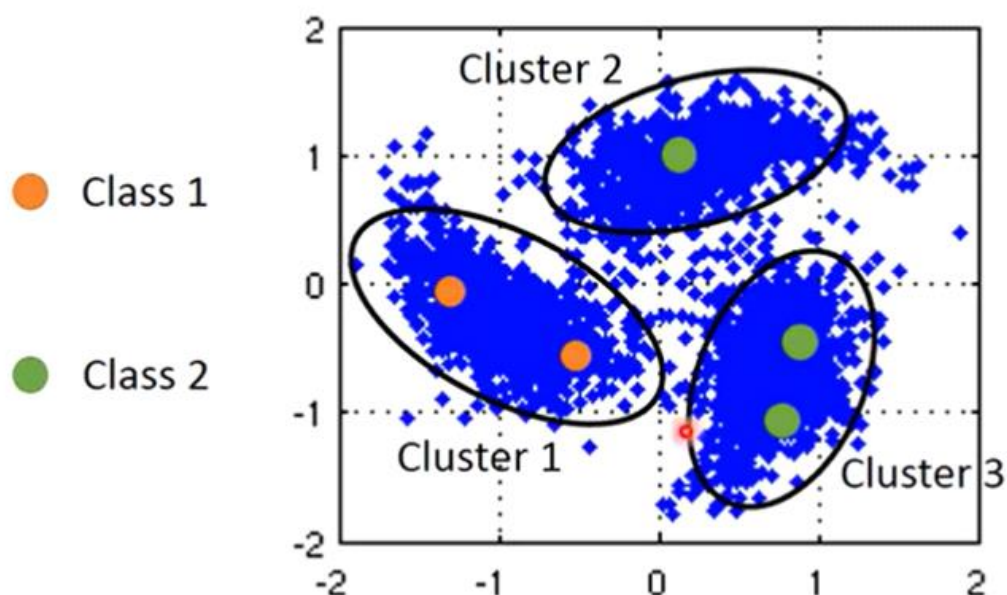# Smoothness Assumption

- Assumption: "similar" x has the same $\hat{y}$

- More precisely:
  - x is not uniform.
  - If x^1 and x^2 are close in a high density region, $\hat{y}^1$ and $\hat{y}^2$ are the same.

connected by a high density path

Not similar?   similar?

Not similar?   similar?

"indirectly" similar with stepping stones

Cluster and then Label

Class 1

Class 2
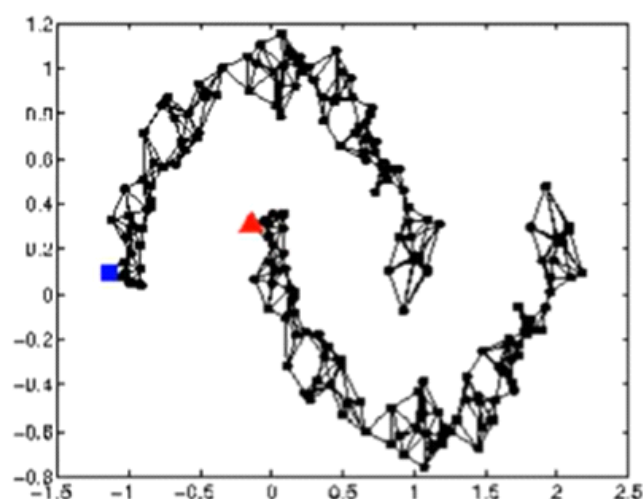
Deep auto-encoding then cluster

Graph-based Approach

- How to know x^1 and x^2 are close in a high density region (connected by a high density path)

Represented the data points as a *graph*

Graph representation is nature sometimes.
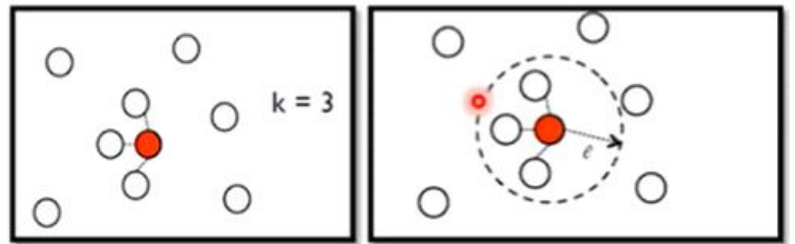
E.g. Hyperlink of webpages, citation of papers
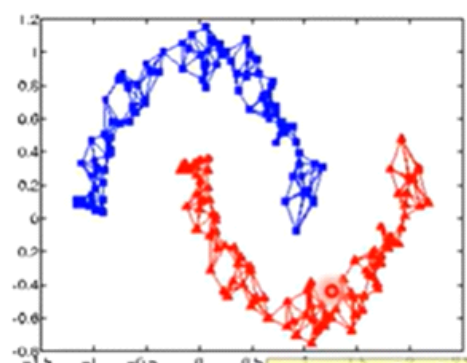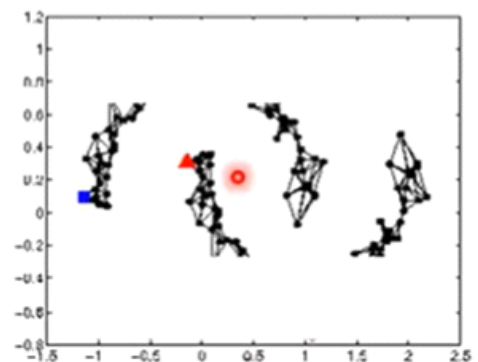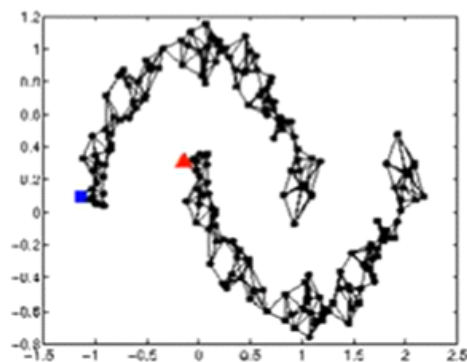
相连的表示一类, 不相连即使离的近也不好使

Graph-based Approach - Graph Construction

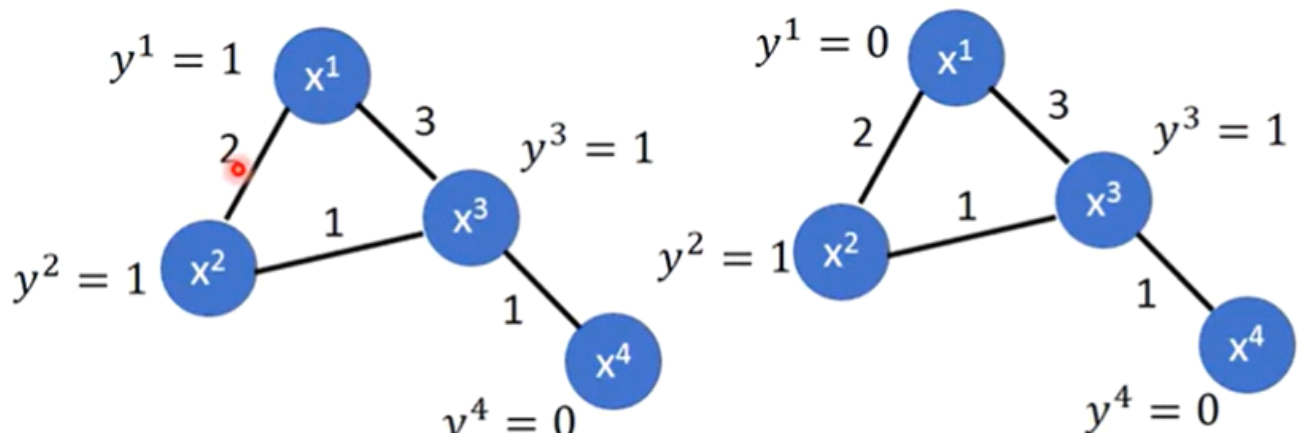The image is from the tutorial slides of Amarnag Subramanya and Partha Pratim Talukdar

- Define the similarity $s(x^i, x^j)$ between x^i and x^j
- Add edge:
  - K Nearest Neighbor
  - e-Neighborhood

需要数据足够多:

- Define the smoothness of the labels on the graph

左面比较smooth,三角形都是1,爪是0

定量描述他有多smooth

- Define the smoothness of the labels on the graph

$$S = \frac{1}{2}\sum_{i,j} w_{i,j}\left(y^i - y^j\right)^2$$