SMAI-M20-L34: Beyond Supervised Learning

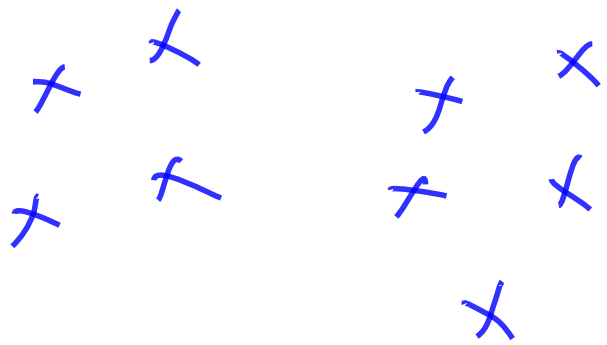
C. V. Jawahar

IIIT Hyderabad

November 9, 2020

Class Review

"Spectral clustering"



1. Consider a hierarchical classifier
   - Create an MST
   - Successively remove the longest or largest edges.

   What objective does it optimizes? maximizes? minimises? Is the answer unique? Does the performance depend on initialization?
2. Is clustering unique? Where do we use clustering?

$$\left[d(x_i, x_j) \right]$$

N - vertices

all pairs

N x N

Blank

Recap:

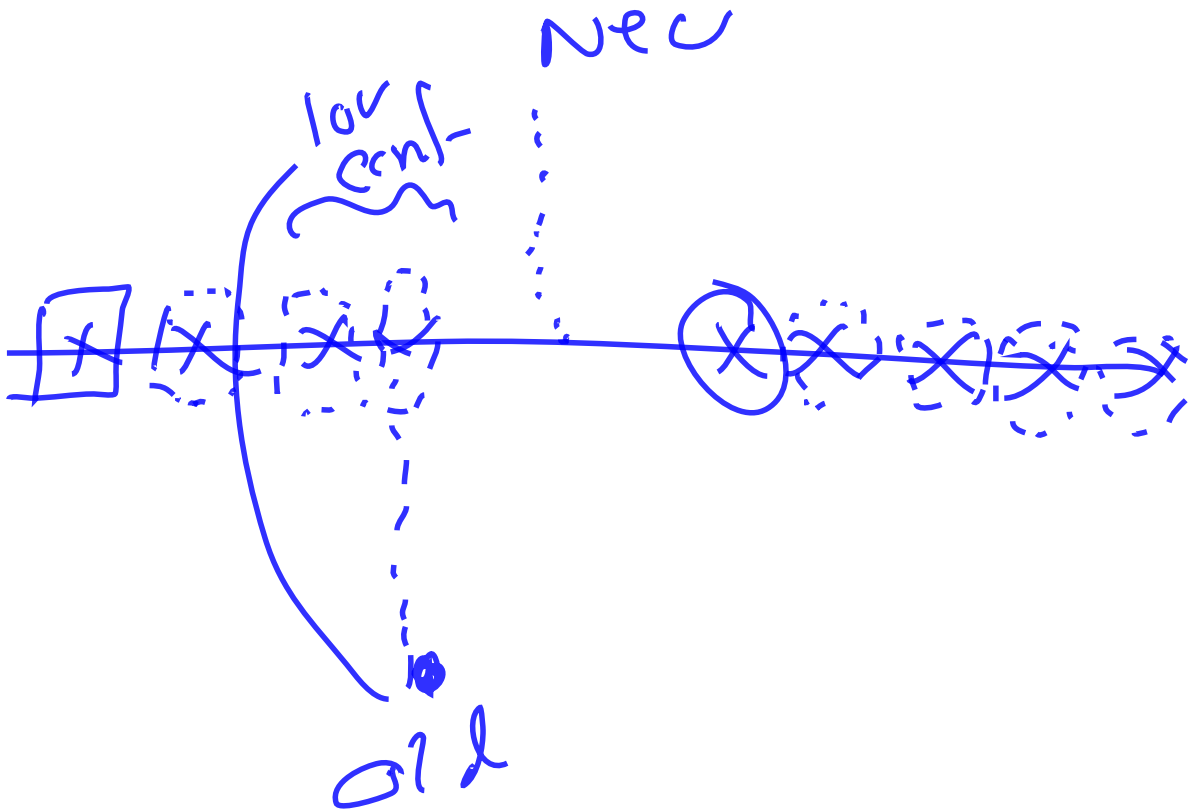
- **Matrix Factorization and Applications:** (i) SVD, (ii) Eigen Decomposition (iii) Matrix Completion (iv) LSI (v) Recommendations
- **Dimensionality Reduction and Applications:** (i) Feature Selection and Extraction (ii) PCA (iii) LDA (iv) Eigen face
- **Supervised Learning:** Formulation, Conceptual Issues, Concerns etc. (i) Loss Functions and Optimization (ii) Probabilistic View, Bayesian View, MLE (iii) Eigen Vector based optimization (iv) Gradient Descent: Stochastic and Batch GD (v) Classification and Regression
- **Classifiers:** (i) Nearest Neighbour, (ii) Notion of a Linear Classifier (iii) Perceptrons (iv) Bayesian Optimal Classifier (v) Logistic Regression (vi) Multiclass classification architectures (v) Decision Trees (vi) SVMs (hard margin, soft margin, kernel) (vii) Kernel trick and kernelized algorithms
- **Neural Network Architectures and Learning** (i) Neuron model, Single Layer Perceptrons (ii) SLP(iii) MLP (iv) Backpropagation (v) Chain rule (vi) Activations (vii) challenges in optimization (viii) Momentum (ix) Convolutional Layer (x) Recurrent/Feedback networks (xi) Auto-encoder and unsupervised learning
- **Beyond Simple Supervised Learning** (i) Paradigms of Learning (ii)

This Lecture:

1. **K-Means Clustering**
   1. Initialize Clusters randomly
   2. Compute Means for each cluster
   3. Re-Assign samples based on the distances to the means
   4. Repeat until no-change
2. **Self-Training**
   1. Small number of Labelled and Large number of unlabelled examples
   2. Build a classififer with labelled examples.
   3. Predict on unlabelled and use these "pseudo-labels" as "labells" and train the new classifier.
   4. Take only selected (say most confident samples for re-training).

Questions? Comments?

Blank



"zero shot"

"few shot"

Blank

Blank

Discussions Point - I

Consider K-Means (with K=2)

1. Six 1D samples

   $$\{-3, -2, -1, +1, +2, +3\}$$

   with $C_1 = \{-3, -2, +1\}$ and $C_2 = \{+3, +2, -1\}$

2. Six 1D samples

   $$\{-3, -2, -1, +1, +2, +3\}$$

   with $C_1 = \{-3, +2, +1\}$ and $C_2 = \{+3, -2, -1\}$

3. Four 2D samples:

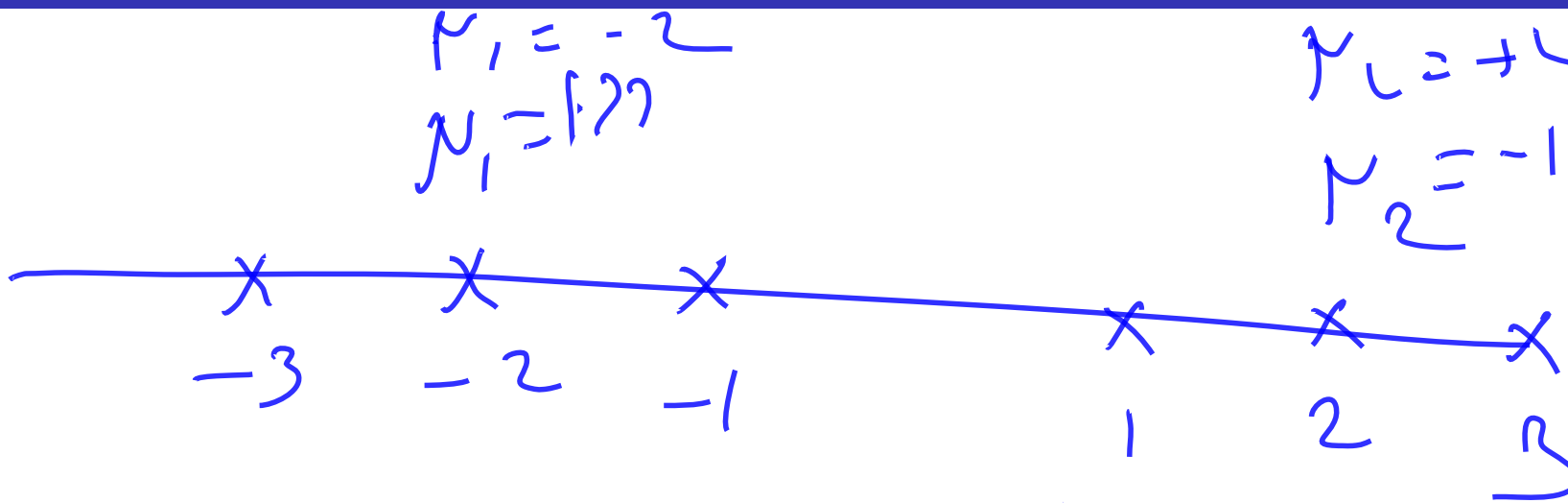
   $$(-1, -1), (-1, +1), (+1, +1), (+1, -1)$$

   with initialization as:

   - $C_1 = \{(-1, -1), (-1, +1)\}$ and $C_2 = \{(+1, -1), (+1, +1)\}$
   - $C_1 = \{(+1, +1), (-1, +1)\}$ and $C_2 = \{(+1, -1), (-1, -1)\}$

4. Comment on the optimization problem that K-Means solves (i) sensitivity to initialization (ii) convergence (iii) Local and global minima etc.

Blank



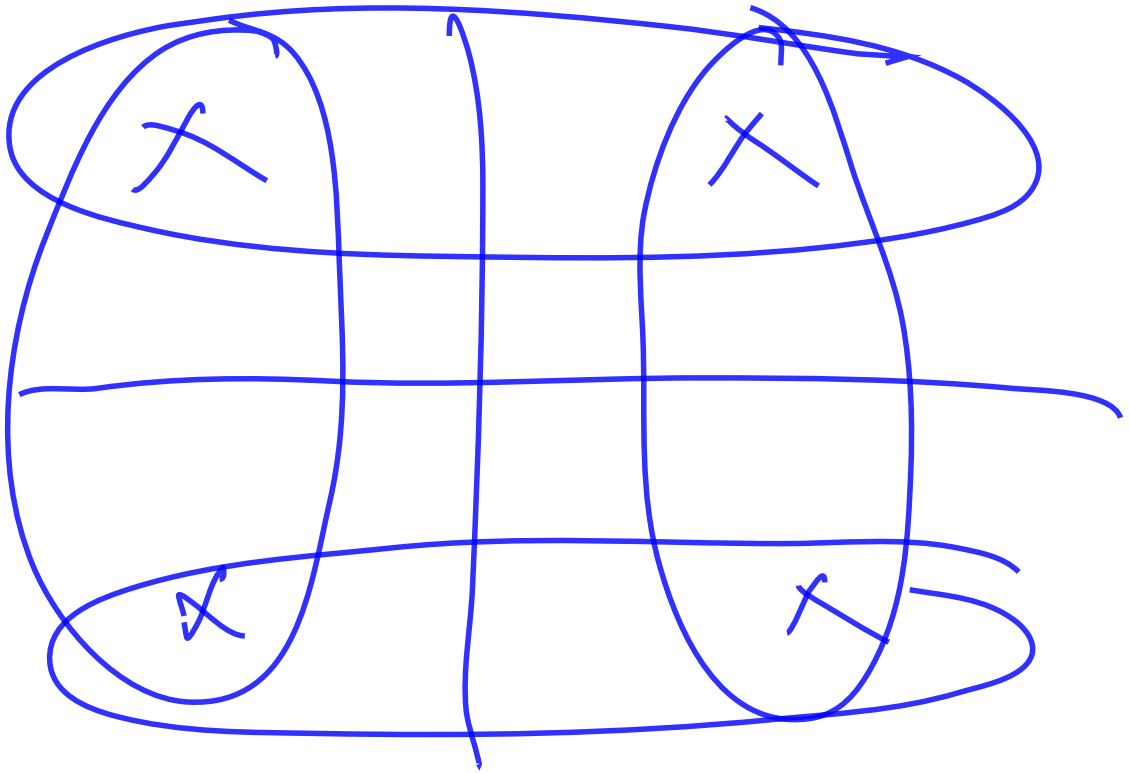
$C_1 = -3, -2, +1 \longrightarrow \mu_1 = -1.33$

$C_2 = 3 \quad 2 \quad -1 \longrightarrow \mu_2 = +1.33$

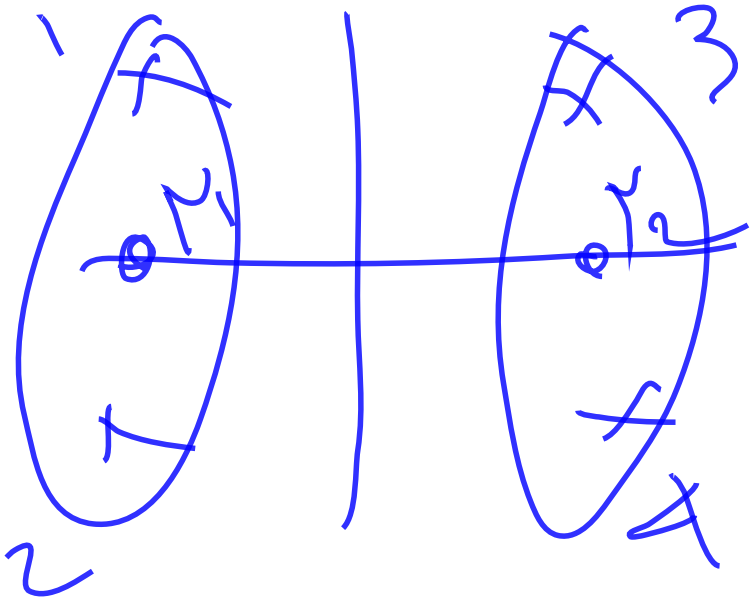
$C_1 = -3 \quad -2 \quad -1 \longrightarrow \mu_1 = -2$

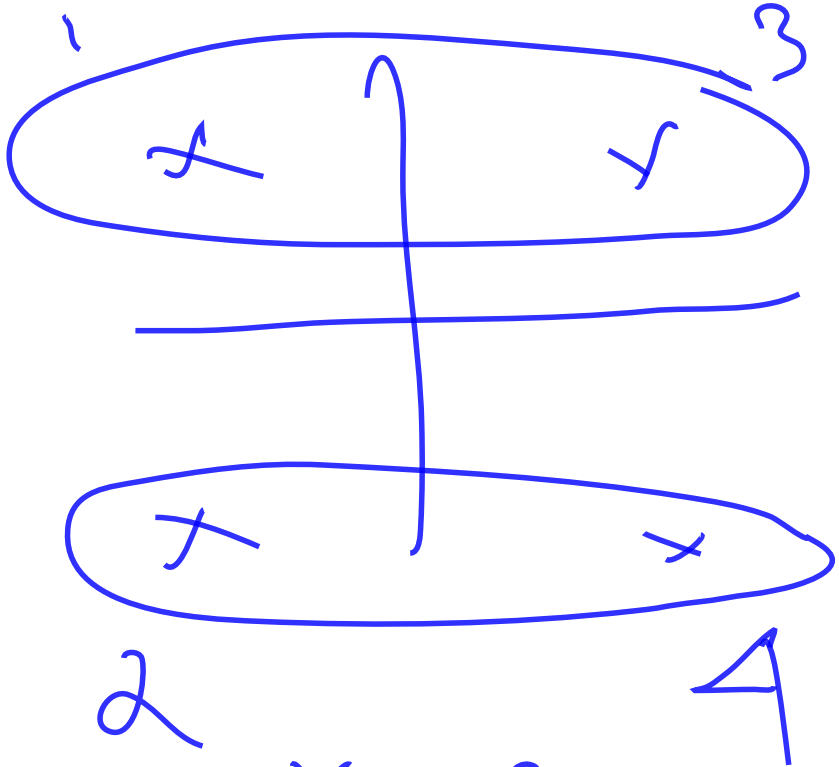
$C_2 = 1 \quad 2 \quad 3 \qquad \mu_2 = +2$

Blank



Blank



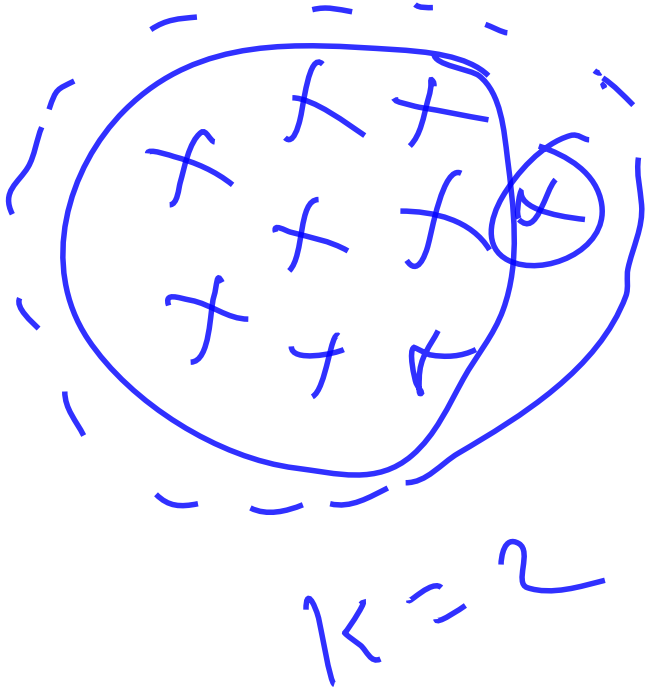


$K = 2$

$K = 2$

I same?

Blank



Blank

Discussions Point -II

Evaluation of Clustering is More Challenging than Classification. Objective is to sensitize the practical issues than exploring fully here.

How do we do performance evaluation in the Clustering problems?

1. How do we evaluate/quantify the quality of a specific cluster?
2. Does the objective (value) of the K-Means Obj. fn act as a good measure? How does it change with K? What should be desirable?
3. How do we evaluate/quantify the computational complexity of clustering?
4. How do we evaluate the final clusters (results)
    1. When there is no ground truth (GT)?
    2. Compare two different outputs, and choose the better one?
    3. When K is same (in both) and different?
    4. When there is full GT (as if from a manually labelled data as in classification; may be only in some course assignments!!)
    5. When there is some partial GT (some samples were manually labelled)
    6. When there is some partial GT (some pairs were labelled as "similar" and some "dissimilar". (or "must link", and "can not link" in graph terminologies)

Suggest situations where some of these will be needed.

Blank

Blank

Blank

Blank

What Next:?

1. Programming for Deep Learning.
2. NN Architectures and NN Learning (winding up)
3. Beyond Simple Supervised Learning