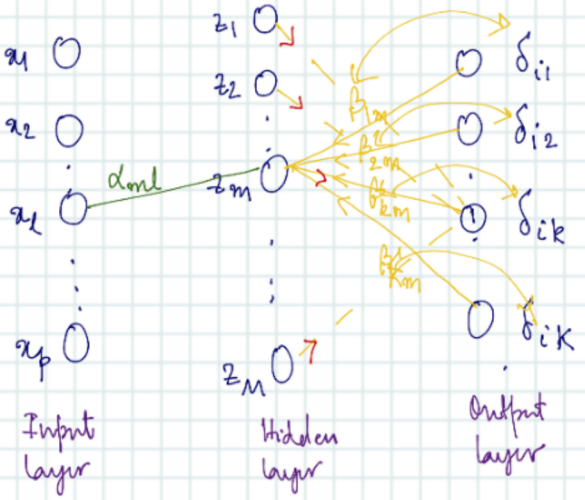
13/7/2021

AS 1604

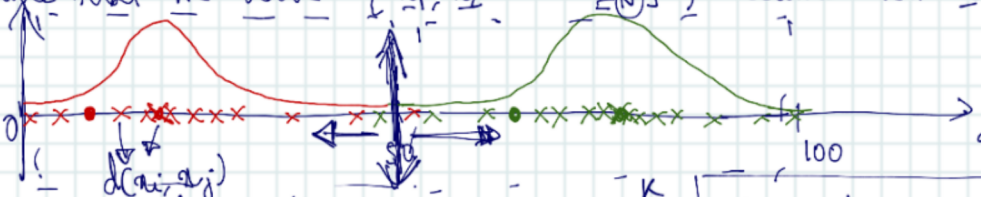
- Recap backprop algorithm
- Unsupervised learning algorithms
  - k-means clustering
  - Principal Component Analysis (PCA)

- Recap:



- Unsupervised learning: Discern patterns in data.

Assume that we have $\{x_1, x_2, \ldots x_N\}$; assume that $x \in \mathbb{R}^p$; K clusters ; $K \ll N$ (N: # data points)



- Within cluster scatter $= W(C) = \sum_{k=1}^{K} \sum_{C(i)=k} \sum_{C(i')=k} d(x_i, x_{i'})$

$C(i)$: clustering rule

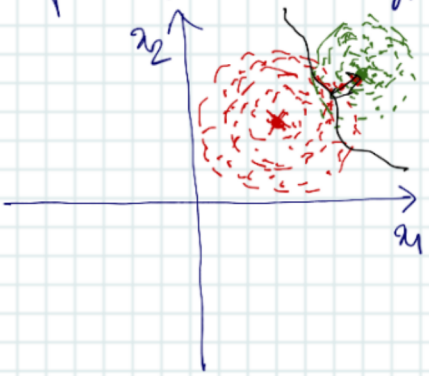
$C(i)$: maps a data point x_i to a cluster k.

- Between cluster scatter $= B(C)$

- Total scatter $= T = W(C) + B(C) \Rightarrow B(C) = T - W(C)$ → independent of C

Goal: Find C that minimizes $W(C)$, In the case of k means clustering, we choose $d(x_i, x_{i'}) = \|x_i - x_{i'}\|_2^2$

$$C^* = \underset{C}{\operatorname{argmin}} \frac{1}{2} W(C) \quad -(1)$$



K means clustering algorithm

- Given at input: $\{\underline{x}_1, \ldots \underline{x}_N\}$ K clusters
- Initialize randomly the means of each of the k clusters $\{\underline{m}_1^0, \underline{m}_2^0 \ldots \underline{m}_K^0\}$; $r = 0$ (iteration index)
- do:
  - $r = r+1$
  - Step 1: For all data points $\{\underline{x}_1, \ldots \underline{x}_N\}$: $C^{(r)}(i) = \underset{1 \le k \le K}{\operatorname{argmin}} \; d(\underline{x}_i, \underline{m}_k)$
  - Step 2: Update centroids or means: $\underline{m}_k^{(r)} = \frac{1}{N_k} \sum_{C^{(r)}(i)=k} \underline{x}_i$ → N_k: # points in cluster k.
- while: $\sum_{k=1}^{K} \| \underline{m}_k^{(r)} - \underline{m}_k^{(r+1)} \|_2^2 > \varepsilon$

- Principal Component Analysis (PCA)