Clustering with DL

① clustering with DL

Reconstruction loss : $L = \underset{\text{AE}}{d}(x_i, f(x_i)) = \sum_i \| x_i - f(x_i) \|^2$

Auto encoder : Non-cluster loss.

cluster losses :

② K-means loss : (Yang et al. 2006)

$$L(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} s_{ik} \| z_i - \mu_k \|^2$$

N : Data point, K : No of cluster, μ_k : cluster center, z_i : embedded Data point, s_{ik} : boolean value for cluster assignment

(Any Soft clase option) ??

② cluster assignment hardening :

$$q_{ij} = \frac{\left(1 + \| z_i - \mu_j \|^2 / v\right)^{-\frac{v+1}{2}}}{\sum_{j'} \left(1 + \| z_i - \mu_{j'} \|^2 / v\right)^{-\frac{v+1}{2}}}$$

v : constant

// lower distance to maximum

q_{ij} : softmax value.

$$P_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_{j'} \left(q_{ij'}^2 / \sum_i q_{ij'} \right)}$$

for all instances ~~in j class~~

// target Distribution P

Finally, loss function (for Network to minimize)

$$L = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

(iii) Balanced assignment loss:

(Dizaji et. al 2017)

$$L_{ba} = KL(G||U)$$

→ uniform Distribution

where, $g_k = P(y=k) = \frac{1}{N} \sum_i q_{ik}$

(iv) Locality preserving loss: (Huang 2014)

$$L_{Lp} = \sum_i \sum_{j \in N_k(i)} S(x_i, x_j) \|z_i - z_j\|^2$$

$S(x_i, x_j)$ ← similarity measure between x_i, x_j

$N_k(i)$ ← set of k nearest Neighbor of x_i

(v) Group Sparsity loss: (Ng... 2002)

$$L_{gs} = \sum_{i=1}^{N} \sum_{g=1}^{G} \lambda_g \|\phi^g(x_i)\|$$

weight of sparsity $\lambda_g = \lambda \sqrt{n_g}$

λ ← constant, n_g ← group size.

Combined losses: $L(\theta) = \alpha \underbrace{L_c(\theta)}_{\text{Clustering}} + (1-\alpha) \underbrace{L_n(\theta)}_{\text{Non clustering loss}}$