7/9/23 Deep Learning

- Recap SNE
- Wrap up SNE
- Issues with SNE: Asymmetry, High dim vs low dim distances (Crowding problem)
- t-SNE: (1) Symmetric SNE, (2) t-distributed SNE

- ANNs & Back prop

- $$P_i = p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|_2^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\|x_i - x_k\|_2^2 / 2\sigma_i^2\right)}$$

- $$Q_i = q_{j|i} = \frac{\exp\left(-\|y_i - y_j\|_2^2 / 2 \cdot \frac{1}{2}\right)}{\sum_{k \neq i} \exp\left(-\|y_i - y_k\|_2^2 / 2 \cdot \frac{1}{2}\right)}$$

- $$C = \sum_i KL(P_i \,\|\, Q_i)$$

- $$y^{(t)} = y^{(t-1)} - \eta \nabla C_{y^{(t-1)}} + \alpha(t)\left(y^{(t-1)} - y^{(t-2)}\right) \quad \text{— (A)}$$

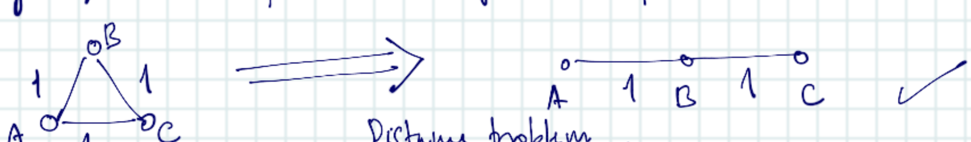
- σ_i is found by doing a search over $Perp(\cdot P_i) = 2^{H(P_i)}$ (which is a function of σ_i).

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}$$

- Remarks:
  - (i) SNE makes use of distance measure
  - (ii) SNE makes use of KL divergence
  - (iii) SNE makes use of entropy

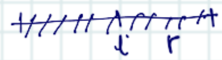
→ Aside

- Mapping equidistant points in high dim space to low dim space



Distance problem

- Area in a norm ball.



Crowding problem.

- Student t-distributed SNE (t-SNE): Two modifications to the original SNE formulations to address the challenges identified above.

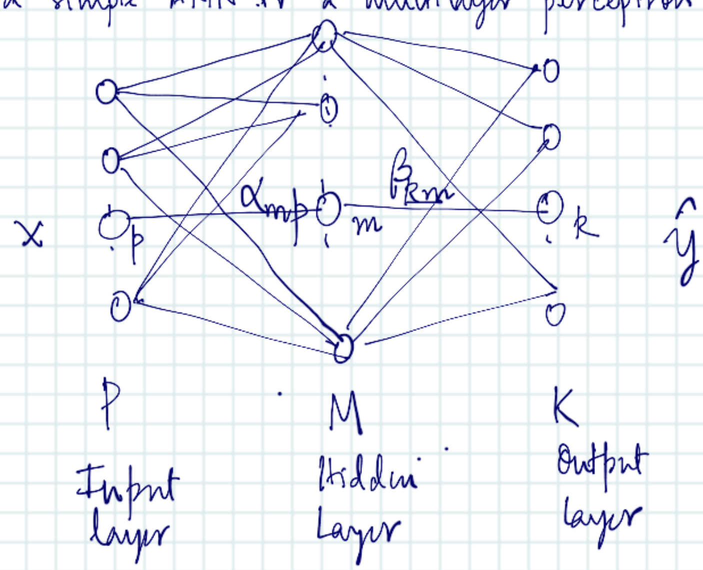
(i) Symmetric SNE: $P_i = \dfrac{p_{j|i} + p_{i|j}}{2n}$ → n is # datapoints

(ii) Mapping to a heavy tailed distributed - Student-t distribution

$$Q_i = q_{j|i} = \frac{\left(1 + \|y_i - y_j\|_2^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|_2^2\right)^{-1}}$$

- Artificial Neural Nets (ANNs): ANNs are popular functions used to model I/O relations in the supervised learning setting: $\hat{y} = f(x;\theta)$

- Let's define a simple ANN or a multilayer perceptron



$\hat{y} = f(x;\theta)$