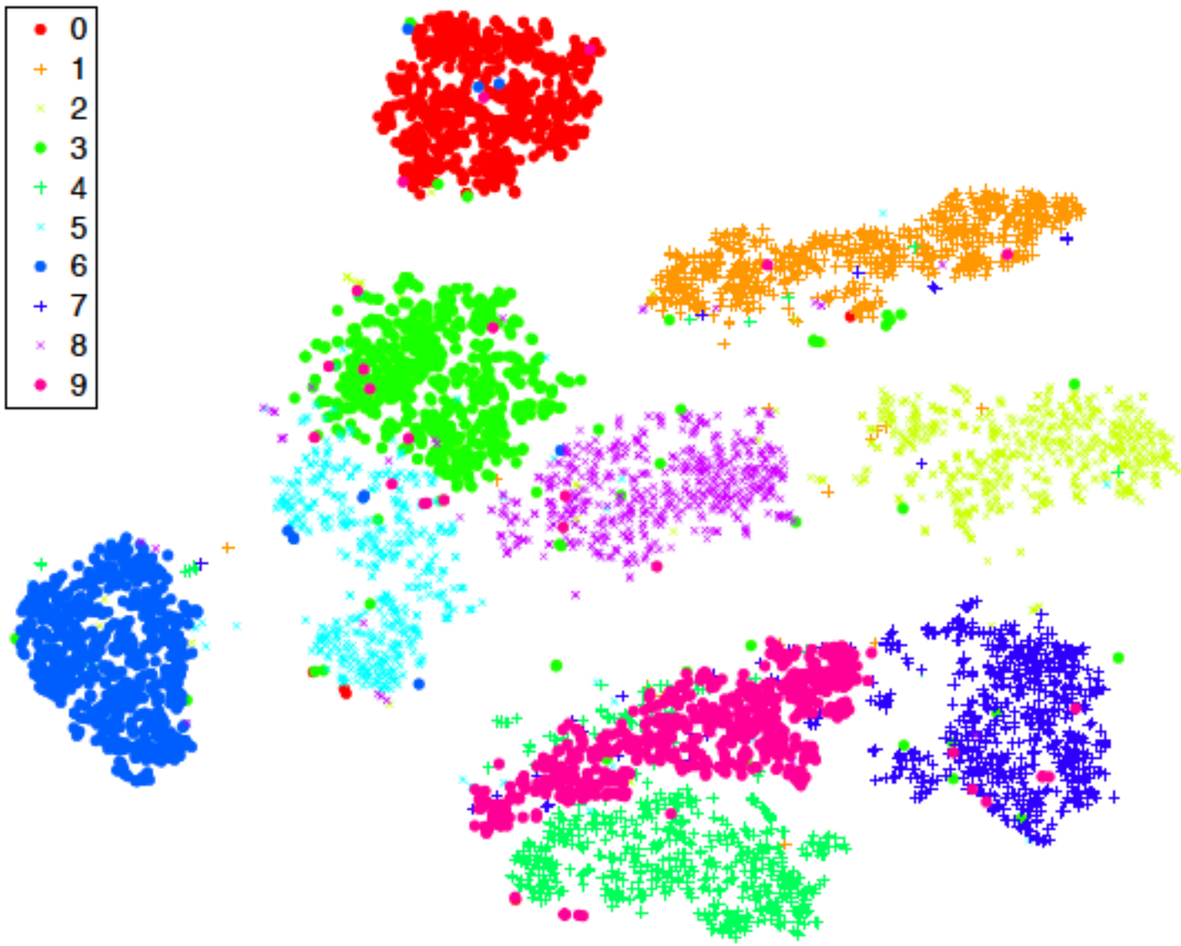
CSE 252D: Advanced Computer Vision

Manmohan Chandraker

Supplement: Visualization with t-SNE

Visualization with t-SNE

- Map high-dimensional data with t-distributed stochastic neighbor embedding



[van der Maaten and Hinton, Visualizing Data Using t-SNE, JMLR 2008]

Visualization with t-SNE

- Map high-dimensional data with t-distributed stochastic neighbor embedding
- Represent similarity for data points as a probability
- Assume neighbor x_j for point x_i picked based on Gaussian density

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\|x_k - x_l\|^2 / 2\sigma^2\right)}$$

"Perplexity" parameter, usual in range [5, 50]

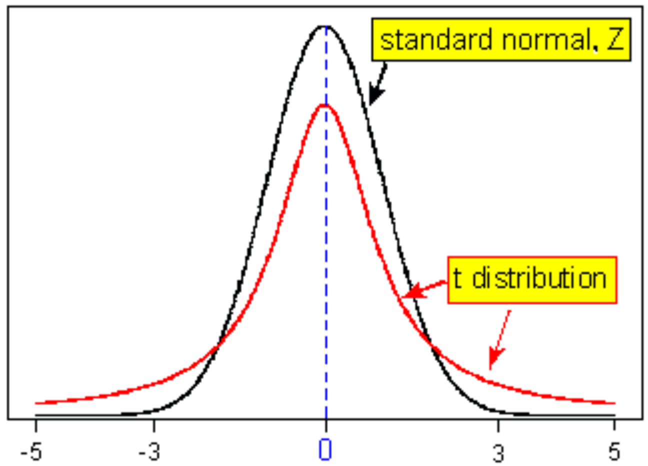
Visualization with t-SNE

- Map high-dimensional data with t-distributed stochastic neighbor embedding
- Represent similarity for data points as a probability
- Assume neighbor x_j for point x_i picked based on Gaussian density

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\|x_k - x_l\|^2 / 2\sigma^2\right)}$$

- Assume a t-student distribution for low-dimensional points

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



Visualization with t-SNE

- Map high-dimensional data with t-distributed stochastic neighbor embedding
- Represent similarity for data points as a probability
- Assume neighbor x_j for point x_i picked based on Gaussian density

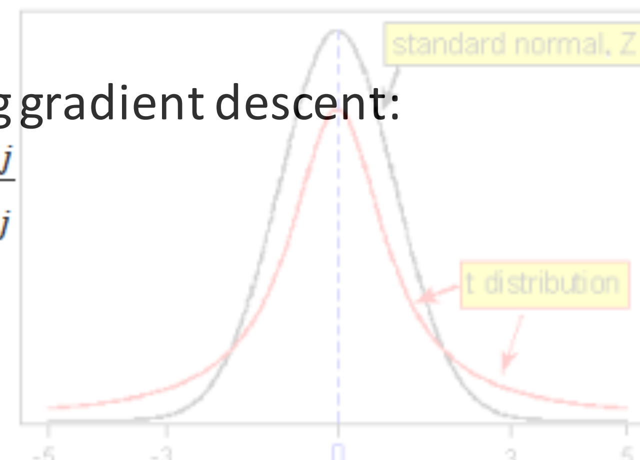
$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\|x_k - x_l\|^2/2\sigma^2\right)}$$

- Assume a t-student distribution for low-dimensional points

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

- Minimize a cost given by the KL-divergence using gradient descent:

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



Visualization with t-SNE

- Map high-dimensional data with t-distributed stochastic neighbor embedding
- Represent similarity for data points as a probability
- Assume neighbor x_j for point x_i picked based on Gaussian density

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\|x_k - x_l\|^2/2\sigma^2\right)}$$

- Assume a t-student distribution for low-dimensional points

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

- Minimize a cost given by the KL-divergence using gradient descent:

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

- Heavier tail means points moderately far in x are mapped further in y
    - More volume at greater distance in high dimensions
    - Mapping nearby points in low dimension requires pushing out far points