Carnegie Mellon University
Heinz College

95-865 Unstructured Data Analytics

t-SNE: some technical details

George Chen

For the purposes of this course, you do *not* need to know these technical details (I’m providing these just to give you a flavor of what some algorithms are like)

Technical Detail for t-SNE

Fleshing out high level idea #1 (from lecture slides)

Suppose there are n high-dimensional points $x_1, x_2, \ldots, x_n$

For a specific point i, point i picks point j ($\neq i$) to be a neighbor with probability:

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

σ_i (depends on i) controls the probability in which point j would be picked by i as a neighbor (think about when it gets close to 0 or when it explodes to ∞)

σ_i is controlled by a knob called **perplexity**
(rough intuition: it is like the “number of nearest neighbors” in Isomap)

Points i and j are "similar" with probability: $p_{i,j} = \frac{p_{j|i} + p_{i|j}}{2n}$

This defines the blue distribution in the lecture slides

Technical Detail for t-SNE

Fleshing out high level idea #2 (from lecture slides)

Denote the n low-dimensional points as x_1', x_2', …, x_n'

Low-dim. points i and j are "similar" with probability: $q_{i,j} = \frac{\frac{1}{1+\|x_i' - x_j'\|^2}}{\sum_{k \neq m} \frac{1}{1+\|x_k' - x_m'\|^2}}$

This defines the green distribution in the lecture slides

Fleshing out high level idea #3 (from lecture slides)

Approximately minimize (with respect to $q_{i,j}$) the following cost:

$$\sum_{i \neq j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}}$$

This cost is called the “KL divergence” between distributions p and q