Analyzing Visualization and Dimensionality-Reduction Algorithms

Oliver Hayman

Mentor: Ashwin Narayan

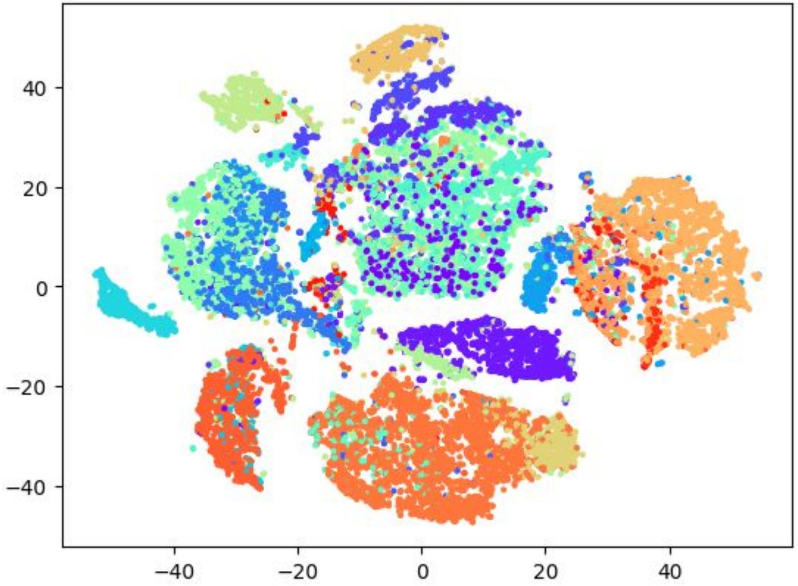
May 18th 2019

MIT PRIMES Conference

Motivation

Algorithms are needed to spot patterns in high dimensional data sets

Not perfect at preserving relationships



Expression patterns from different mouse brain cells

t-distributed Stochastic Neighbor Embedding

Probability distribution on points in high-dimensional space:

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

- probability of picking x_j in Gaussian distribution centered at x_i

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2n}$$

- modified probability of picking points in joint Gaussian distribution

similarly in the embedded space:

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

- probability of picking points in joint Student's t-Distribution

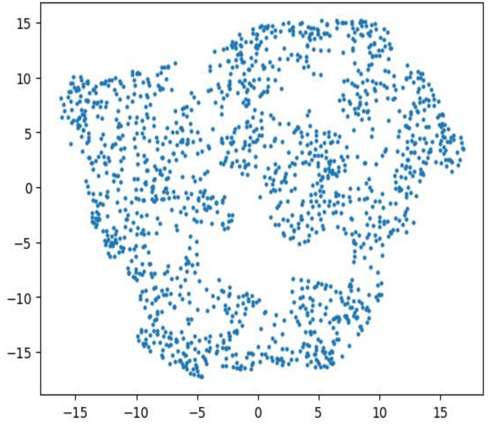
t-SNE minimizes the distance b/w these two distributions

Choice of *perplexity* is key

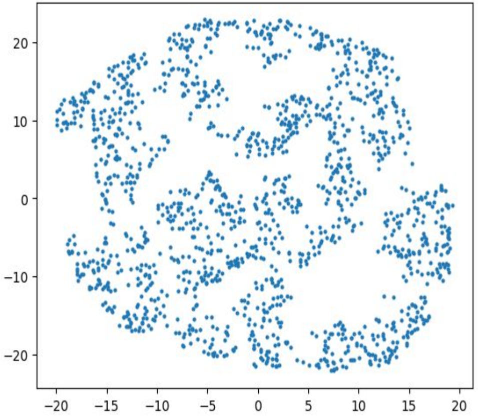
Choose σ_i so that it is bigger in sparse regions and smaller in dense regions

Governed by a parameter called perplexity - can be thought of as number of neighbors for each point grouped together

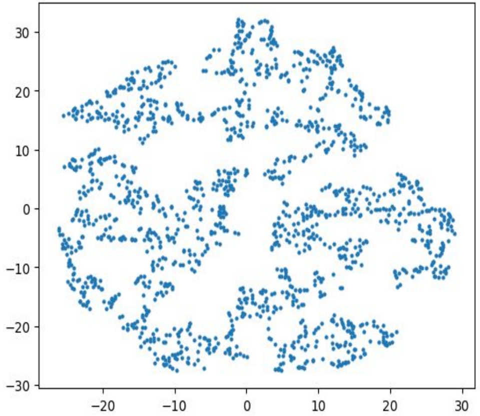
Example



Perplexity: 80



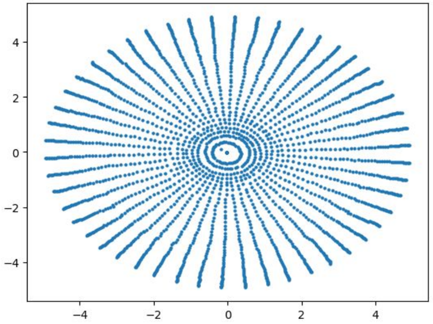
Perplexity: 50



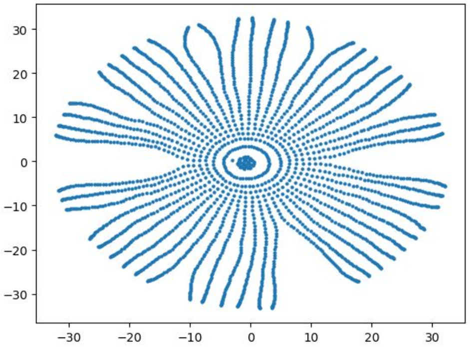
Perplexity: 30

Example runs

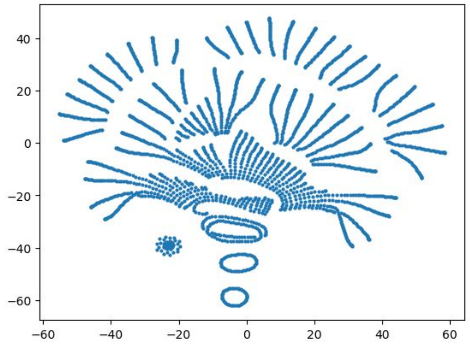
Perplexity: 80



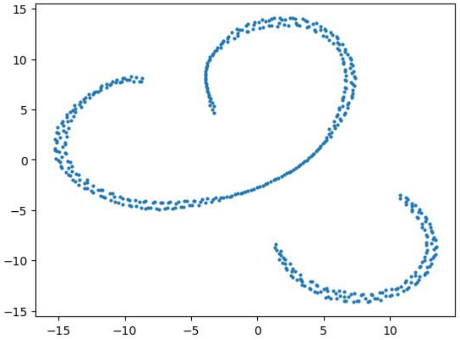
Perplexity: 30



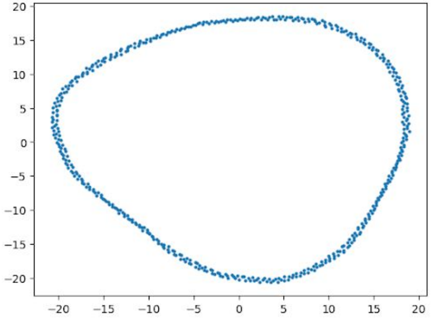
Perplexity: 10



Perplexity: 50



Perplexity: 30



What makes a visualization “good”?

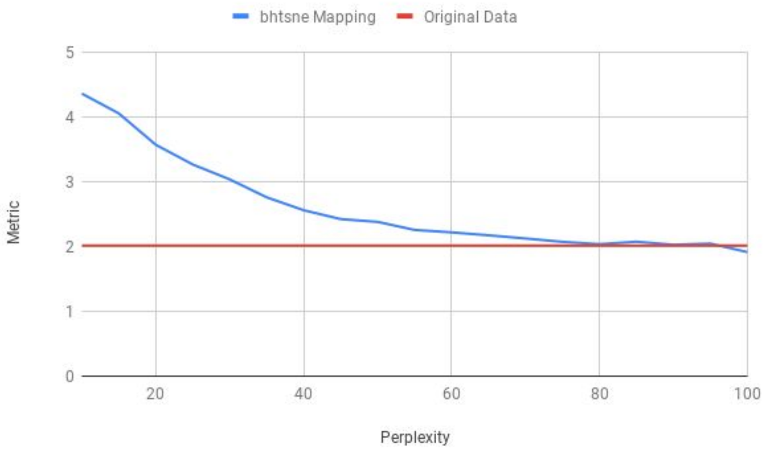
Need metric to measure clustering

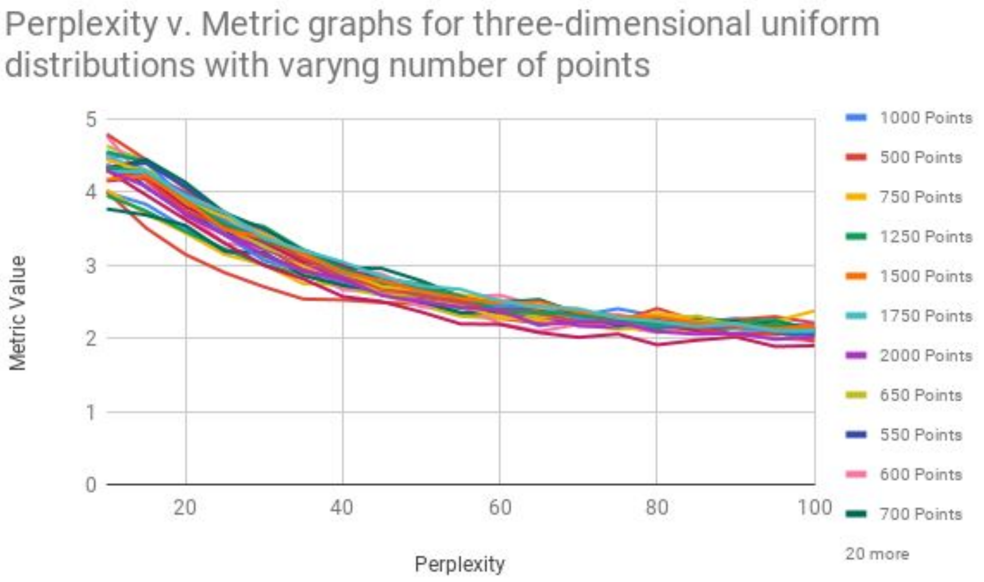
$\alpha(x_i, a)$ - distance from x_i to ath closest data point

$\beta(x_i, a)$ - number of points whose distance from x_i is less than a

For set X, metric is $\frac{1}{n}\sum_{x_i \in X} \beta\left(x_i, \left(\frac{2\alpha(x_i, c)}{c}\right)^{\frac{1}{d}}\right)$ - 2 hyperspheres, metric based on number of points in smaller one

Perplexity choice affects clustering





Fit to equation $F(x) = ae^{-bx} + c$

Towards theoretical results

For uniform distributions in unit hypercube,

$V(r)$ - volume of hypersphere

$f(x,r)$ - volume of hypersphere cut a distance x from center

Formulas for P(larger hypersphere having radius r), P(point in smaller hypersphere | larger hypersphere has radius r) used to determine expected value

Made assumption that hypersphere will only intersect one edge of hypercube (gives an approximation)

Future work

Find and prove relationship between perplexity and clustering

Find modification to algorithm

Apply methods to other algorithms

Define other metrics for different properties

Acknowledgements

I would like to thank my mentor, Ashwin Narayan, for suggesting this project, helping focus my research, and being extremely flexible with meetings. I would also like to thank the MIT PRIMES-USA program for giving me this opportunity to work on the type of research most people only get involved in later in their academic careers. Dr. Tanya Khovanova was especially helpful in guiding me through the planning process for my project and in working with me on my presentation.

References

MacKay, D. J. (2003). Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press.

Van der Maaten, L., & Hinton, G. (2008, November). Visualizing Data using t-SNE (Y. Bengio, Ed.) In JMLR Volume 9. Retrieved May 16, 2019.