Script

Taran Lynn

1. Start off by asking what the purpose of data visualization is.
   (a) The purpose is to see how data is structured.
   (b) This is done through manifold learning. A manifold is a lower-dimensional Euclidean object embedded and distorted into higher dimensional space. [Show example of 2D manifold in 3D space being put on a 2D space]
2. Broad overview of how Stochastic Network Embedding works.
   (a) SNE tries to preserve the higher dimensional distances between point on a 2D or 3D plain.
   (b) If we imagine that each data pair of points in higher dimensional space is connected by a spring of strength $p_{j|i}$. SNE can be thought of as trying to squash points down into lower dimensional space. The spring lengths will $q_{j|i}$ probably be different, but their force, caused by gradient descent, will make them to approach $p_{j|i}$. Not a completely accurate analogy because $p_{j|i} \neq p_{i|j}$, so spring length is different when viewed from x_i and x_j (important later). [Show spring graphic]
   (c) First, each data point is assigned a similarity $p_{j|i}$ to every other point based on where it is in a higher-dimensional Gaussian probability density centered on x_i. Ignore how σ_i is chosen for now. [Use 2D Gaussian as an example]
   (d) Then we come with create a set of lower dimensional points equal in number to the higher dimensional ones. These point get similarity values $q_{j|i}$. We then perform gradient descent to minimize the difference between every $p_{j|i}$ and $q_{j|i}$ for every i and j, as measured by the Kullbeck-Leibler divergence. [Note that since P_i and Q_i are probability distributions, KL is ≥ 0]
   (e) σ_i is chosen via binary search on perplexity. Can be interpreted as a smooth measure the number of neighbors. [$H(P_i)$ is the entropy in the neighborhood sample, i.e. the number of bits needed to represent the uncertainty, so $2^{H(P_i)}$ is like the number of neighbors.]

3.

Problems with SNE