Notes

- Symmetric SNE uses the sum of all distance pairs, should be constant for every p_{ji}. Since $||x_i - x_j||^2 = ||x_j - x_i||^2$ we get $p_{ij} = p_{ji}$.
- $p_{j|i}$ is the similarity of x_j to x_i. In Euclidean space this is the probability x_i will pick x_j as its neighbor. Neighbor probability is given using a Gaussian distribution from x_i.
- Lower dimension similarities are normalised so that $\sigma_i = 1/\sqrt{2}$.
- KL divergence, aka relative entropy, is a measure of dissimilarity between two distributions.
- Perplexity kind of controls the number of neighbors of a point, controls springy-ness.

1. Xiaoli tSNE
2. Taran SNE
3. Xiaoxing programming and evaluation