salient-SNE effectively interprets non-linear clustering

Ethan S. Dyer (edyer@stanford.edu), Karthik A. Jagadeesh (kjag@stanford.edu)

Stanford University

Introduction

Various techniques exist to reduce the dimensionality of large sets of points. PCA is frequently used to linearly transform data by identifying the principal components in the direction of the largest variance in the data. PCA falls short trying to tease structure out of non-trivial high-dimensional data sets. t-SNE, a non-linear reduction method, is effective at separating many high dimensional datasets [1].

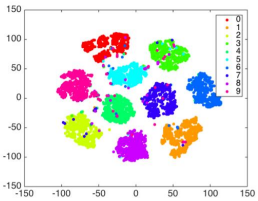

Figure 1: MNIST digits clustering very nicely using unsupervise t-SNE approach

t-SNE is not well suited for on-line processing of additional data points and the results of t-SNE are not easily interpretable in terms of meaningful aspects of the original data. We introduce salient-SNE, a parametric t-SNE approach [2], to address these challenges.

Materials

Dataset #1: MNIST - 50,000 hand written digits ranging from 0-9.

Figure 2: MNIST digits datasets

Methods

Step 1: Pre-train a 1 hidden layer network

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{m} e^{z_k}}, L(z_j) = y \cdot -\log(\sigma(z_j))$$

$$J_{\text{softmax}} = \sum_{j=0}^{n} L(z_j)$$

Step 2: Fine tune using t-SNE loss function

$$P(i|j) = \frac{e^{\frac{-||x_i - x_j||^2}{2\sigma_i^2}}}{\sum_{i \neq k} e^{\frac{-||x_i - x_k||^2}{2\sigma_i^2}}}, \quad P_{i,j} = P(i|j) + P(j|i)$$

$$Q(i|j) = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum_{i \neq k}(1 + ||y_i - y_k||^2)^{-1}}$$

$$J_{\text{t-SNE}} = KL(P||Q) = \sum_{i \neq j} P_{i,j} \cdot \frac{P_{i,j}}{Q_{i,j}}$$

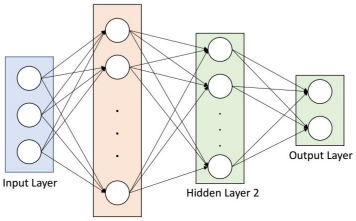

Figure 3: salient-SNE network architecture

Step 3: Interpret the salient-SNE model

$$\vec{v}_{a,b} = a - b, \quad \textit{salient-SNE}(a \to b) = \frac{\partial Y}{\partial X}\Big|_a \cdot \vec{v}_{a,b}$$

Results

$\vec{v}_{a,b}$ is defined using a, any source image, and b, the central image of any target class (Figure 4). The saliency map identifies the changes to the input that are indicative of the output. Transforming a 1 to 2 (Figure 5. top left) requires down weighting a vertical line of pixels at the center of the image while transforming a 0 to 6 (Figure 5. bottom right) requires up-weighting pixels in the center of the image.

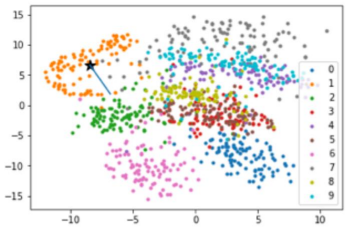

Figure 4: salient-SNE clusters

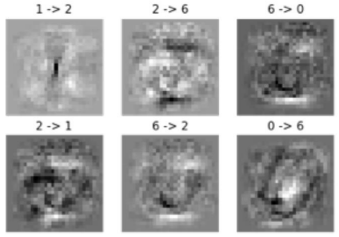

Figure 5: Saliency map between clusters

salient-SNE effectively clusters MNIST digits so that similar images are near each other in the mapped 2 dimensional space. The 0, 1, 2, 6, 7 and 8 clusters are separated from all other clusters. As expected, the 9 and 4 clusters are next to each other and the 3 and 5 clusters are next to each other (see Figure 4).

Training

We noticed a large improvement in the model performance from tuning the hyperparameters. Optimal parameters: regularization $\lambda - 0.2$, hidden layers - 2, hidden units - 200,10. We performed 10000 steps of pretraining, using the 10 hidden units as classification logits before switching to the t-SNE loss.

Future Work

We plan to improve the saliency maps by:

- Reduce noise using ensemble models
- Utilizing integrated gradients

We will apply salient-SNE to single cell RNAseq data [3].

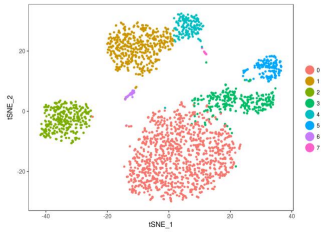

Figure 6: Single cell RNAseq effectively clusters using t-SNE

References

[1] L.J.P. van der Maaten et al. Visualizing High-Dimensional Data Using t-SNE. JMLR 2008.

[2] L.J.P. van der Maaten. Learning a Parametric Embedding by Preserving Local Structure. JMLR 2009.

[3] Macosko et al. Highly Parallel Expression Profiling. Cell 2015.

Stanford University