Machine Learning applied to Planetary Sciences

PTYS 595B/495B

Leon Palafox

https://leonpalafox.github.io/MLClass/

Manifold Learning

- As we have more dimensions, visualization of the datasets become difficult.
- Manifold learning techniques are used mostly for analyzing the data.
- Nevertheless, these are still important to find interesting relationships in the data.

T-SNE

t-distributed stochastic neighbor embedding

http://everynoise.com/engenremap.html

http://cs.stanford.edu/people/karpathy/tsnejs/csvdemo.html

T-SNE theory

- Given two spaces X and Y, where:
  - Dimensionality of X >> Dimensionality of Y (2-3)
- T-SNE calculates probability distributions over X and Y, in which similar objects are near.

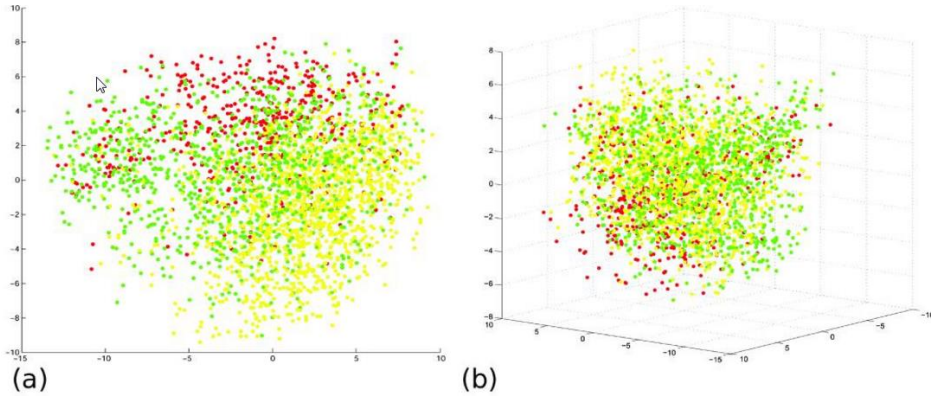
$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2)}, \qquad q_{ij} = \frac{(1+\|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k\neq i}(1+\|\mathbf{y}_k - \mathbf{y}_i\|^2)^{-1}}$$

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

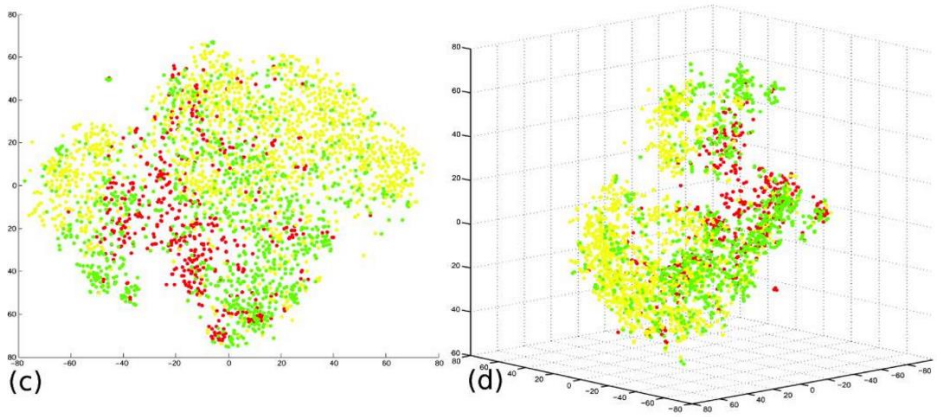
Applications

- Cancer Prognosis

PCA

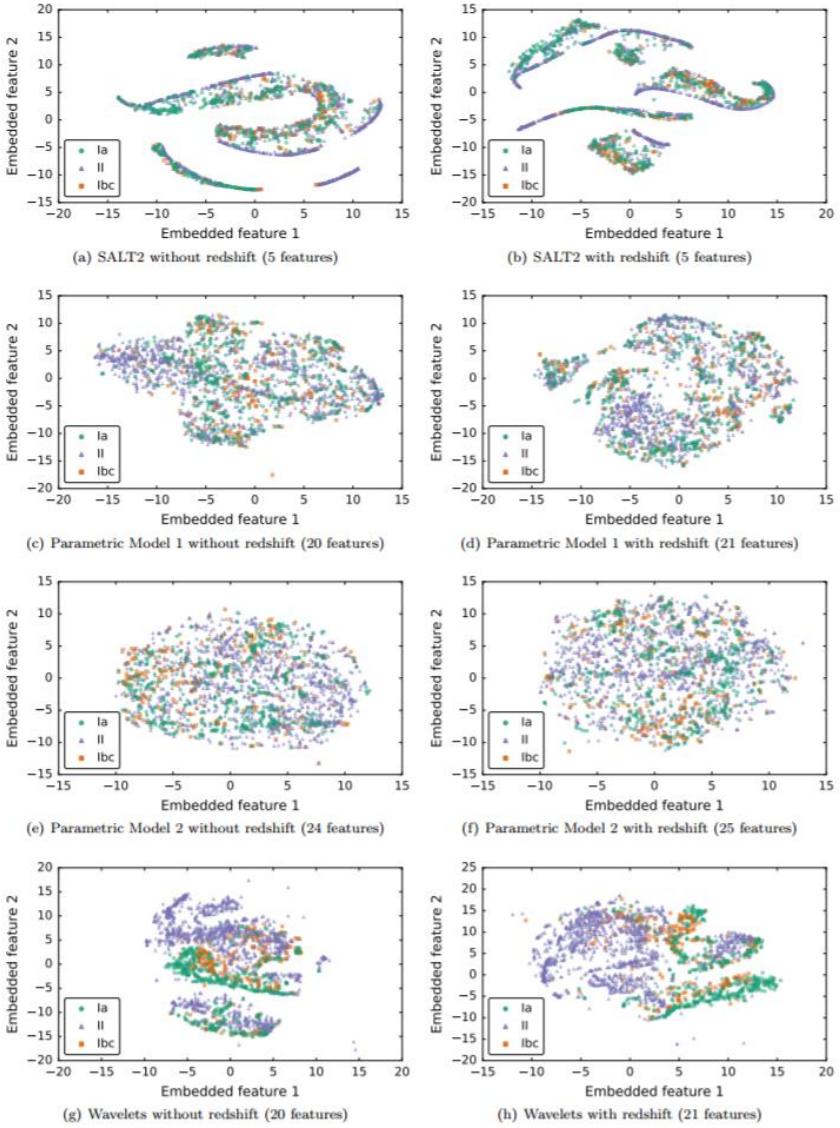


T-SNE



Jamieson, Andrew R., et al. "Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE." *Medical physics* 37.1 (2010): 339-351.

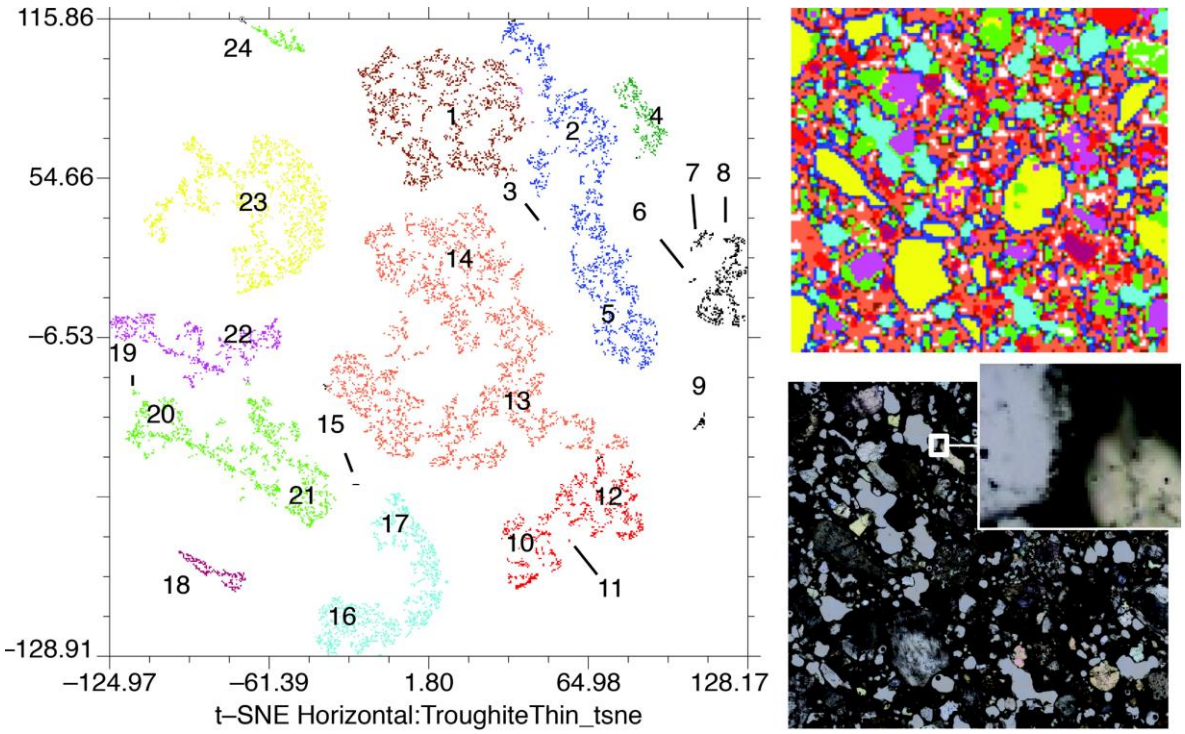
Applications



(a) SALT2 without redshift (5 features)
(b) SALT2 with redshift (5 features)
(c) Parametric Model 1 without redshift (20 features)
(d) Parametric Model 1 with redshift (21 features)
(e) Parametric Model 2 without redshift (24 features)
(f) Parametric Model 2 with redshift (25 features)
(g) Wavelets without redshift (20 features)
(h) Wavelets with redshift (21 features)

Lochner, Michelle, et al. "Photometric Supernova Classification with Machine Learning." *arXiv preprint arXiv:1603.00882* (2016).

Applications



Thompson, David R., et al. "Automating X-ray Fluorescence Analysis for Rapid Astrobiology Surveys." *Astrobiology* 15.11 (2015): 961-976.

Matlab/Python

- sklearn.manifold.TSNE
  - N_components: 2/3
  - Perplexity (how big is the Gaussian neighborhood)
- https://lvdmaaten.github.io/drtoolbox/
  - Has a ton of different tools aside from TSNE