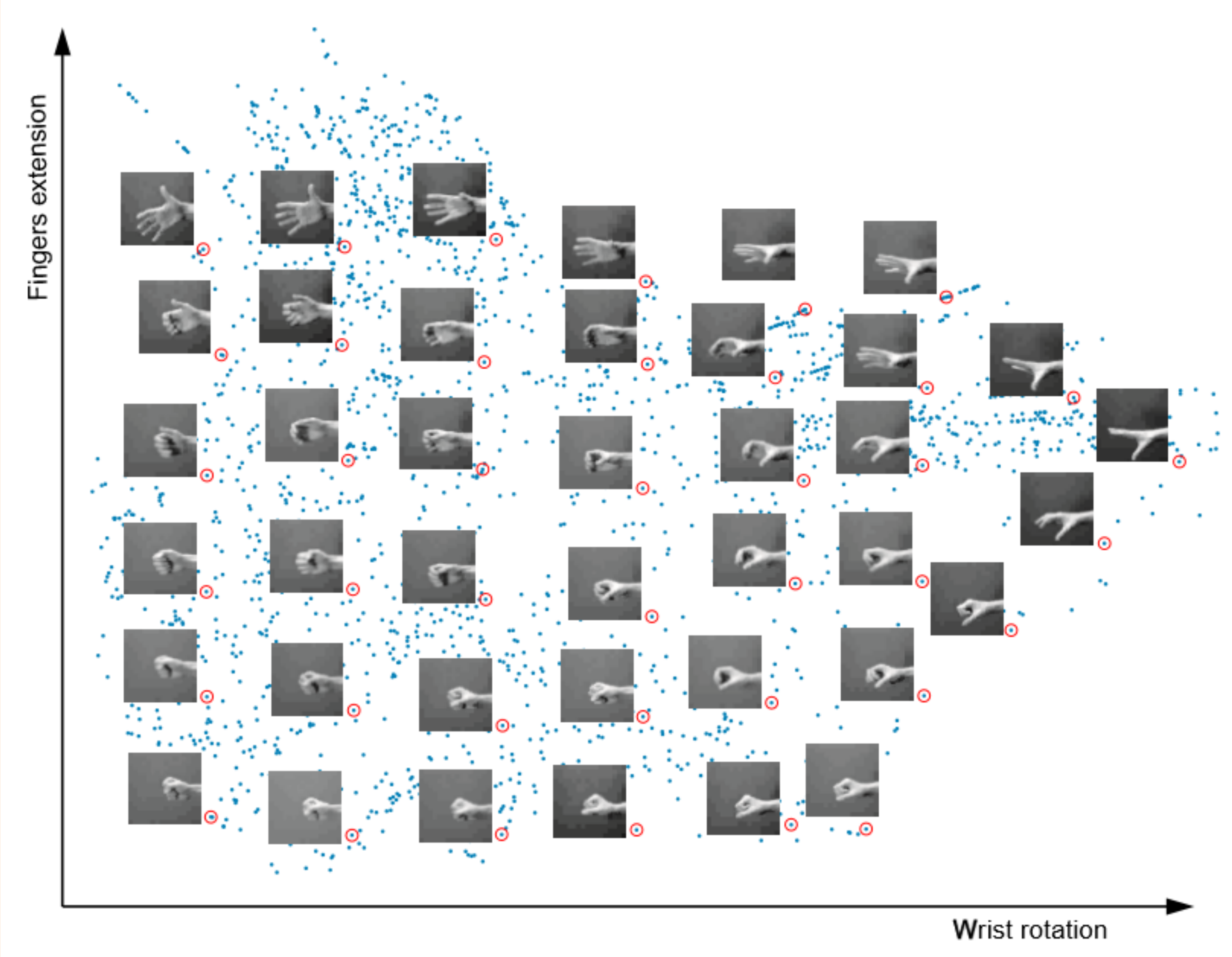
CPSC 340: Machine Learning and Data Mining

Multi-Dimensional Scaling

Bonus Slides

ISOMAP on Hand Images



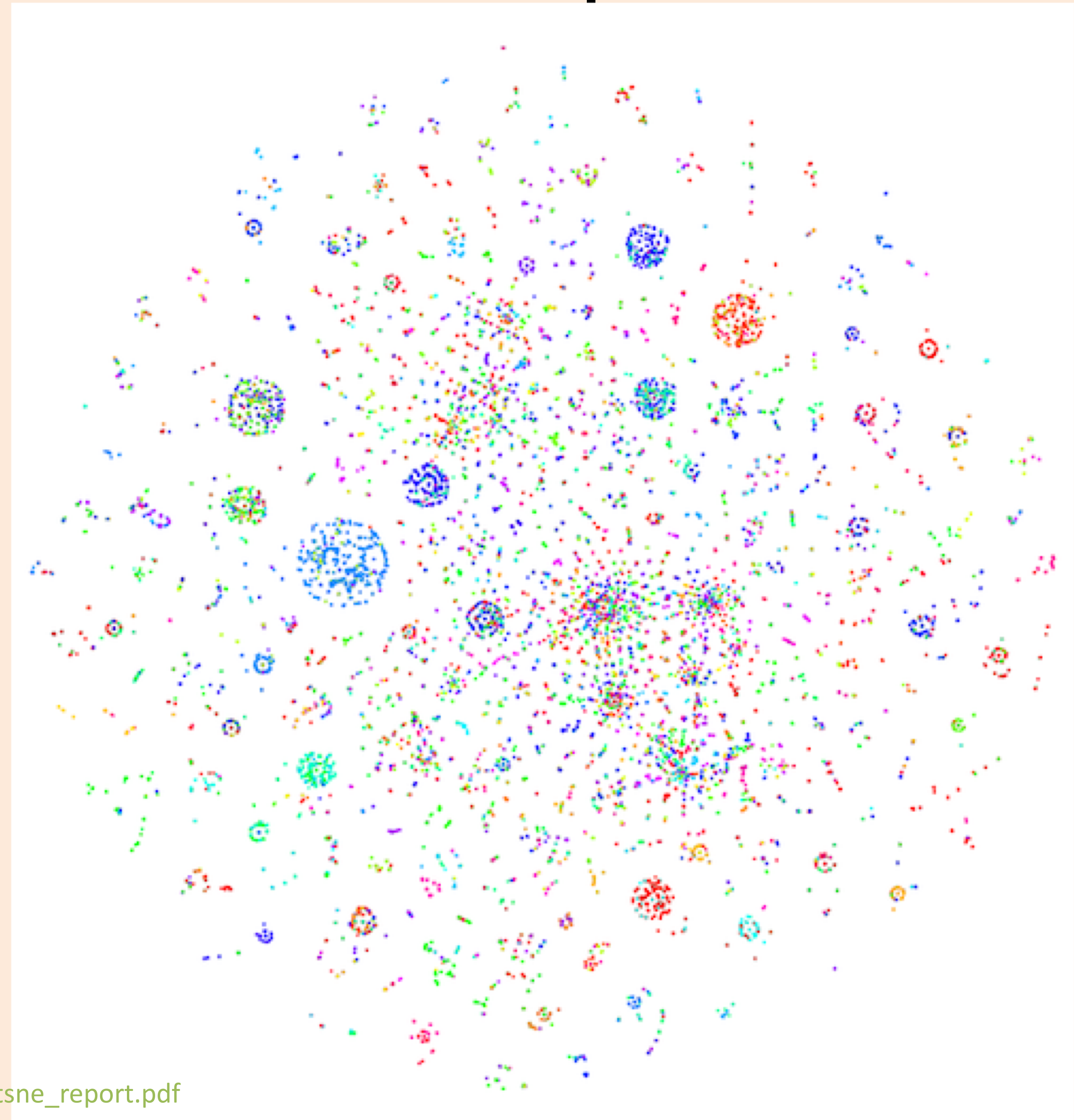
- Related method is “local linear embedding”.

http://wearables.cc.gatech.edu/paper_of_week/isomap.pdf

t-Distributed Stochastic Neighbour Embedding

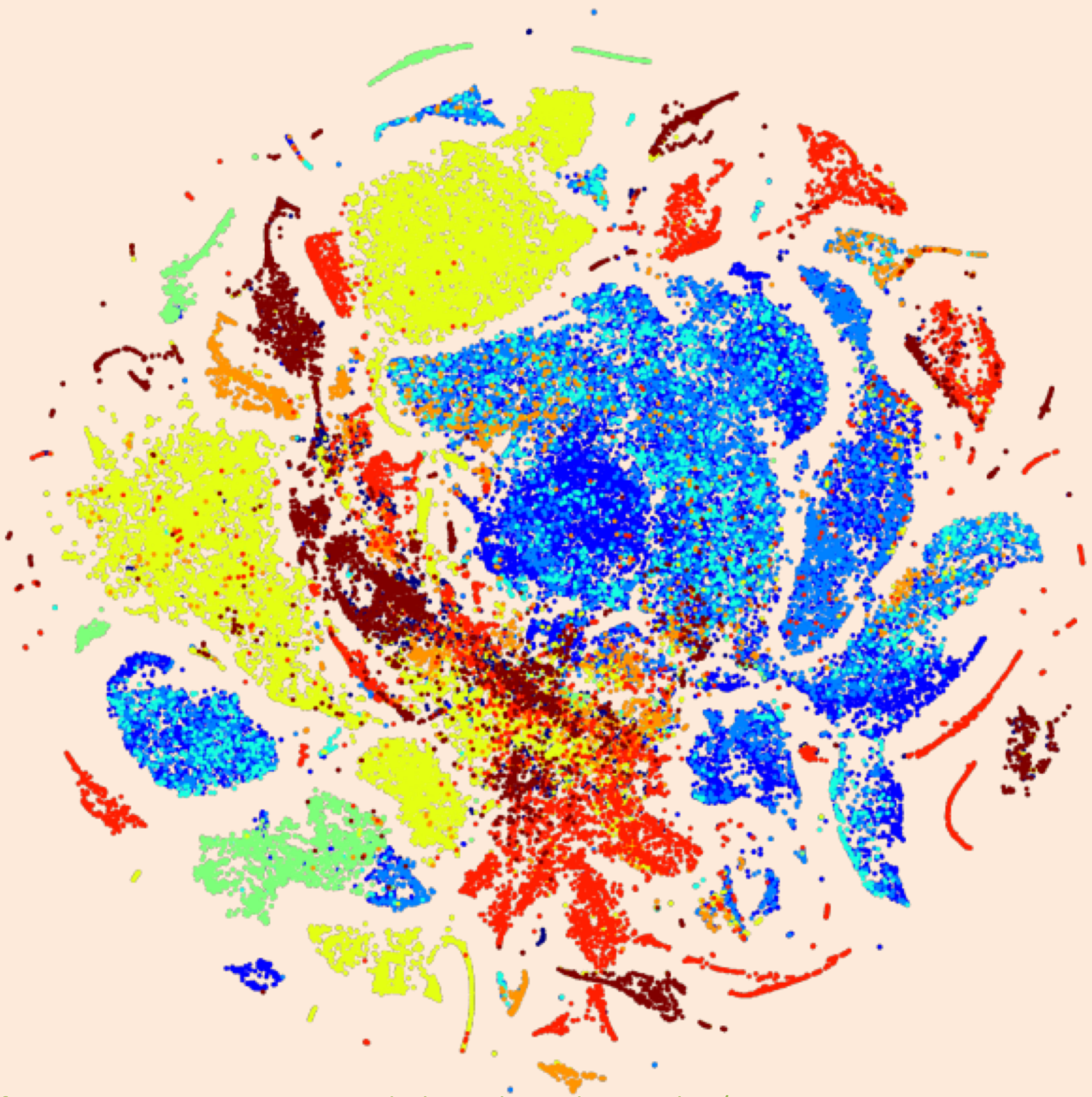
- t-SNE is a special case of MDS (specific d_1, d_2, and d_3 choices):
  - d_1: for each x_i, compute probability that each x_j is a 'neighbour'.
    - Computation is similar to k-means++, but most weight to close points (Gaussian).
    - Doesn't require explicit graph.
  - d_2: for each z_i, compute probability that each z_j is a 'neighbour'.
    - Similar to above, but uses student's t (grows really slowly with distance).
    - Avoids 'crowding', because you have a huge range that large distances can fill.
  - d_3: Compare x_i and z_i using an entropy-like measure:
    - How much 'randomness' is in probabilities of x_i if you know the z_i (and vice versa)?
- Interactive demo: https://distill.pub/2016/misread-tsne

t-SNE on Wikipedia Articles



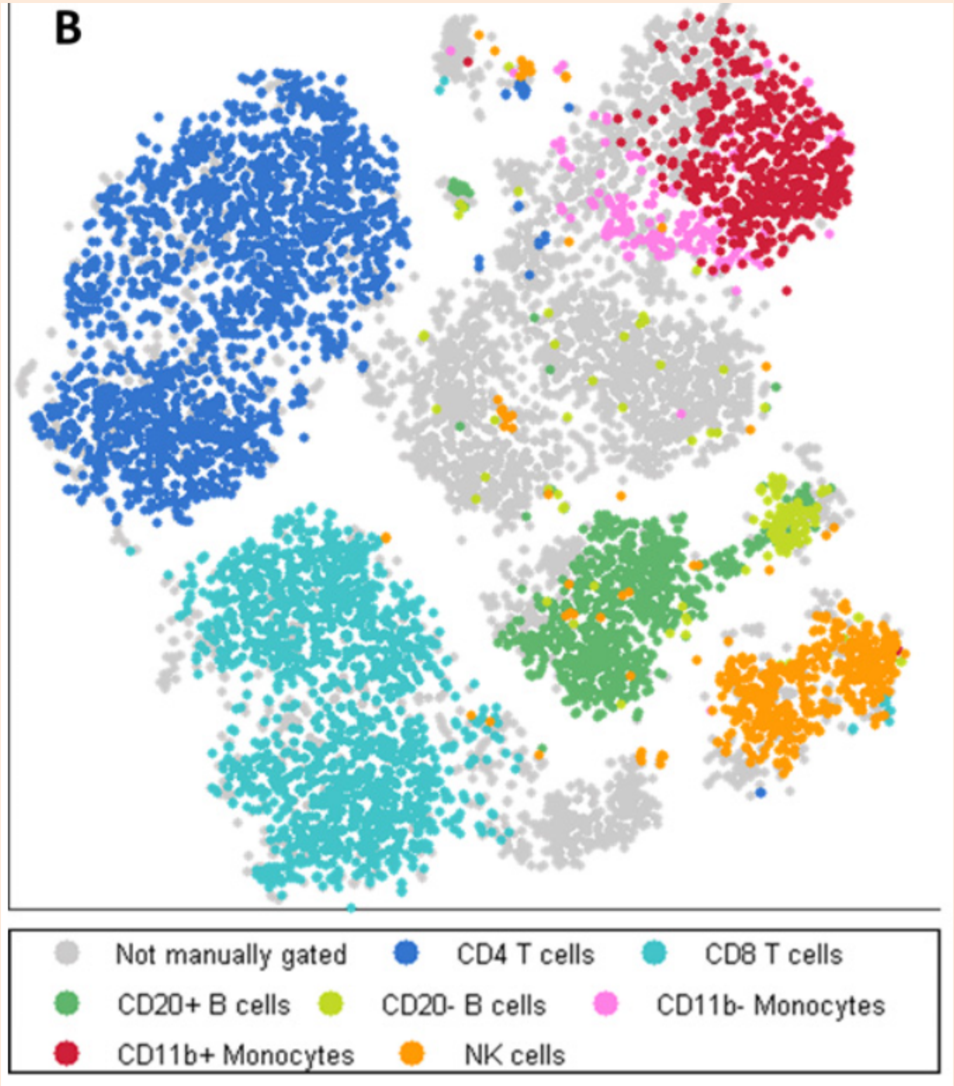
http://jasneetsabharwal.com/assets/files/wiki_tsne_report.pdf

t-SNE on Product Features



http://blog.kaggle.com/2015/06/09/otto-product-classification-winners-interview-2nd-place-alexander-guschin/

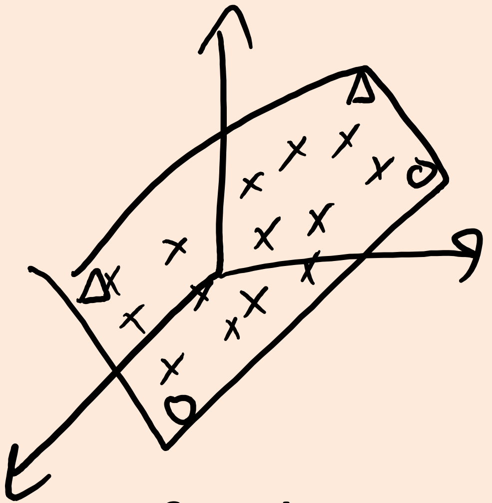
t-SNE on Leukemia Heterogeneity

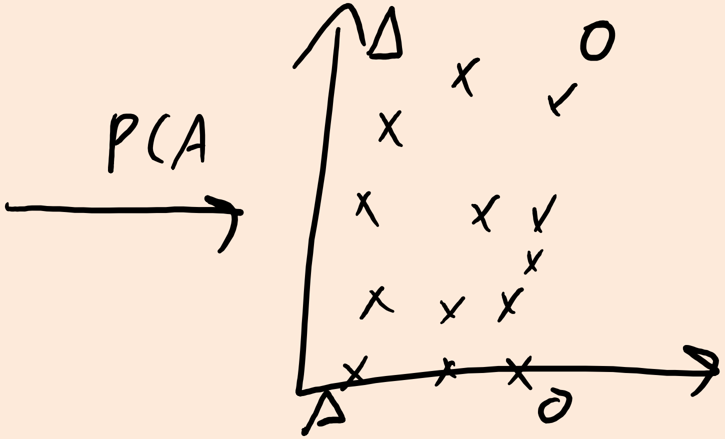


http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4076922/

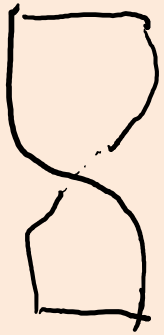
Does t-SNE always outperform PCA?

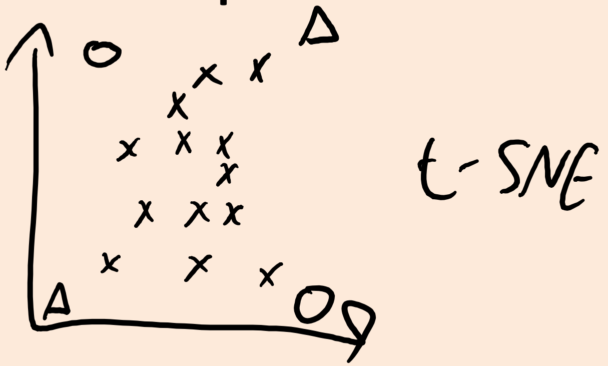
- Consider 3D data living on a 2D hyper-plane:





- PCA can perfectly capture the low-dimensional structure.
- T-SNE can capture the local structure, but can “twist” the plane.
  – It doesn’t try to get long distances correct.





Graph Drawing

- A closely-related topic to MDS is graph drawing:
  - Given a graph, how should we display it?
  - Lots of interesting methods: https://en.wikipedia.org/wiki/Graph_drawing

Bonus Slide: Multivariate Chain Rule

- Recall the univariate chain rule: $\frac{d}{dw}[f(g(w))] = f'(g(w))\,g'(w)$

- The multivariate chain rule: $\underbrace{\nabla[f(g(w))]}_{d\times 1} = \underbrace{f'(g(w))}_{1\times 1}\,\underbrace{\nabla g(w)}_{d\times 1}$

- Example:

$$\nabla\left[\tfrac{1}{2}(w^Tx_i - y_i)^2\right]$$

$$= \nabla[f(g(w))]$$

with $g(w) = w^Tx_i - y_i$ $\longrightarrow$ $\nabla g(w) = x_i$

and $f(r_i) = \frac{1}{2}r_i^2$ $\longrightarrow$ $f'(r_i) = r_i$

$$\nabla[f(g(w))] = r_i\, x_i$$

$$= (w^Tx_i - y_i)x_i$$

Bonus Slide: Multivariate Chain Rule for MDS

- General MDS formulation:

$$\underset{Z \in \mathbb{R}^{n \times k}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=i+1}^{n} g(d_1(x_i, x_j), d_2(z_i, z_j))$$

- Using multivariate chain rule we have:

$$\nabla_{z_i} g(d_1(x_i, x_j), d_2(z_i, z_j)) = g'(d_1(x_i, x_j), d_2(z_i, z_j)) \nabla_{z_i} d_2(z_i, z_j)$$

- Example: If $d_1(x_i, x_j) = \|x_i - x_j\|$ and $d_2(z_i, z_j) = \|z_i - z_j\|$ and $g(d_1, d_2) = \frac{1}{2}(d_1 - d_2)^2$

$$\nabla_{z_i} g(d_1(x_i, x_j), d_2(z_i, z_j)) = \underbrace{-(d_1(x_i, x_j) - d_2(z_i, z_j))}_{\substack{g'(d_1, d_2) \\ \text{(move distances closer)}}} \underbrace{\left[- \frac{(z_i - z_j)}{2\|z_i - z_j\|} \right]}_{\substack{\nabla_{z_i} d_2(z_i, z_j) \\ \text{(how distance changes in } z\text{-space)}}}$$

Assuming $z_i \neq z_j$