



ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ

Faculty of Mathematics and Computer Science  
Department of Natural Language Processing

## Redukcja wymiarowości

PCA, (t-)SNE, UMAP

14-05-2019

Kuba Pokrywka



## Redukcja wymiarowości

---

przestrzeń wysokowymiarowa  $\rightarrow$  przestrzeń niskowymiarowa

- ▶ redukcja niepotrzebnych informacji
- ▶ wizualizacja

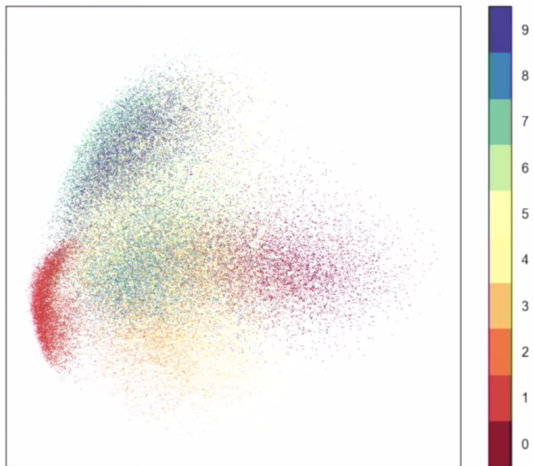


## mnist

---

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

# mnist





## Redukcja wymiarowości

---

### Typy algorytmów

- ▶ faktoryzacja macierzy (PCA, word2vec, GloVe)
- ▶ grafy sąsiedztwa (SNE, t-SNE, Isomap, UMAP)



1901- Karl Pearson

Zadanie Mamy przestrzeń wektorową. Znaleźć jej przekształcenie liniowe (jej obrót) tak, żeby pierwszy wymiar nowej przestrzeni miał możliwie największą wariancję. Następnie drugi wymiar nowej przestrzeni miał możliwie największą wariancję, itp.

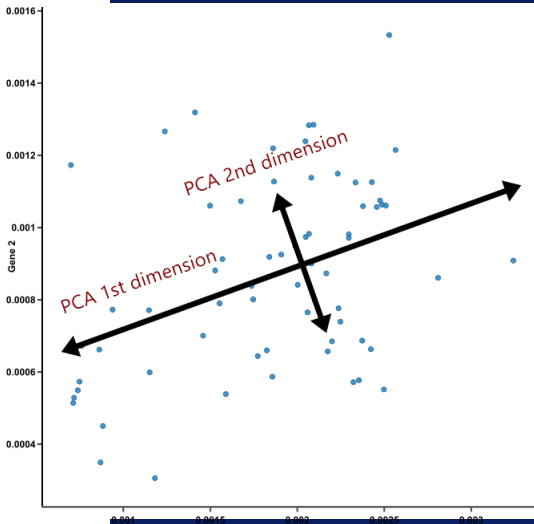


## PCA

---

Dostajemy przestrzeń, w której pierwsza składowa wyjaśnia np. 40 procent wariancji, druga 30 procent, itp.

# PCA







## Algorytm

---

- ▶ skalowanie lub normalizacja danych
- ▶ przesunięcie do środka układu współrzędnych
- ▶ znalezienie prostej przechodzącej przez środek układu, że po rzutowaniu danych na nią dostajemy największą wariancję
- ▶ znalezienie drugiej ortogonalnej prostej o tej samej właściwości
- ▶ itd aż skończą się wymiary



## Jak znaleźć

---

Macierz kowariancji/korelacji

$$A = U W V^t$$
$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} \\ a_{21} & a_{22} & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1j} \\ u_{21} & u_{22} & \cdots & u_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{i1} & u_{i2} & \cdots & u_{ij} \end{bmatrix} \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{jj} \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & \cdots & v_{j1} \\ v_{12} & v_{22} & \cdots & v_{j2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1j} & v_{2j} & \cdots & v_{jj} \end{bmatrix}$$

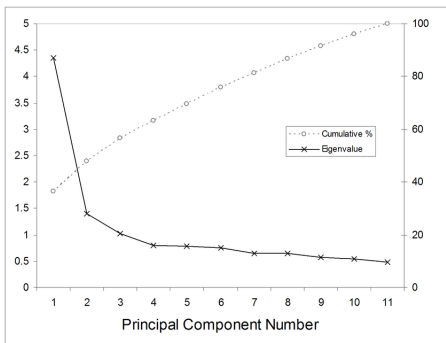


## Charakterystyka

---

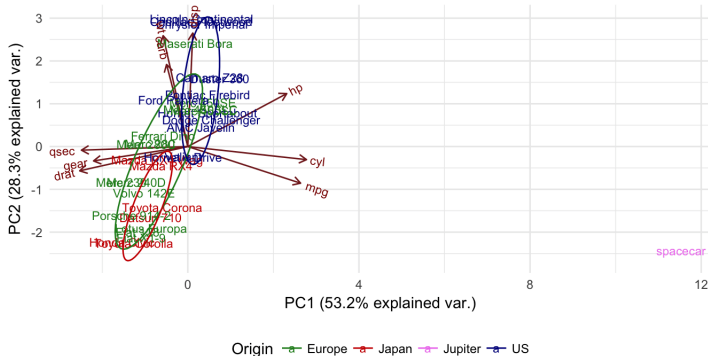
- ▶ szybko się oblicza- zwykle mnożenie przez macierz
- ▶ możliwość inferencji
- ▶ daje przestrzeń o tej samej wymiarowości
- ▶ dobrze wyjaśnialna
- ▶ jedyny parametr to tylko transformacja na danych wejściowych- skalowanie/normalizacja
- ▶ zachowuje globalną strukturę danych

Macierz kowariancji/korelacji



# wizualizacja

PCA of mtcars dataset, with extra sample added





## Stochastic Neighbor Embedding

---

Neighbor Embedding (Hinton and Roweis, 2002)

Zadanie: Znaleźć przestrzeń w zadanej (niskiej) wymiarowości, tak żeby rozkłady przestrzeni wysokowymiarowej i niskowymiarowej były jak zbliżone.

---

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)},$$

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)}.$$

$$C = \sum_i KL(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}},$$





---

$$\text{Perp}(P_i) = 2^{H(P_i)},$$

Shannon entropy 5-50

$$H(P_i) = - \sum_j p_{j|i} \log_2 p_{j|i}.$$

$$\frac{\delta C}{\delta y_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j).$$



## t-SNE

---

Visualizing Data using t-SNE (2008)

Laurens van der Maaten, Geoffrey Hinton



## t-SNE

---

- ▶ używa symetrycznej funkcji prawdopodobieństwa
- ▶ używa rozkładu t-Studenta zamiast Gaussowskiego w przestrzeni niskowymiarowej

---

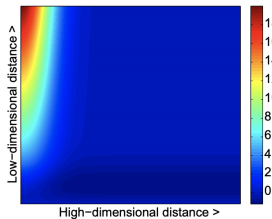
$$q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq l} \exp(-\|y_k - y_l\|^2)},$$

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

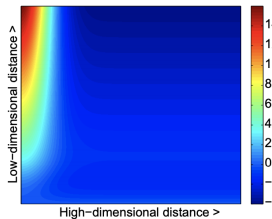
---

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}}$$

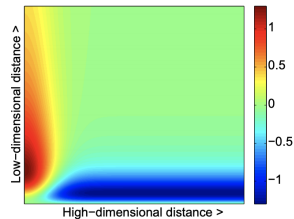
$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$



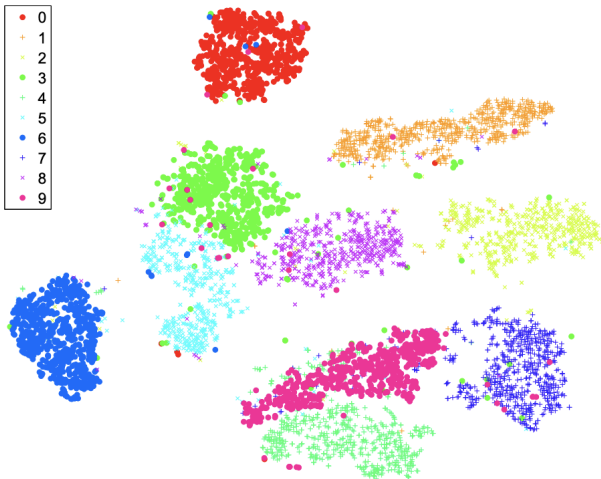
(a) Gradient of SNE.



(b) Gradient of UNI-SNE.



(c) Gradient of t-SNE.



(a) Visualization by t-SNE.





## t-SNE Charakterystyka

---

- ▶ nie działa w inferencji
- ▶ jest wolne i dobrze działa dla 2D, 3D
- ▶ nie zachowuje struktury globalnej i może pokazać zależności tam, gdzie ich nie ma



UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction (2018)

Leland McInnes, John Healy, James Melville

Uzasadnienie teoretyczne w artykule.



0-simplex



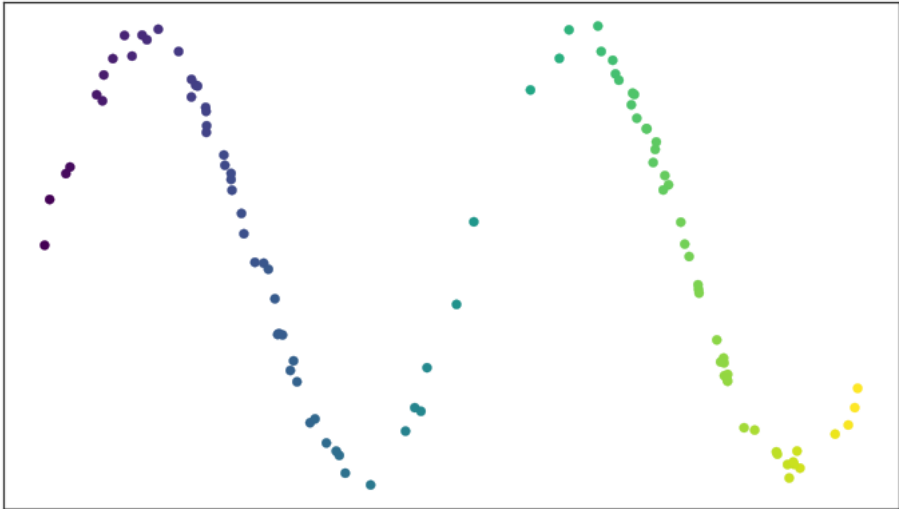
1-simplex

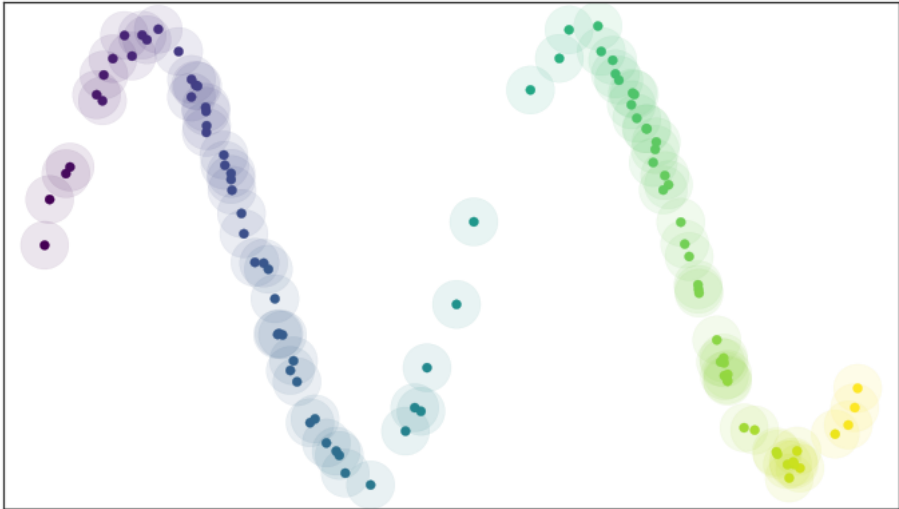


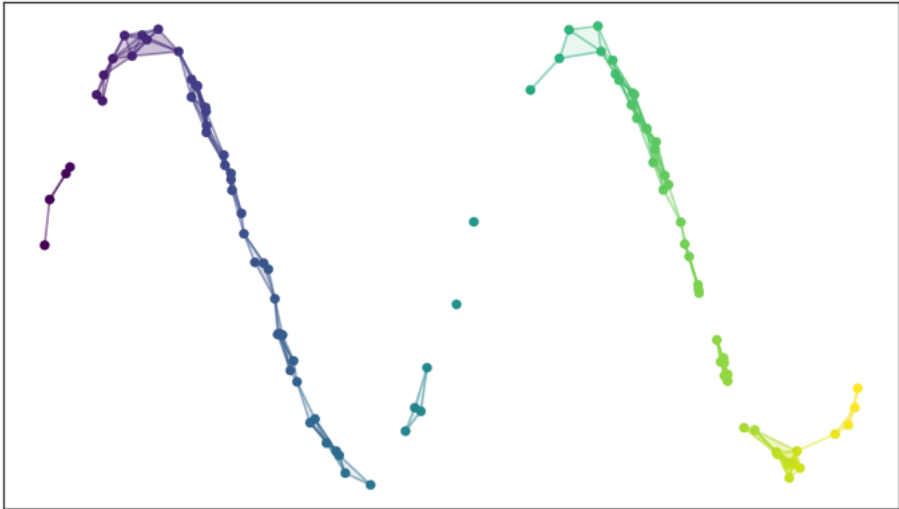
2-simplex

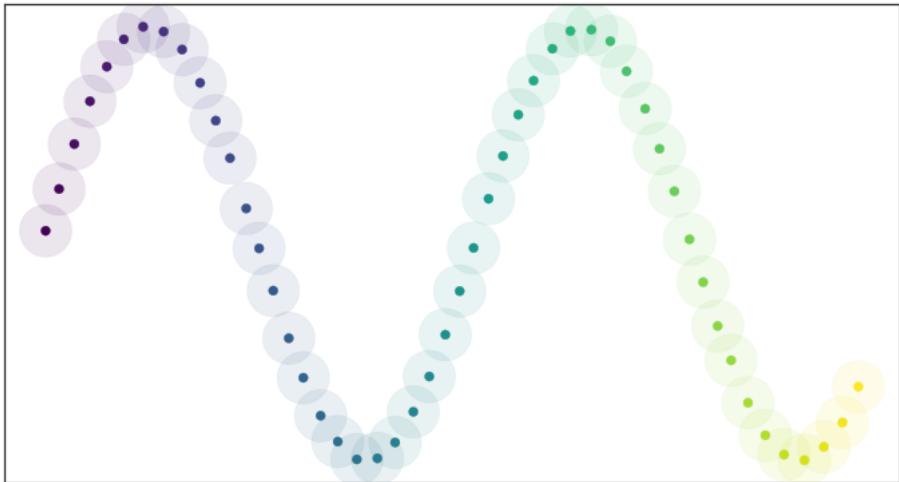


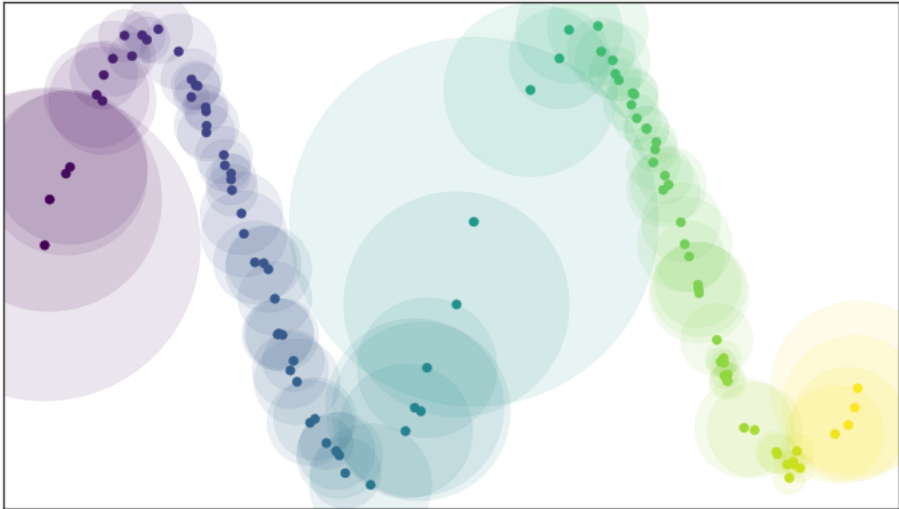
3-simplex



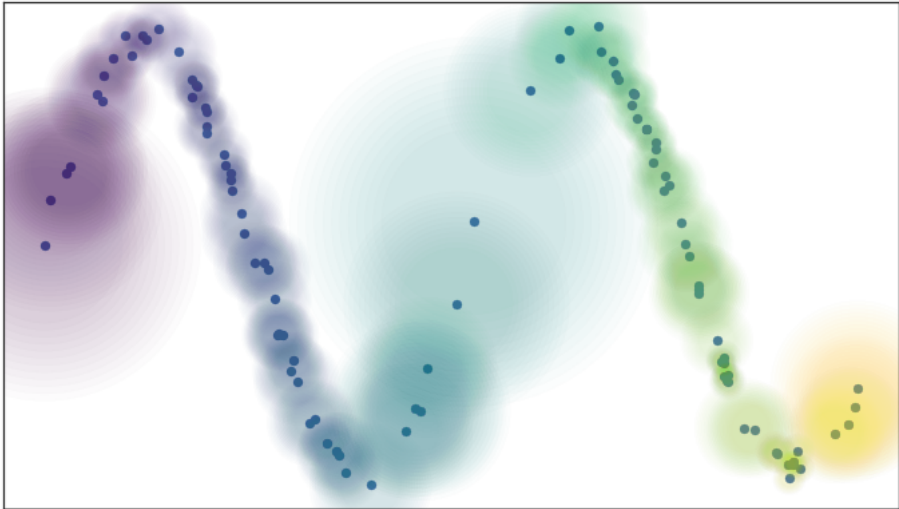


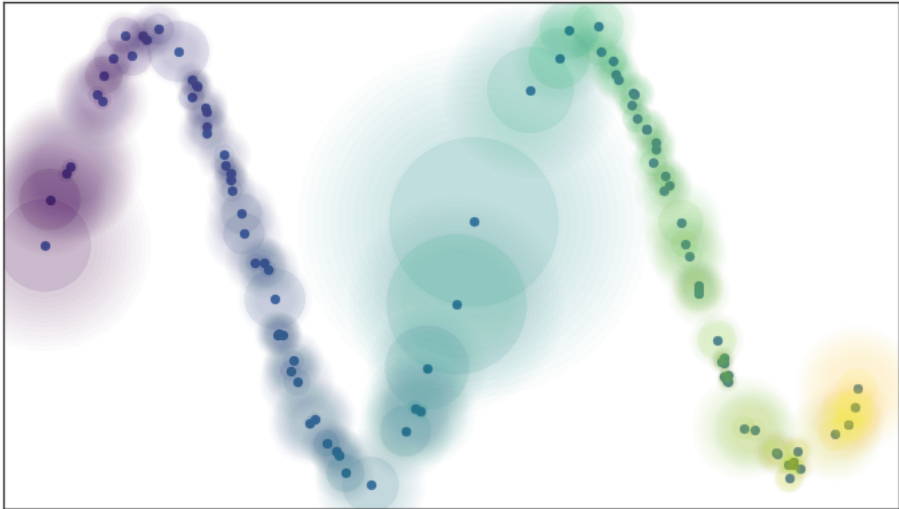


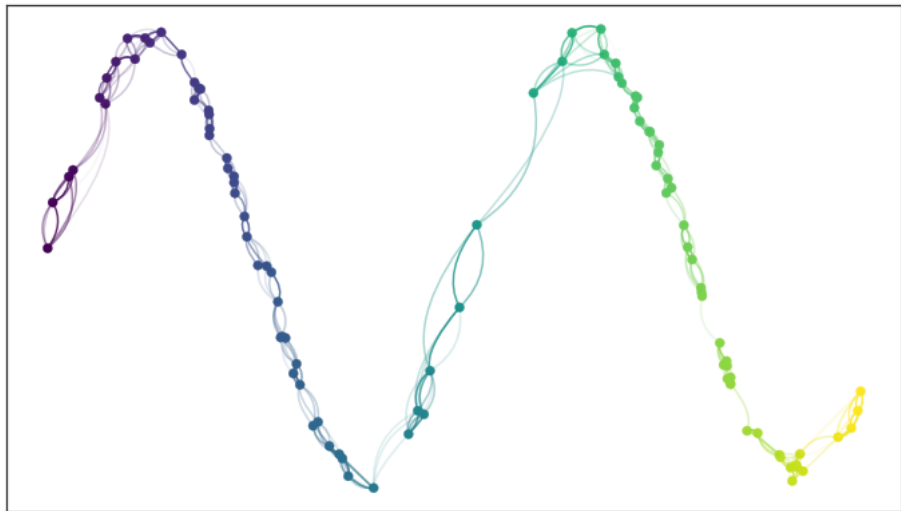


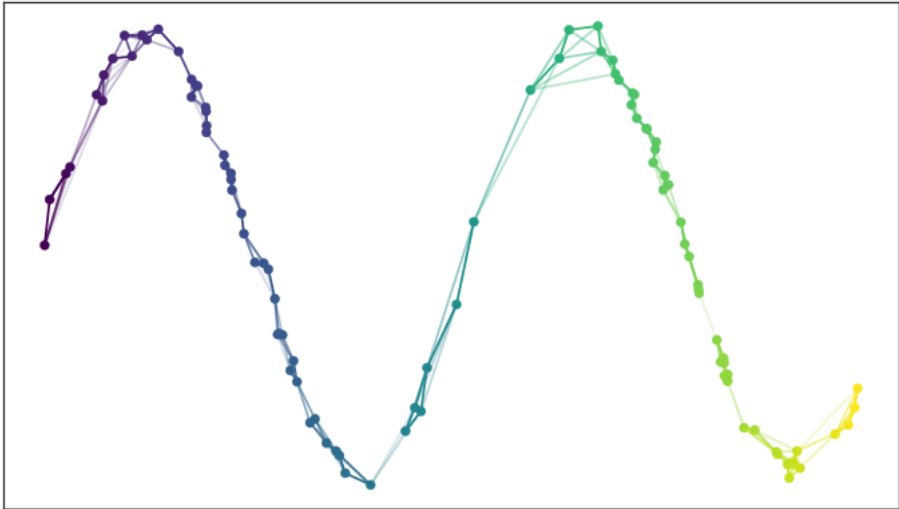












$$\sum_{e \in E} w_h(e) \log\left(\frac{w_h(e)}{w_l(e)}\right) + (1 - w_h(e)) \log\left(\frac{1 - w_h(e)}{1 - w_l(e)}\right)$$

---

**Algorithm 1** UMAP algorithm

---

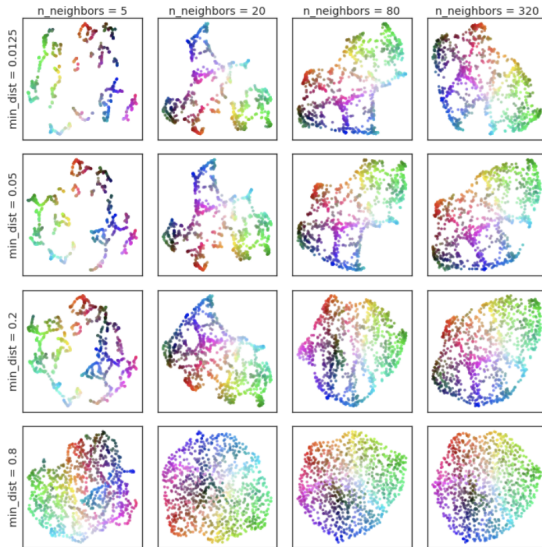
```
function UMAP( $X, n, d, \text{min-dist}, \text{n-epochs}$ )  
  for all  $x \in X$  do  
     $\text{fs-set}[x] \leftarrow \text{LOCALFUZZYSIMPLICIALSET}(X, x, n)$   
   $\text{top-rep} \leftarrow \bigcup_{x \in X} \text{fs-set}[x]$   $\triangleright$  We recommend the probabilistic t-conorm  
   $Y \leftarrow \text{SPECTRALEMBEDDING}(\text{top-rep}, d)$   
   $Y \leftarrow \text{OPTIMIZEEMBEDDING}(\text{top-rep}, Y, \text{min-dist}, \text{n-epochs})$   
  return  $Y$ 
```

---

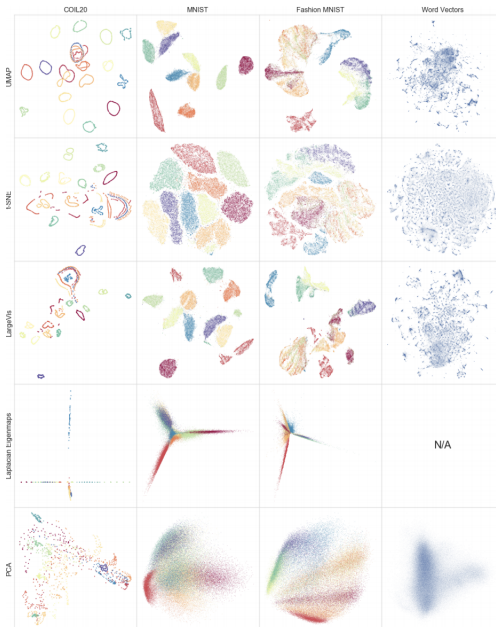
### 4.3 Hyper-parameters

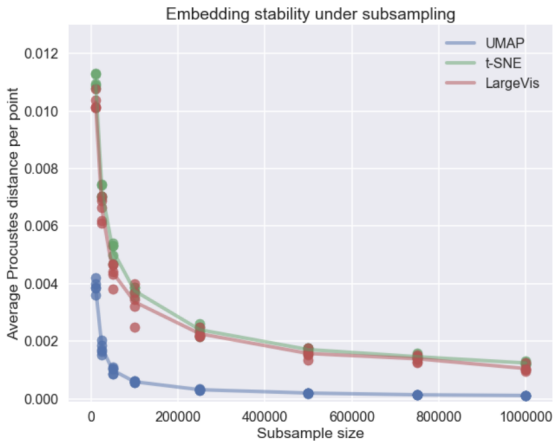
As described in Algorithm 1, the UMAP algorithm takes four hyper-parameters:

1.  $n$ , the number of neighbors to consider when approximating the local metric;
2.  $d$ , the target embedding dimension;
3. min-dist, the desired separation between close points in the embedding space; and
4. n-epochs, the number of training epochs to use when optimizing the low dimensional representation.











## Podsumowanie

---

- ▶ zachowuje globalną strukturę
- ▶ szybsze
- ▶ stabilniejsze
- ▶ działa dla wymiarów  $> 3$
- ▶ działa w inferencji
- ▶ może być wykorzystane w supervised metric learning