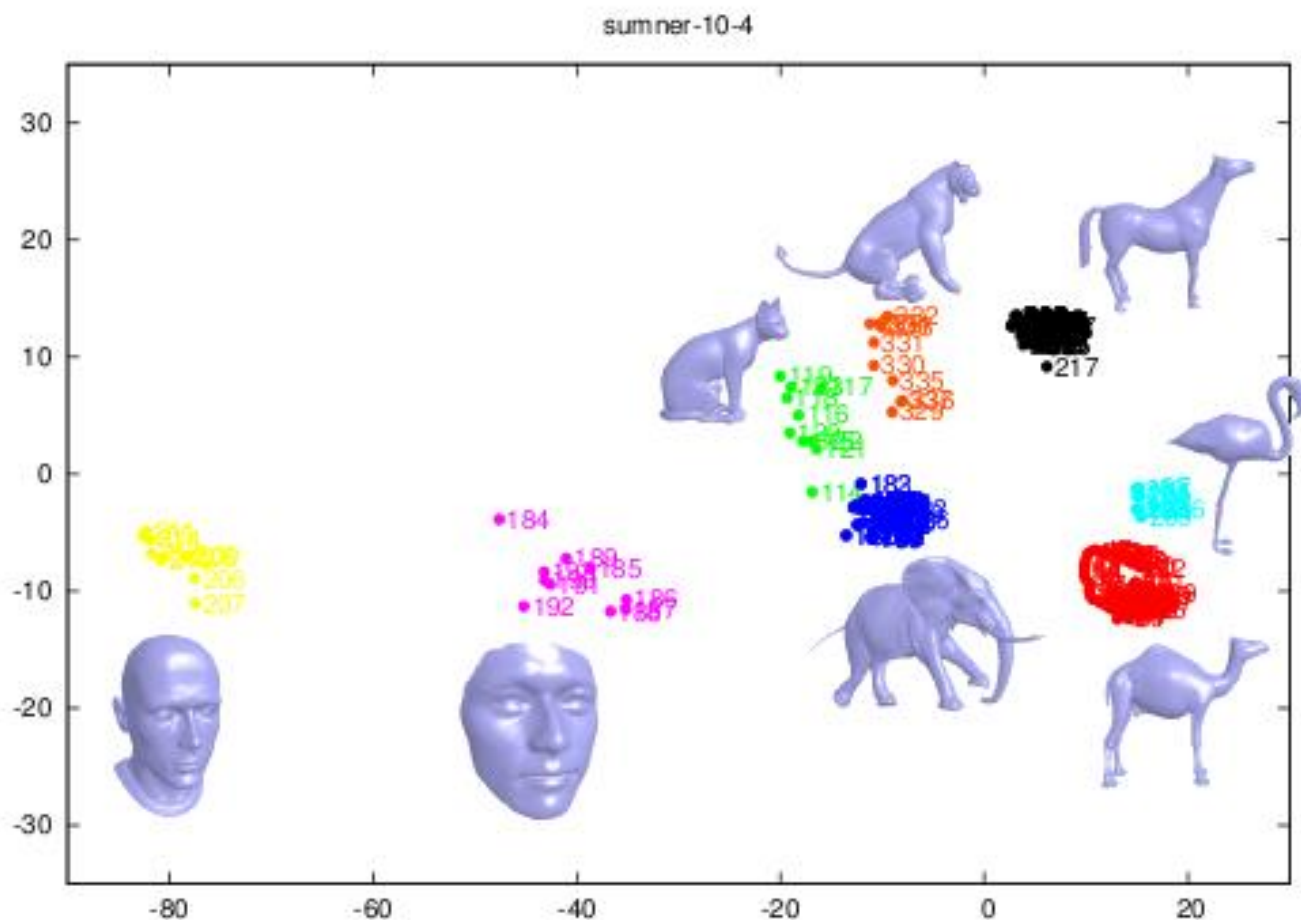


Manifold embeddings

ВИЗУАЛИЗАЦИЯ

Задача визуализации состоит в отображении объектов в 2х- или 3хмерное пространство с сохранением отношений между ними.



MULTIDIMENSIONAL SCALING (MDS)

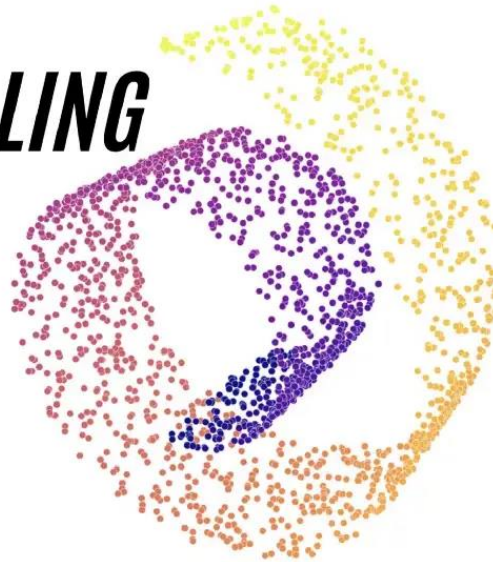
Идея метода – *минимизация квадратов отклонений между исходными и новыми попарными расстояниями:*

$$\sum_{i \neq j}^l (\rho(x_i, x_j) - \rho(z_i, z_j))^2 \rightarrow \min_{z_1, \dots, z_l}$$

ISOMAP

- Isomap – комбинация нескольких алгоритмов для нелинейного снижения размерности, сохраняющего локальную структуру данных.

***UNROLLING
THE
SWISS
ROLL***



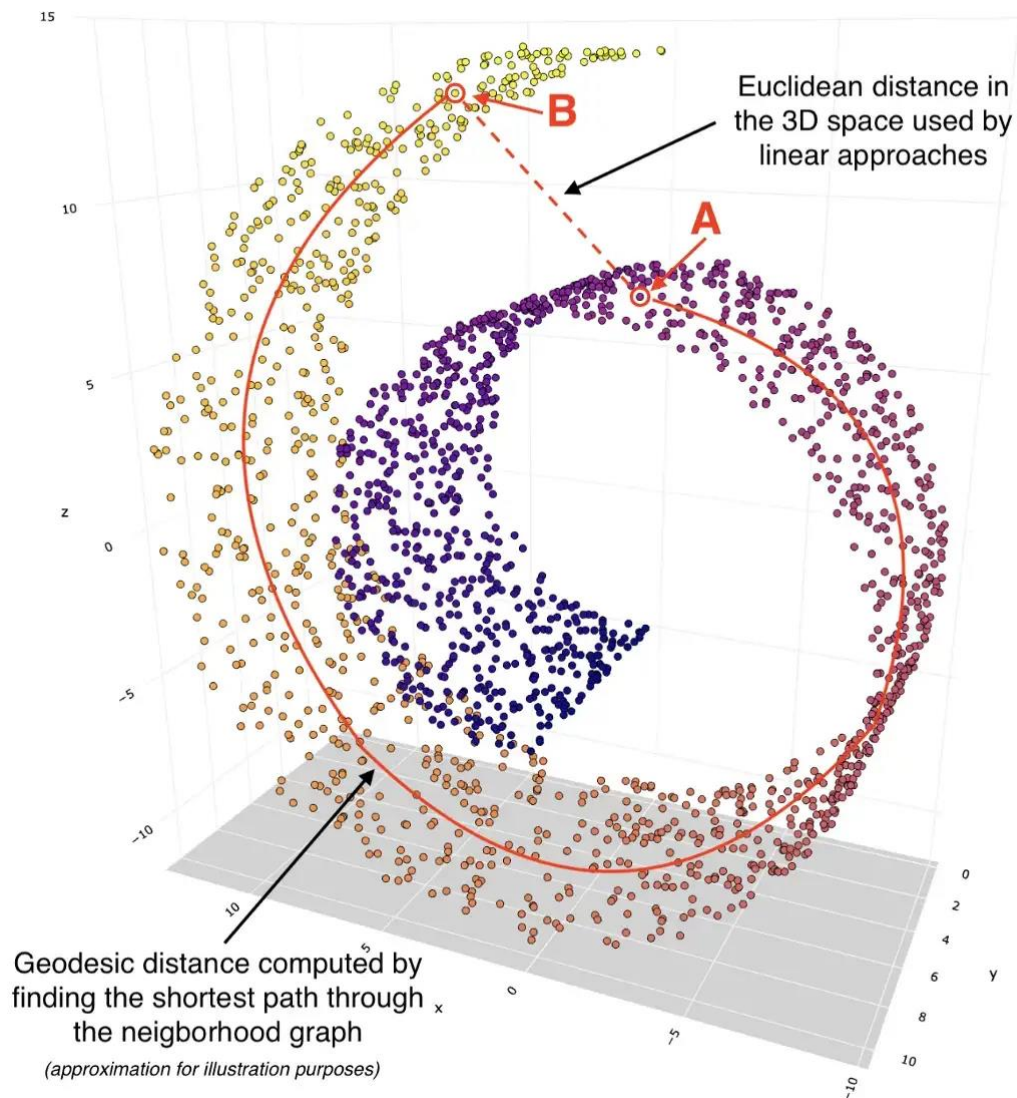
<https://towardsdatascience.com/isomap-embedding-an-awesome-approach-to-non-linear-dimensionality-reduction-fc7efbca47a0>

ISOMAP

Алгоритм:

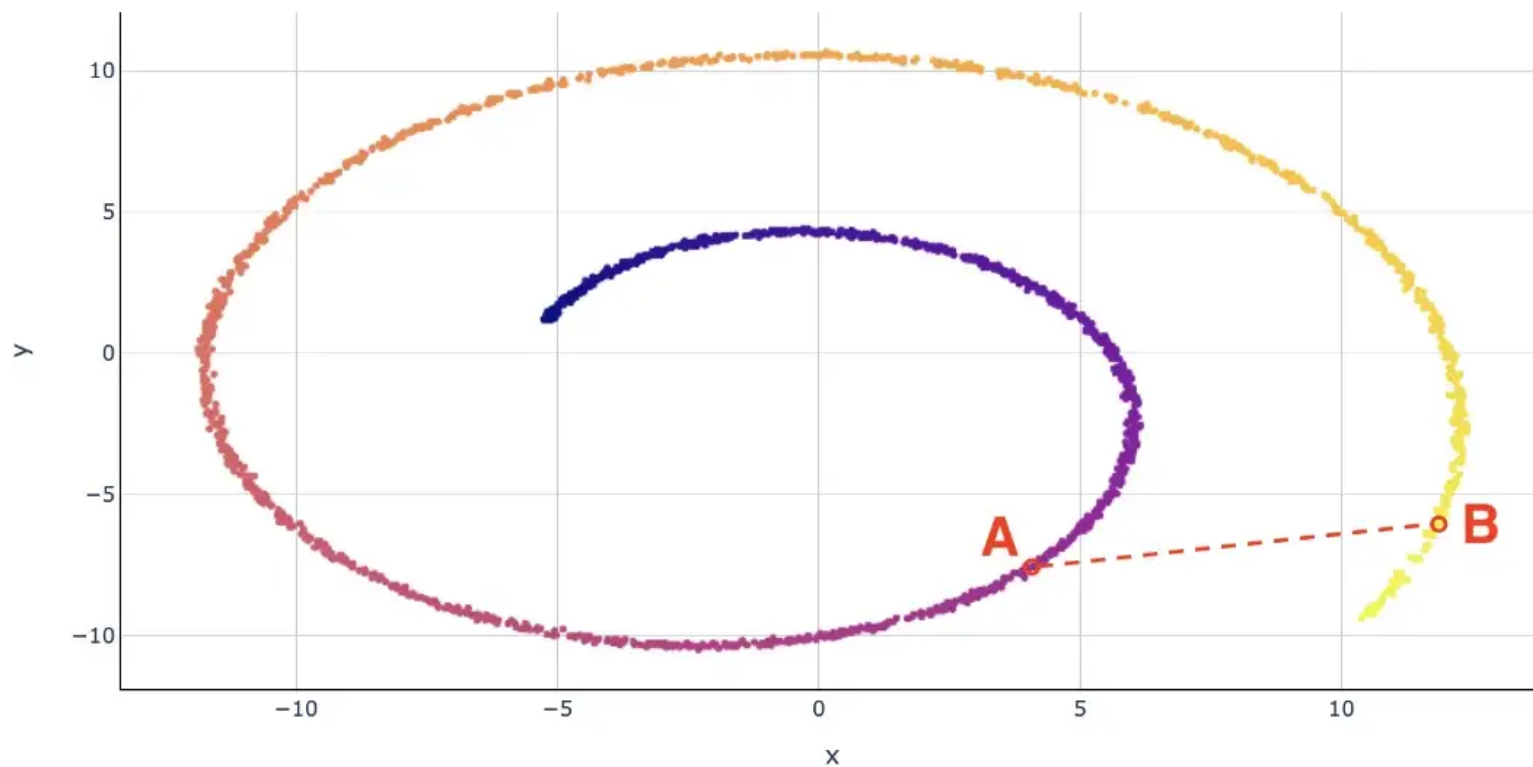
- 1) Используем KNN для вычисления расстояний между ближайшими соседями
- 2) Создаем граф, где только соседи соединены ребрами друг с другом
- 3) Вычисляем кратчайший путь между двумя точками по ребрам графа
- 4) Используем MDS для построения проекции в низкоразмерное пространство.

ISOMAP: ПРИМЕР



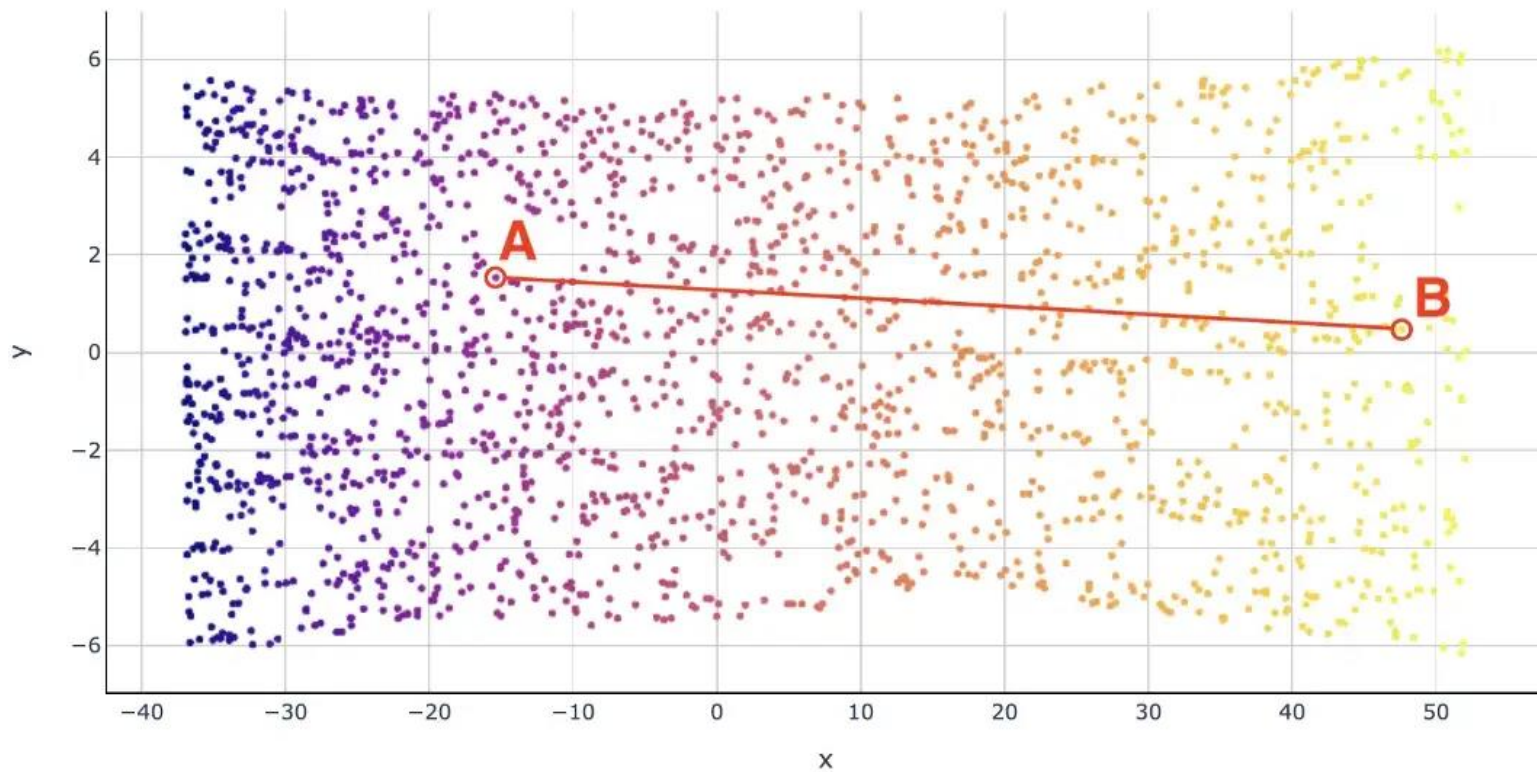
ISOMAP: ПРИМЕР

PCA Transformation



ISOMAP: ПРИМЕР

Isomap transformation



TSNE

t-SNE – t-distributed stochastic neighbor embedding

- *При проекции нам важно не сохранение расстояний между объектами, а сохранение пропорций:*

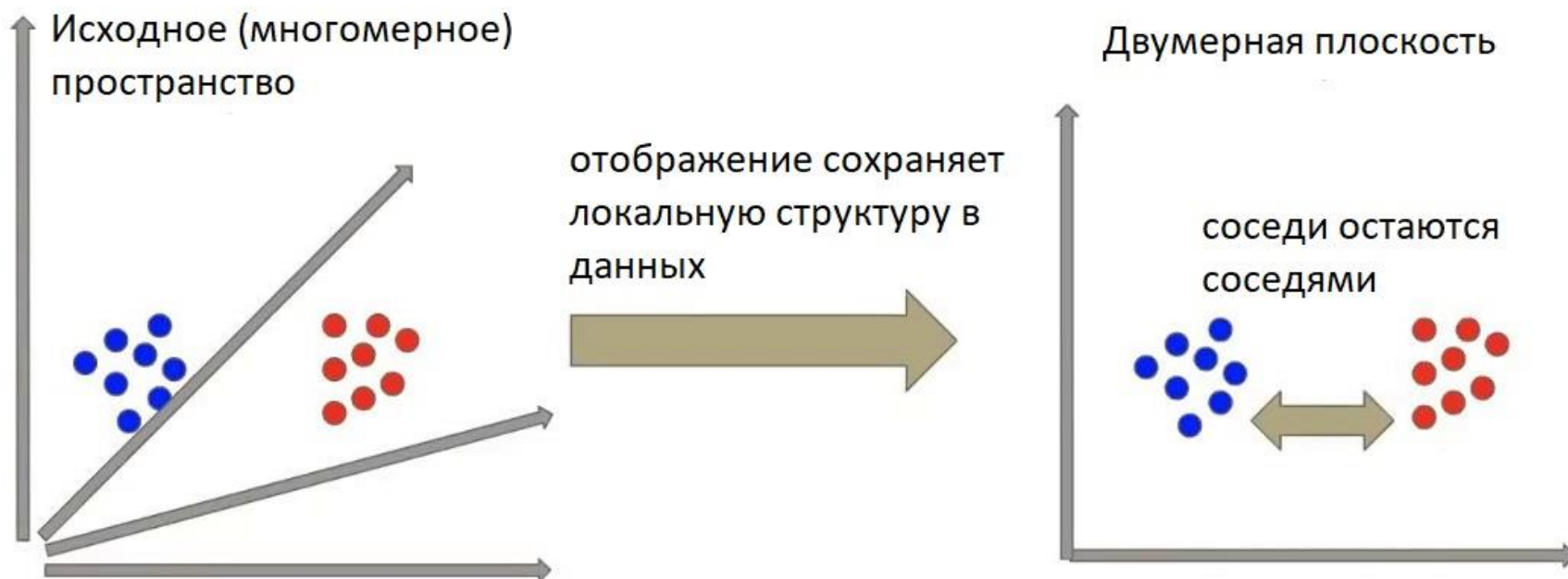
$$\rho(x_1, x_2) = \alpha \rho(x_1, x_3) \Rightarrow \rho(z_1, z_2) = \alpha \rho(z_1, z_3)$$

TSNE

t-SNE – t-distributed stochastic neighbor embedding

- При проекции нам важно не сохранение расстояний между объектами, а сохранение пропорций:

$$\rho(x_1, x_2) = \alpha \rho(x_1, x_3) \Rightarrow \rho(z_1, z_2) = \alpha \rho(z_1, z_3)$$



БЛИЗОСТЬ ОБЪЕКТОВ В ИСХОДНОМ ПРОСТРАНСТВЕ

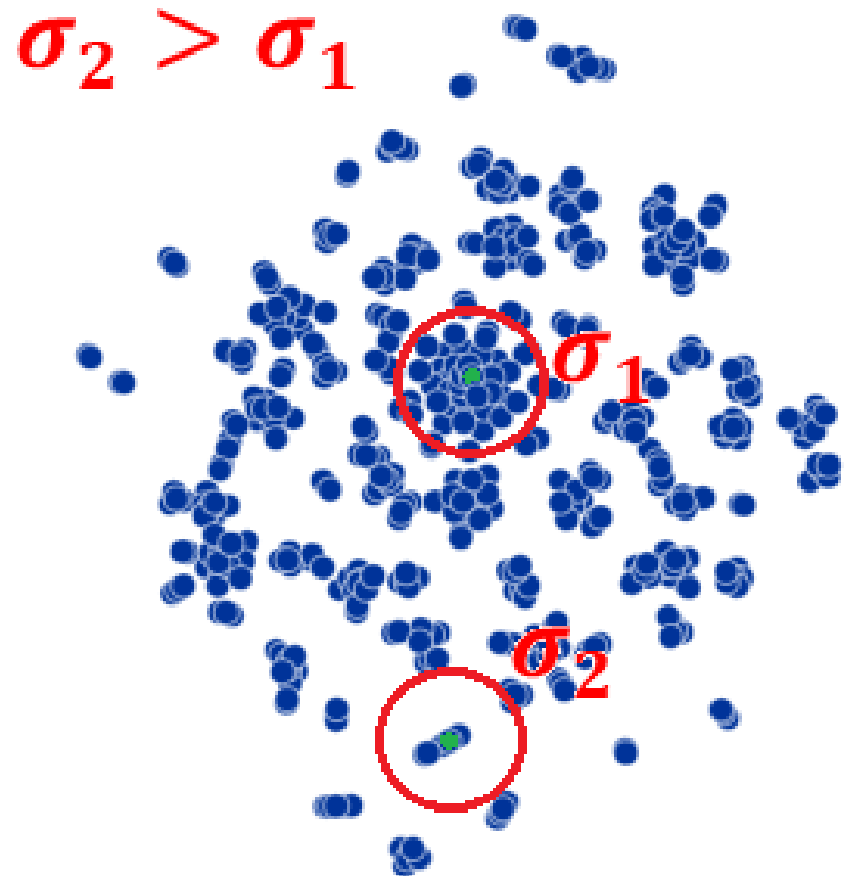
$$p(i|j) = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_j^2)}{\sum_{k \neq j} \exp(-\|x_k - x_j\|^2 / 2\sigma_j^2)}$$

(затем симметризуем $p(i|j)$)

- объекты из окрестности x_j

приближаются нормальным
распределением

- чем кучнее объекты
из этой окрестности,
тем меньше берётся
значение σ_j^2



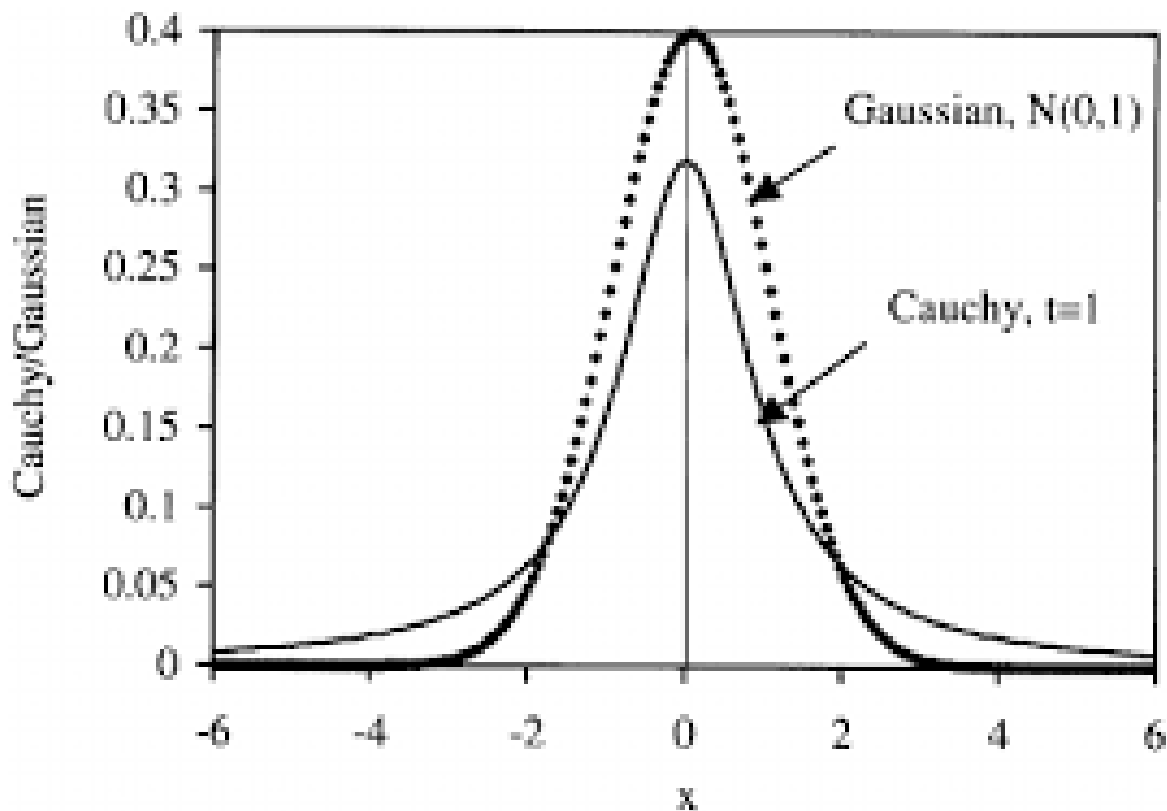
БЛИЗОСТЬ ОБЪЕКТОВ В НОВОМ ПРОСТРАНСТВЕ

- *В пространстве большой размерности можно разместить несколько объектов так, чтобы расстояния между ними были малы, но сохранить это свойство в низкоразмерном пространстве довольно сложно.*
- Будем измерять сходство объектов в новом пространстве с помощью распределения Коши, так как оно не так сильно штрафует за увеличение расстояний между объектами:

$$q_{ij} = \frac{\left(1 + \left\|z_i - z_j\right\|^2\right)^{-1}}{\sum_{k \neq j} \left(1 + \left\|z_k - z_j\right\|^2\right)^{-1}}$$

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И РАСПРЕДЕЛЕНИЕ КОШИ

- Будем измерять сходство объектов в новом пространстве с помощью распределения Коши, так как оно не так сильно штрафует за увеличение расстояний между объектами:



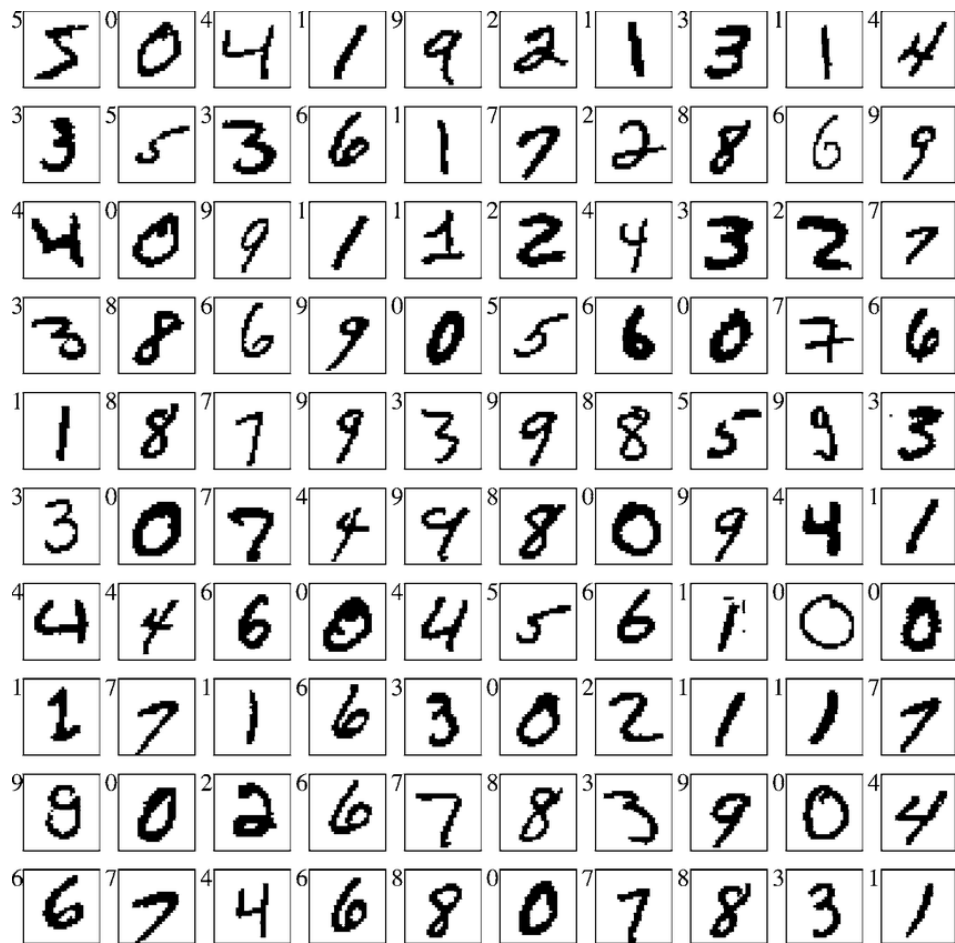
ОБУЧЕНИЕ TSNE

- Для построения проекций z_i объектов x_i будем минимизировать расстояние между исходным и полученным распределениями (минимизируем дивергенцию Кульбака-Лейблера).

$$KL(p||q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \rightarrow \min_{z_1, \dots, z_l}$$

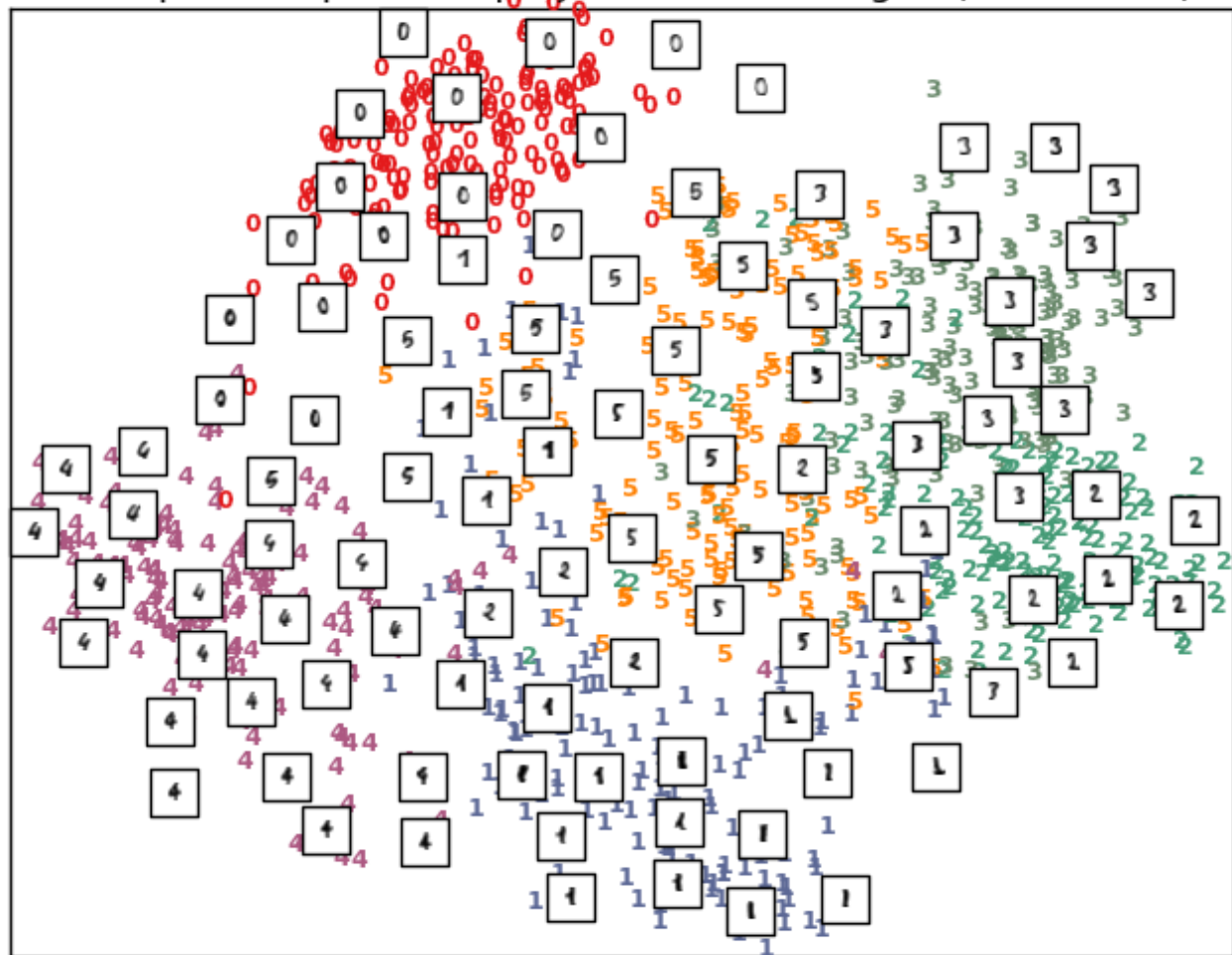
TSNE (ПРИМЕР)

- MNIST – датасет из различных написаний десятичных цифр, где каждая картинка размера 28x28.



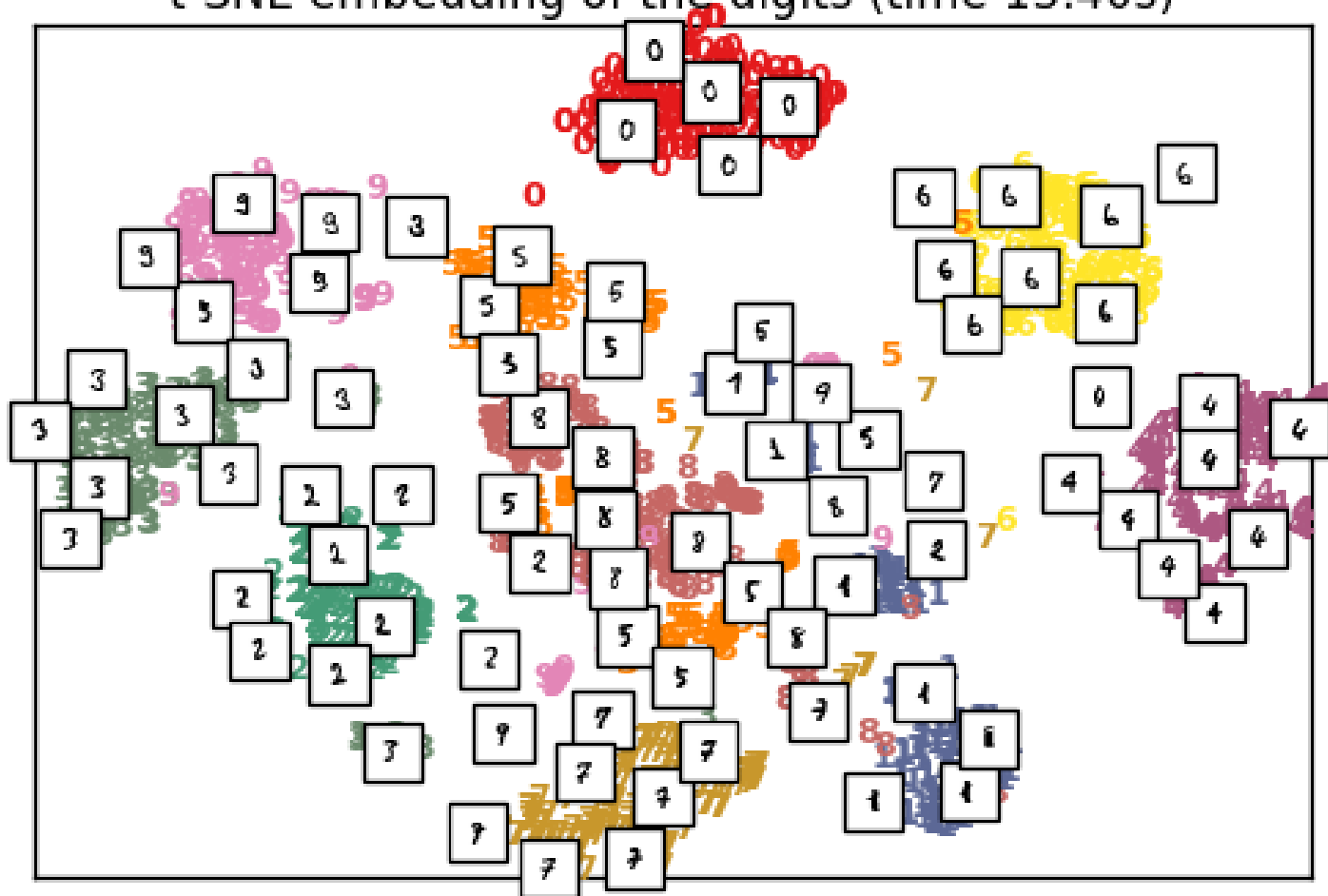
РСА (ПРИМЕР)

Principal Components projection of the digits (time 0.01s)



TSNE (ПРИМЕР)

t-SNE embedding of the digits (time 13.40s)



ВИЗУАЛИЗАЦИЯ PCA И TSNE

<http://projector.tensorflow.org/>