



# ТЕХНОСФЕРА

## Лекция 4 Визуализация результатов кластеризации

Николай Анохин

19 марта 2015 г.

# Краткое содержание предыдущих лекций

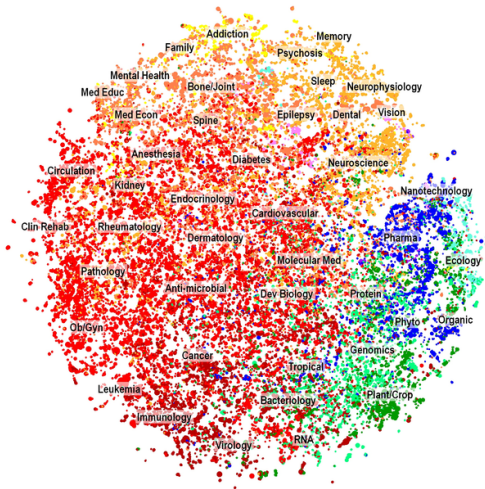
**Дано.**  $N$  обучающих  $D$ -мерных объектов  $\mathbf{x}_i \in \mathcal{X}$ , образующих тренировочный набор данных (training data set)  $X$ .

**Найти.** Модель  $h^*(\mathbf{x})$  из семейства параметрических функций  $H = \{h(\mathbf{x}, \theta) : \mathcal{X} \times \Theta \rightarrow \mathbb{N}\}$ , ставящую в соответствие произвольному  $\mathbf{x} \in \mathcal{X}$  один из  $K$  кластеров так, чтобы объекты внутри одного кластера были похожи, а объекты из разных кластеров различались.

# Краткое содержание предыдущей лекции

Рассмотрели классические алгоритмы кластеризации

1. Смесь гауссовских распределений и k-means
2. Hierarchical Clustering
3. DBSCAN



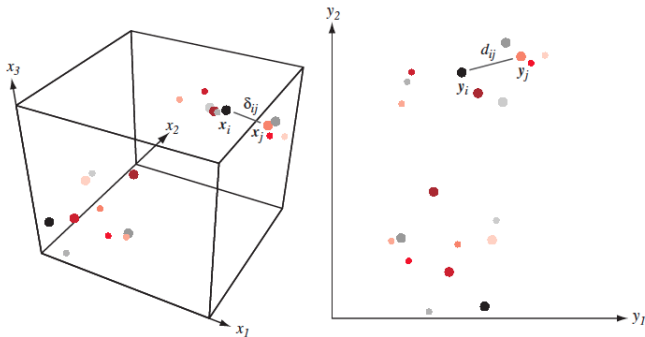
## Multidimensional Scaling

# Идея метода

Перейти в пространство меньшей размерности так, чтобы расстояния между объектами в новом пространстве были подобны расстояниям в исходном пространстве.

# Обозначения

- ▶  $\mathbf{x}_i \in \mathcal{X} \subset R^D$  – объекты в исходном многомерном пространстве
- ▶  $\delta_{ij}$  – расстояние между  $\mathbf{x}_i$  и  $\mathbf{x}_j$
- ▶  $\mathbf{x}_i \in \mathcal{Y} \subset R^E$  – объекты в целевом пространстве ( $E = 2$  или  $E = 3$ )
- ▶  $d_{ij}$  – расстояние между  $\mathbf{y}_i$  и  $\mathbf{y}_j$



# Критерии

Выбираем конфигурацию  $\mathbf{y}_i$ , соответствующую минимуму критерия



$$J_{ee} = \frac{\sum_{i < j} (d_{ij} - \delta_{ij})^2}{\sum_{i < j} \delta_{ij}^2}$$



$$J_{ff} = \sum_{i < j} \frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}^2}$$



$$J_{ef} = \frac{1}{\sum_{i < j} \delta_{ij}} \sum_{i < j} \frac{(d_{ij} - \delta_{ij})^2}{\delta_{ij}}$$

►

$$\nabla_{\mathbf{y}_k} J_{ee} = \frac{2}{\sum_{i < j} \delta_{ij}^2} \sum_{j \neq k} (d_{kj} - \delta_{kj}) \frac{\mathbf{y}_k - \mathbf{y}_j}{d_{kj}}$$

►

$$\nabla_{\mathbf{y}_k} J_{ff} = 2 \sum_{j \neq k} \frac{d_{kj} - \delta_{kj}}{\delta_{kj}^2} \frac{\mathbf{y}_k - \mathbf{y}_j}{d_{kj}}$$

►

$$\nabla_{\mathbf{y}_k} J_{ef} = \frac{2}{\sum_{i < j} \delta_{ij}} \sum_{j \neq k} \frac{d_{kj} - \delta_{kj}}{\delta_{kj}} \frac{\mathbf{y}_k - \mathbf{y}_j}{d_{kj}}$$



T-SNE

Вопросы

