

# Обобщения метрик и основные методы многомерного шкалирования

Александр Сенин

11 марта 2020 г.

# Аксиомы метрики

Формализуем интуитивные представления о расстоянии.

Пусть на некотором множестве  $X$  определена функция  $\rho : X \times X \rightarrow \mathbb{R}$  такая, что  $\forall x, y, z \in X$ :

- ▶  $\rho(x, y) \geq 0$ , причем  $\rho(x, y) = 0 \Leftrightarrow x = y$
- ▶  $\rho(x, y) = \rho(y, x)$
- ▶  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Тогда функция  $\rho$  называется метрикой.

# Применение метрик в анализе данных

Метрические методы для решения задач классификации или регрессии (например, kNN)

Кластеризация, обнаружение выбросов

Понижение размерности, многомерное шкалирование

## Примеры метрик в анализе данных

Расстояние Минковского:  $l_p(x, y) = \left( \sum_{i=1}^d |x^i - y^i|^p \right)^{1/p}$

Частные случаи:

- ▶ Евклидово расстояние при  $p = 2$ .
- ▶ Расстояние городских кварталов при  $p = 1$ .
- ▶ Метрика Чебышева при  $p = \infty : l_\infty(x, y) = \max_{i=1\dots d} |x_i - y_i|$ .

Не является метрикой при  $p < 1$  (нарушается неравенство треугольника).



Рис.: Единичная окружность

# Обобщения метрик

Далеко не все расстояния, используемые в анализе данных являются метриками.

## Псевдометрика

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  $x = y \Rightarrow \rho(x, y) = 0$
4.  $\rho(x, y) = \rho(y, x)$
5.  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Разрешаем равенство нулю не только на совпадающих объектах.

Пример:

$$\rho_M(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)}, S \succeq 0$$

## Пример псевдометрики в анализе данных

Расстояние Махаланобиса:

$$l_2(x, y) = \sqrt{(x - y)^T(x - y)} \longrightarrow \rho_M(x, y) = \sqrt{(x - y)^T S^{-1}(x - y)},$$

где  $S$  — симметричная неотрицательно определенная матрица.

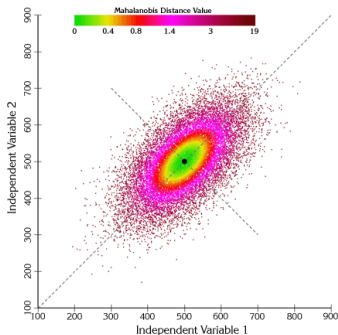
Смысл: для  $S$  можно построить ортонормированный базис из собственных векторов  $\Rightarrow SQ = Q\Lambda \Rightarrow S = Q\Lambda Q^T \Rightarrow S^{-1} = Q\Lambda^{-1}Q^T \Rightarrow \rho_M(x, y) = l_2(\tilde{x}, \tilde{y})$ , где  $\tilde{x} = \Lambda^{-1/2}Q^T x$ .

Пусть  $S = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{N-1} X^T X$  — несмещенная выборочная оценка ковариационной матрицы (выборка предварительно центрирована,  $\bar{x} = 0$ ). Тогда собственные векторы описывают «направления разброса» данных, а собственные значения — степень разброса в направлении соответствующего вектора.

## Пример псевдометрики в анализе данных

Расстояние Махаланобиса  $\rho_M(x, y) = \sqrt{(x - y)^T S^{-1}(x - y)}$

«автоматический» приводит пространство к тому, в котором данные имеют единичную ковариационную матрицу (whitening).



Аксиомы псевдометрики легко проверяются с учетом неотрицательной определенности  $S$ , неравенство треугольника следует из неравенства Коши-Буняковского для скалярного произведения по отношению к  $S$ :  $(x, y)_S = x^T S y$ .

Если потребовать  $S \succ 0$  получим метрику  $\Rightarrow \rho_M$  будет метрикой, если матрица объекты-признаки имеет полный ранг.

# Обобщения метрик

## Метаметрика

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  ~~$x = y \Rightarrow \rho(x, y) = 0$~~
4.  $\rho(x, y) = \rho(y, x)$
5.  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Разрешаем расстоянию между совпадающими объектами не равняться нулю.

## Квазиметрика

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  $x = y \Rightarrow \rho(x, y) = 0$
4.  ~~$\rho(x, y) = \rho(y, x)$~~
5.  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Разрешаем расстоянию быть несимметричным.



# Обобщения метрик

## Полуметрика

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  $x = y \Rightarrow \rho(x, y) = 0$
4.  $\rho(x, y) = \rho(y, x)$
5.  ~~$\rho(x, y) \leq \rho(x, z) + \rho(z, y)$~~

Разрешаем невыполнение неравенства треугольника.

Пример: Расстояние Минковского

$$l_p(x, y) = \left( \sum_{i=1}^d |x^i - y^i|^p \right)^{1/p} \quad \text{при } p < 1.$$

Достаточно рассмотреть контрпример:

$$a = (0, 0); b = (0, 1); c = (1, 1)$$

$$l_p(a, c) = 2^{1/p} > l_p(a, b) + l_p(b, c) = 2$$

при  $p < 1$ .

# Еще обобщения метрик

## Праметрика

1.  $\rho(x, y) \geq 0$
2.  ~~$\rho(x, y) = 0 \Rightarrow x = y$~~
3.  ~~$x = y \Rightarrow \rho(x, y) = 0$~~
4.  ~~$\rho(x, y) = \rho(y, x)$~~
5.  ~~$\rho(x, y) \leq \rho(x, z) + \rho(z, y)$~~

## Дивергенция

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  $x = y \Rightarrow \rho(x, y) = 0$
4.  ~~$\rho(x, y) = \rho(y, x)$~~
5.  ~~$\rho(x, y) \leq \rho(x, z) + \rho(z, y)$~~

Например, расстояние Кульбака — Лейблера между распределениями:

$$D_{KL}(p||q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

# Независимость системы аксиом метрики

Система аксиом независима, если каждая аксиома системы не является логическим следствием остальных аксиом.

1.  $\rho(x, y) \geq 0$
2.  $\rho(x, y) = 0 \Rightarrow x = y$
3.  $x = y \Rightarrow \rho(x, y) = 0$
4.  $\rho(x, y) = \rho(y, x)$
5.  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Аксиома неотрицательности следует из 3, 4, 5 аксиом:

$$0 = \rho(x, x) \leq \rho(x, y) + \rho(y, x) = 2\rho(x, y) \Rightarrow \rho(x, y) \geq 0$$

# Независимость системы аксиом

Независимая система:

1.  $\rho(x, y) = 0 \Rightarrow x = y$
2.  $x = y \Rightarrow \rho(x, y) = 0$
3.  $\rho(x, y) = \rho(y, x)$
4.  $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

Как доказать? От противного: предполагаем выводимость одной аксиомы из остальных, находим контрпример, где взятая аксиома не выполнена, а остальные выполнены.

Контрпримеры:

1.  $\rho(x, y) \equiv 0$
2.  $\rho(x, y) \equiv 1$
3.  $\rho(x, y) = x - y$  при  $x - y \geq 0$  и  $\rho(x, y) = 1$  при  $x - y < 0$
4.  $\rho(x, y) = l_p(x, y)$  при  $p < 1$

# Напоминание: метод главных компонент (PCA)

Линейный метод понижения размерности без учителя.

Задача понижения размерности: по выборке  $x_1, \dots, x_N \in \mathbb{R}^D$  построить новое признаковое представление данных  $\tilde{x}_1, \dots, \tilde{x}_N \in \mathbb{R}^d$ , где  $d < D$ .

- ▶ Борьба с переобучением
- ▶ Снижение вычислительных затрат
- ▶ Визуализация
- ▶ Сжатие данных

# Идея: метод главных компонент

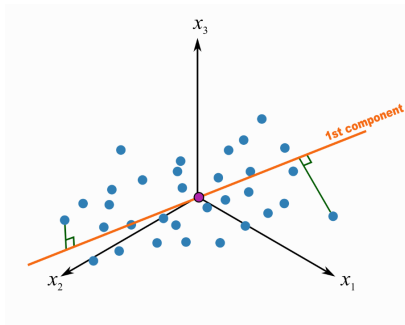
Найти подпространство размерности  $d$ , имеющее минимальную сумму квадратов ошибки:

$$\sum_{i=1}^N \|h_i\|^2 \rightarrow \min_{v_1, \dots, v_d}$$

~

Найти подпространство размерности  $d$ , проекции объектов на которое будут иметь наибольший разброс:

$$\sum_{i=1}^N \|p_i\|^2 \rightarrow \max_{v_1, \dots, v_d}$$

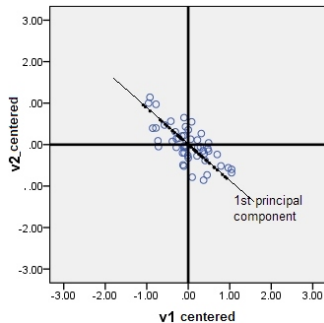
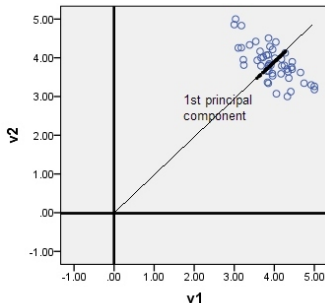


$$\begin{aligned} x_i &= h_i + p_i, \quad p_i \in L, \quad h_i \perp L; \\ L &= \ell(v_1, \dots, v_d); \\ \tilde{x}_i &= p_i; \end{aligned}$$

## Решение: метод главных компонент

Предварительно обрабатываем данные:

Ищем подпространство, проходящее через 0  $\Rightarrow$  центрируем до применения метода.



## Решение: метод главных компонент

Решаем задачу максимизации жадно:

1.  $\|Xa_1\|^2 \rightarrow \max_{a_1}, \text{ т.ч. } \|a_1\| = 1$
2.  $\|Xa_2\|^2 \rightarrow \max_{a_2}, \text{ т.ч. } \|a_2\| = 1, a_2 \perp a_1$
3.  $\|Xa_3\|^2 \rightarrow \max_{a_3}, \text{ т.ч. } \|a_3\| = 1, a_3 \perp a_2, a_3 \perp a_1$

и т.д.

$$\frac{\partial}{\partial a_1} (a_1^T X^T X a_1 - \lambda (a_1^T a_1 - 1)) = 2X^T X a_1 - 2\lambda a_1 = 0$$

$\Rightarrow a_1$  — с.в.  $X^T X$ , причем  $\|Xa_1\|^2 = \lambda \|a_1\|^2 = \lambda$

$\Rightarrow a_1$  соотв. наиб. с.з.

и т.д.



## Решение: метод главных компонент

Можно поставить эквивалентную задачу максимизации разброса проекций на первую главную компоненту:

$$\text{Var}(Xa_1) = \frac{1}{N-1}(Xa_1)^T Xa_1 = a_1^T \frac{1}{N-1} X^T X a_1 = a_1^T S a_1, \|a_1\| = 1.$$

Главные компоненты последовательно описывают направления, вдоль которых данные имеют наибольший разброс, в порядке невозрастания с.з.

Можно показать, что жадный подбор компонент дает искомое оптимальное подпространство.

Т.о. выборочная ковариационная матрица симметрична  $\Rightarrow \exists$  о.н. базис из с.в.  $\Rightarrow$  упорядочиваем с.з. по невозрастанию и в качестве главных компонент берем первые  $d$  соответствующих с.в.

# Многомерное шкалирование

Задача:

Про исходные данные  $\hat{X} = \hat{x}_1, \dots, \hat{x}_N$  известны лишь попарные различия, требуется найти такое признаковое представление данных в маломерном пространстве ( $p = 2, 3$ )  $x_i = (x_i^1, x_i^2)$ , что  $d_{ij} = \text{dissim}(\hat{x}_i, \hat{x}_j) \approx \|x_i - x_j\|_2$ .

Применение: визуализация.



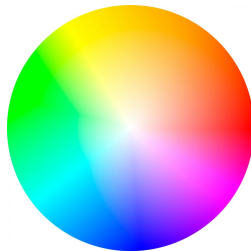
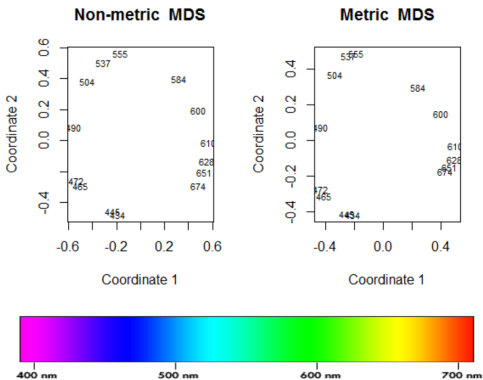
## Цветовое восприятие человека

Получили симметричную матрицу различий с нулями на диагонали.

	434	445	465	472	490	504	537	555	584	600	610	628	651
445	0.14												
465	0.58	0.50											
472	0.58	0.56	0.19										
490	0.82	0.78	0.53	0.46									
504	0.94	0.91	0.83	0.75	0.39								
537	0.93	0.93	0.90	0.90	0.69	0.38							
555	0.96	0.93	0.92	0.91	0.74	0.55	0.27						
584	0.98	0.98	0.98	0.98	0.93	0.86	0.78	0.67					
600	0.93	0.96	0.99	0.99	0.98	0.92	0.86	0.81	0.42				
610	0.91	0.93	0.98	1.00	0.98	0.98	0.95	0.96	0.63	0.26			
628	0.88	0.89	0.99	0.99	0.99	0.98	0.98	0.97	0.73	0.50	0.24		
651	0.87	0.87	0.95	0.98	0.98	0.98	0.98	0.98	0.80	0.59	0.38	0.15	
674	0.84	0.86	0.97	0.96	1.00	0.99	1.00	0.98	0.77	0.72	0.45	0.32	0.24

# Цветовое восприятие человека

MDS → двумерный цветовой круг



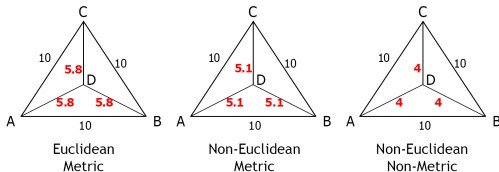
## Евклидовы и неевклидовы расстояния

Вопрос: заданная матрица различий  $\equiv$  матрица попарных расстояний?

Если да, существует ли такое признаковое представление данных  $x_i$ , что выполнено точное равенство:  $d_{ij} = \text{dissim}(\hat{x}_i, \hat{x}_j) = \|x_i - x_j\|_2$ ?

Если  $\exists \{x_i\}_{i=1}^N \in \mathbb{R}^p$ , т.ч.  $d_{ij} = \|x_i - x_j\|_2$ , то исходное расстояние  $d(x_i, x_j)$  — **евклидово** (Euclidean distance).

Если  $\forall p$  и  $\forall \{x_i\}_{i=1}^N \in \mathbb{R}^p$  выполнено  $d_{ij} \neq \|x_i - x_j\|_2$ , то исходное расстояние  $d(x_i, x_j)$  — **неевклидово** (Non-euclidean distance).



$$5.8 \rightarrow \frac{2}{3}\sqrt{75}$$

# Классическая теория многомерного шкалирования

Пусть входная матрица различий — матрица попарных расстояний для евклидовой метрики  $D = (d_{ij})$ .

Задача классического MDS (cMDS) — найти  $X = (x_1, \dots, x_N)^T$ , т.ч.  $d_{ij} = \|x_i - x_j\|_2$ .

Решение не единственно:  $X^* = X + c^T$  тоже решение, т.к.

$d_{ij} = \|x_i - x_j\| = \|(x_i + c) - (x_j + c)\|$ . Ищем центрированную конфигурацию  $\bar{x} = 0$ .

# Восстановление координат в cMDS

Матрица  $D$  евклидова, т.е.  $\exists \{x_i\}_{i=1}^N \in \mathbb{R}^p$ , т.ч.

$$d_{ij}^2 = (x_i - x_j)^T (x_i - x_j)$$

Идея: восстановить  $\{x_i\}_{i=1}^N \in \mathbb{R}^p$ , при условии  $\bar{x} = 0$ .

Решение: восстановим матрицу Грама  $B = (b_{ij})$ , где  $b_{ij} = x_i^T x_j$

Обозначим  $X = (x_1, \dots, x_N)^T \Rightarrow B = XX^T$

Т. о. если получим матрицу Грама, то сможем по ее спектральному разложению получить  $X$ .



## Восстановление матрицы Грама в cMDS

- ▶  $d_{ij}^2 = (x_i - x_j)^T (x_i - x_j) = x_i^T x_i + x_j^T x_j - 2x_i^T x_j = b_{ii} + b_{jj} - 2b_{ij}$
- ▶  $\bar{x} = 0 \Rightarrow \sum_{i=1}^N b_{ij} = 0$
- ▶  $\frac{1}{N} \sum_{i=1}^N d_{ij}^2 = \frac{1}{N} \sum_{i=1}^N b_{ii} + b_{jj}$ 
$$\frac{1}{N} \sum_{j=1}^N d_{ij}^2 = b_{ii} + \frac{1}{N} \sum_{j=1}^N b_{jj}$$
$$\frac{1}{N^2} \sum_{i,j=1}^N d_{ij}^2 = \frac{2}{N} \sum_{i=1}^N b_{ii}$$
- ▶  $b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2)$

## Решение cMDS

Строим по  $D$  матрицу Грама  $B$ :

$$\blacktriangleright b_{ij} = -\frac{1}{2}(d_{ij}^2 - d_{i\bullet}^2 - d_{\bullet j}^2 + d_{\bullet\bullet}^2)$$

$$\blacktriangleright B = C_N A C_N, \text{ где } A = -\frac{1}{2}D^2 \text{ — поэлементно, } C_n = E - \frac{1}{n}\mathbf{1}\mathbf{1}^T.$$

$B = XX^T$ ,  $X \in \mathbb{R}^{N \times p} \Rightarrow B$  — симметричная, неотрицательно определенная, ранга  $rgB = rgXX^T = rgX = p$

$B$  имеет  $p$  положительных с.з. и  $n - p$  нулевых с.з.

$B = \Gamma \Lambda \Gamma^T$ , где  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$  и  $\Gamma = (\gamma_1, \dots, \gamma_p)$  — матрица с.в.

$$X = \Gamma \Lambda^{\frac{1}{2}}$$

# cMDS

cMDS позволяет восстановить  $\{x_i\}_{i=1}^N \in \mathbb{R}^p$ , т.ч.  $\bar{x} = 0$  и  $d_{ij} = \|x_i - x_j\|_2$  — *точное равенство*.

Размерность  $p$  восстановленных координат может не удовлетворять нашим целям.

Для нахождения координат  $\{x_i\}_{i=1}^N \in \mathbb{R}^q$  меньшей размерности  $q < p$  берем  $\Gamma = (\gamma_1, \dots, \gamma_q)$   $q$  с.в. матрицы Грама с наибольшими с.з. и находим искомые координаты  $X = \Gamma \Lambda^{\frac{1}{2}} \in \mathbb{R}^{N \times q}$

Тогда  $d_{ij} \approx \|x_i - x_j\|$  — *точное равенство не гарантируется*.

## cMDS ~ PCA

## cMDS ~ PCA

## cMDS

Находим матрицу Грама  $XX^T$ , используем ее спектральное разложение.

## PCA

Находим выборочную ковариационную матрицу  $\frac{1}{N-1}X^TX$ , используем ее спектральное разложение.

$XX^T \in \mathbb{R}^{N \times N}$ ,  $X^TX \in \mathbb{R}^{D \times D}$ , но спектры совпадают с точностью до добавления нулевых с.з.:

$XX^Tv = \lambda v \Rightarrow X^T(XX^Tv) = X^T(\lambda v) \Rightarrow$  если  $v$  - с.в.  $XX^T$ , то  $X^Tv$  - с.в.  $X^TX$  с тем же с.з.

# cMDS

cMDS можно использовать и для неевклидовых матриц, и вообще для произвольных матриц различий (не обязательно расстояний).

Вообще говоря, могут быть отрицательные с.з. матрицы Грама  $XX^T$ . Тогда можно оставить только с.в., отвечающие положительным с.з.

cMDS будет находить  $\{x_i\}_{i=1}^N \in \mathbb{R}^q$ , т.ч.  $\bar{x} = 0$  и  $d_{ij} \approx \|x_i - x_j\|$ .

## сMDS

Вопрос: как понять, что матрица попарных расстояний  $D$  евклидова?

Theorem (Критерий «евклидовости» матрицы различий)

Пусть  $A = (a_{ij})$ ,  $a_{ij} = -\frac{1}{2}d_{ij}^2$  и  $B = C_N A C_N$ , где  $C_N$  — центрирующая матрица.

$D$  — евклидова тогда и только тогда, когда  $B$  неотрицательно определена.

По сути пытаемся к  $D$  применить алгоритм сMDS.

# Пример применения cMDS

**TABLE 13.2.** *Airline distances (km) between 18 cities. Source: Atlas of the World, Revised 6th Edition, National Geographic Society, 1995, p. 131.*

	Beijing	Cape Town	Hong Kong	Honolulu	London	Melbourne
Cape Town	12947					
Hong Kong	1972	11867				
Honolulu	8171	18562	8945			
London	8160	9635	9646	11653		
Melbourne	9093	10338	7392	8862	16902	
Mexico	12478	13703	14155	6098	8947	13557
Montreal	10490	12744	12462	7915	5240	16730
Moscow	5809	10101	7158	11342	2506	14418
New Delhi	3788	9284	3770	11930	6724	10192
New York	11012	12551	12984	7996	5586	16671
Paris	8236	9307	9650	11988	341	16793
Rio de Janeiro	17325	6075	17710	13343	9254	13227
Rome	8144	8417	9300	12936	1434	15987
San Francisco	9524	16487	11121	3857	8640	12644
Singapore	4465	9671	2575	10824	10860	6050
Stockholm	6725	10334	8243	11059	1436	15593
Tokyo	2104	14737	2893	6208	9585	8159
	Mexico	Montreal	Moscow	New Delhi	New York	Paris
Montreal	3728					
Moscow	10740	7077				
New Delhi	14679	11286	4349			
New York	3362	533	7530	11779		
Paris	9213	5522	2492	6601	5851	

# Пример применения cMDS

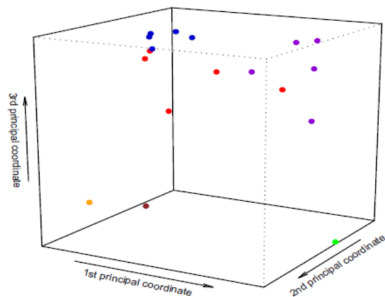
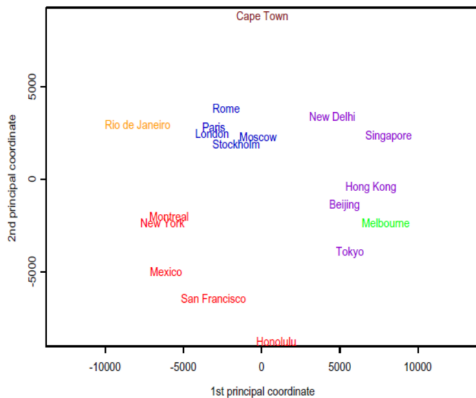
**TABLE 13.6.** *Eigenvalues of  $B$  and the eigenvectors corresponding to the first three largest eigenvalues (in red) for the airline distances example.*

	Eigenvalues	Eigenvectors		
1	471582511	0.245	-0.072	0.183
2	316824787	0.003	0.502	-0.347
3	253943687	0.323	-0.017	0.103
4	-98466163	0.044	-0.487	-0.080
5	-74912121	-0.145	0.144	0.205
6	-47505097	0.366	-0.128	-0.569
7	31736348	-0.281	-0.275	-0.174
8	-7508328	-0.272	-0.115	0.094
9	4338497	-0.010	0.134	0.202
10	1747583	0.209	0.195	0.110
11	-1498641	-0.292	-0.117	0.061
12	145113	-0.141	0.163	0.196
13	-102966	-0.364	0.172	-0.473
14	60477	-0.104	0.220	0.163
15	-6334	-0.140	-0.356	-0.009
16	-1362	0.375	0.139	-0.054
17	100	-0.074	0.112	0.215
18	0	0.260	-0.214	0.173

Авиарасстояние (Airline distance) не является евклидовым.



# Пример применения сMDS



## Другой подход к MDS

Решаем ту же задачу:

Найти  $\{x_i\}$  такие, что  $d_{ij} = \text{dissim}(\hat{x}_i, \hat{x}_j) \approx \|x_i - x_j\|_2$ .

Другой подход: определим функционал

$$\text{Stress}(x_1, \dots, x_N) = \left( \sum_{i \neq j=1, \dots, N} (d_{ij} - \|x_i - x_j\|)^2 \right)^{1/2}$$

Будем его оптимизировать:  $\text{Stress} \rightarrow \min_{x_1, \dots, x_N}$

Например, градиентным спуском.

Такой подход называется метрическим многомерным шкалированием (*metric MDS* — *mMDS*).

## Другой подход к MDS

По сути cMDS решает аналогичную задачу:

$$Strain(x_1, \dots, x_N) = \left( \frac{\sum_{i,j} (b_{ij} - x_i^T x_j)^2}{\sum_{i,j} b_{ij}^2} \right)^{1/2} \rightarrow \min_{x_1, \dots, x_N},$$

где  $B$  - матрица Грама.

## Пример mMDS

$$Sammons\ stress(x_1, \dots, x_N) = \frac{1}{\sum_{i < j} d_{ij}} \sum_{i < j} \frac{(\|x_i - x_j\| - d_{ij})^2}{d_{ij}}$$

Идея: даем ошибкам веса, т.ч. метод лучше воспроизводит малые расстояния.

Оптимизируем итерационно, например, градиентным спуском. Начальные координаты  $x_1, \dots, x_N$  можно получить методом классического многомерного шкалирования (cMDS).

## Еще пример задачи MDS

Wolford и Hollingsworth собрали данные, как часто человек ошибается при распознавании букв, показанных ему в течение нескольких миллисекунд. Получена матрица ошибок (ее кусок):

Letter	C	D	G	H	M	N	Q	W
C	—							
D	5	—						
G	12	2	—					
H	2	4	3	—				
M	2	3	2	19	—			
N	2	4	1	18	16	—		
Q	9	20	9	1	2	8	—	
W	1	5	2	5	18	13	4	—

*Можно ли считать эту матрицу матрицей различий (dissimilarities matrix)?*

## non-metric MDS

Как перевести матрицу сходств (similarities matrix) в матрицу различий (dissimilarities matrix)?

Можно найти наибольшее значение из сходств  $c = \max \delta_{ij}$  (наибольшее число в матрице) и преобразовать матрицу:  
 $\delta_{ij} \rightarrow c - \delta_{ij}$ .

Проблема: полученные расстояния «плавают» в зависимости от выбора  $c$ . Сохраняется лишь порядок между расстояниями.

⇒ Идея неметрического многомерного шкалирования (non-metric MDS) — лучше воспроизводить порядковые отношения между расстояниями, чем сами расстояния.

# non-metric MDS

Постановка проблемы (Shepard-Kruskal Scaling):

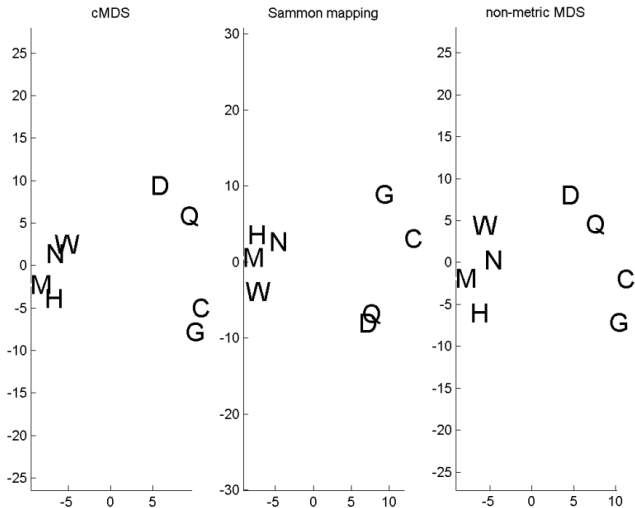
Дана симметричная матрица с нулями на диагонали  $\Delta = (\delta_{ij})$

Найти  $X = (x_1, \dots, x_N) \in \mathbb{R}^{d \times n}$ , т.ч.  $\forall i, j, k, l$   
 $\|x_i - x_j\|_2^2 < \|x_k - x_l\|_2^2 \iff \delta_{ij} < \delta_{kl}$

Аналогично cMDS,  $\exists X$  — точное решение, то есть выполнено:  
 $\forall i, j, k, l \quad \|x_i - x_j\|_2^2 < \|x_k - x_l\|_2^2 \iff \delta_{ij} < \delta_{kl}$

Идея та же, что в cMDS: восстанавливаем матрицу Грама  $B$  и боремся с отрицательными с.з.:  $\hat{B} = B - \min(\lambda_{\min}(B), 0)I$ .

## Еще пример задачи MDS





## Резюме

- ▶ Метрические методы встречаются в большинстве задач анализа данных, позволяют решать как задачи машинного обучения, так и преобразовывать и визуализировать данные.
- ▶ Спектральное разложение — ключевая идея метода главных компонент, классического многомерного шкалирования, расстояния Махаланобиса.
- ▶  $PCA \sim MDS$ .
- ▶ Основные методы многомерного шкалирования: cMDS, mMDS, nMDS.