

# Динамическое выравнивание многомерных временных рядов

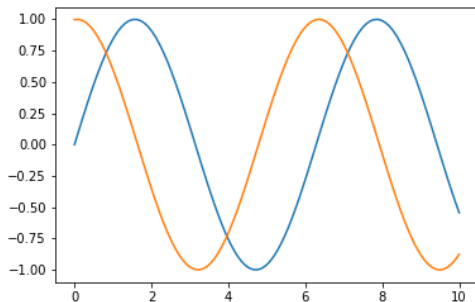
Моргачев Г., Смирнов В., Липницкая Т.,  
Руководитель: Гончаров А.

10 декабря 2018 г.

# Сравнение рядов

## Проблемы

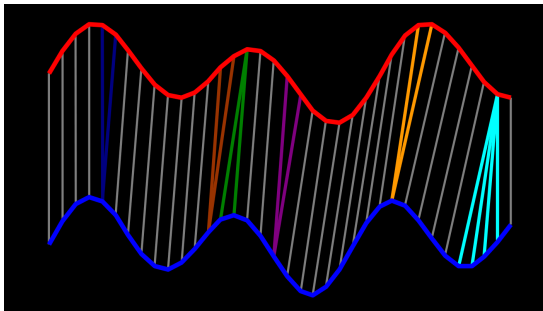
- Растяжение
- Сдвиги



# DTW

## DTW

- Выравнивание рядов друг относительно друга
- Позволяет задать функцию расстояния
- Использует матрицу попарных расстояний между точками рядов



# Многомерное DTW

## Особенность

Необходимость выбора функции расстояния между соответственными точками рядов

## Постановка задачи

Зависимость качества кластеризации временных рядов от выбора функции расстояния между ними

# Задача

Множество временных рядов  $\mathbb{S} \subset \mathbb{R}^{l \times n}$ , где  $l$  - количество каналов,  $n$  - длина ряда.

$\forall s_i \in \mathbb{S}$  задано  $y_i \in \mathbb{Y}$  - множество меток классов.

Функции расстояния между векторами  $\mathbb{R}$ :

$$\mathbb{R} = \{\rho : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^+\}$$

Соответствующие **DTW**

$$g_\rho : \mathbb{S} \times \mathbb{S} \rightarrow \mathbb{R}^+$$

# Задача

Пусть  $S \subset \mathbb{S}$ ,  $|S| = N$  - выборка

Матрица попарных расстояний:

$$D(g_\rho(S)) = ||D_{ij}||, \quad D_{ij} = g_\rho(s_i, s_j), \quad s_i, s_j \in S$$

Кластеризатор:

$$f : D \rightarrow Z^N, \quad Z - \text{множество меток кластеров}$$

# Задача

## Функции качества

$$Q_1(f(D), S) = \frac{1}{|Z|} \sum_{z \in Z} \max_y \frac{N_z^y}{N_z}$$

$$Q_2(f(D), S) = \frac{1}{|Z|} \sum_{z \in Z} \max_y \frac{(N_z^y)^2}{N_z N^y}$$

## Постановка задачи

$$Q_i(D(g_\rho(S), S) \rightarrow \max_{\rho}$$

## Кластеризация

**Иерархическая** с функциями расстояния между кластерами:

- 1 *complete*:  $d(A, B) = \max_{a \in A, b \in B} (dist(a, b))$
- 2 *weighted*:  $d(A, B) = \frac{(dist(S, B) + dist(T, B))}{2}$ , где кластер  $A = S \cup T$
- 3 *weighted*:  $d(u, v) = \sum_{a \in A, b \in B} \frac{d(a, b)}{(|A| * |B|)}$



## Данные

Размеченные данные ускорений акселерометра телефона

- 6 состояния человека
- 3 канала
- Разбиты на ряды по 50 точек
- Размер выборки - 2048
- Производится нормализация данных

# Результаты

$\rho$	n	$Q_1$			$Q_2$		
		<i>compl</i>	<i>aver</i>	<i>weigh</i>	<i>compl</i>	<i>aver</i>	<i>weight</i>
$L_1$	24	0.5059	0.5854	0.6384	0.2732	0.3761	0.4488
	36	0.5325	0.6196	0.6163	0.2988	0.4246	0.4140
	48	0.5563	0.6388	0.6308	0.3303	0.4432	0.4306
$L_2$	24	0.4876	0.6216	0.6258	0.2701	0.4173	0.4246
	36	0.4982	0.6459	0.6433	0.2701	0.4545	0.4489
	48	0.5336	0.6486	0.6530	0.2701	0.4546	0.4615

# Результаты

## Выводы

Лучшие результаты в случае выбора функции расстояния, порожденной  $L_2$  нормой. Требуется продолжение исследования.

Спасибо за внимание!