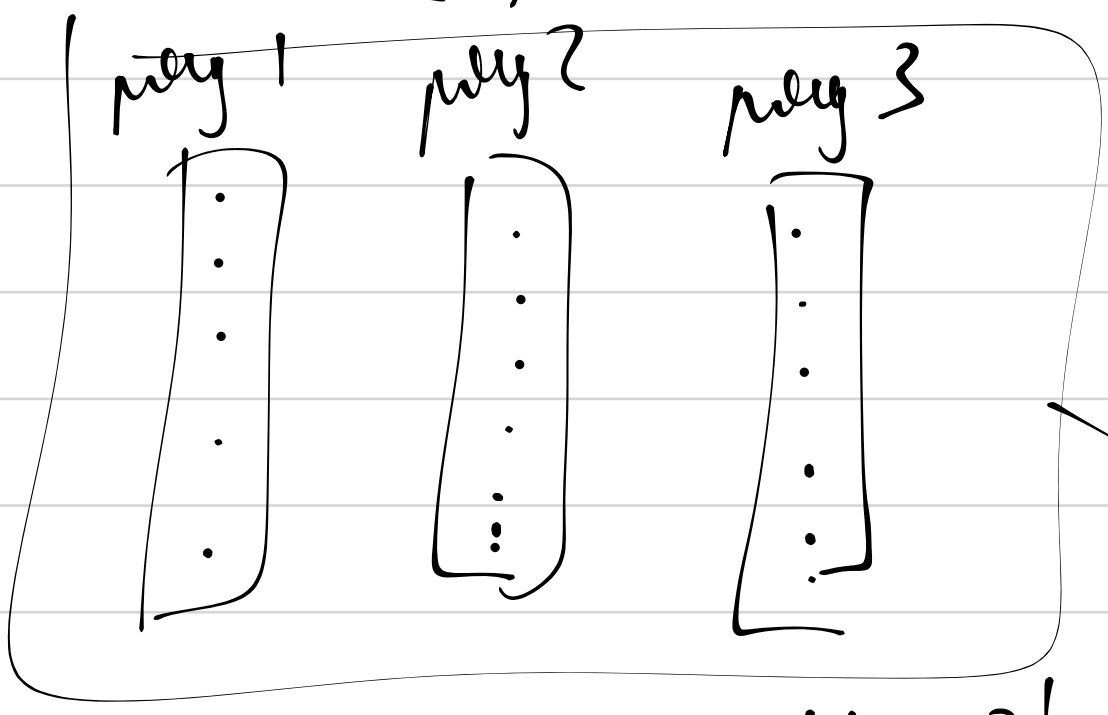
Привет!!

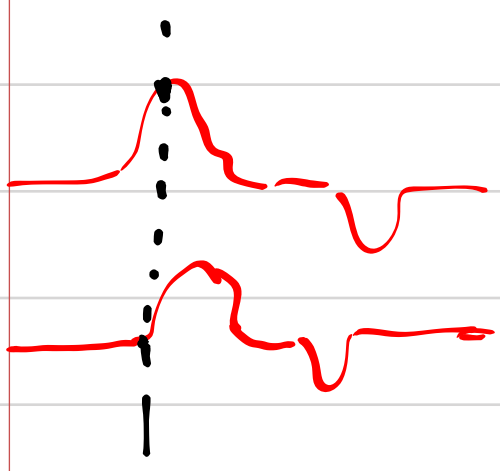
Классификация / кластеризация.

кластериз.

много рядов → [ ] → ряды, поделенные на нек-ко групп

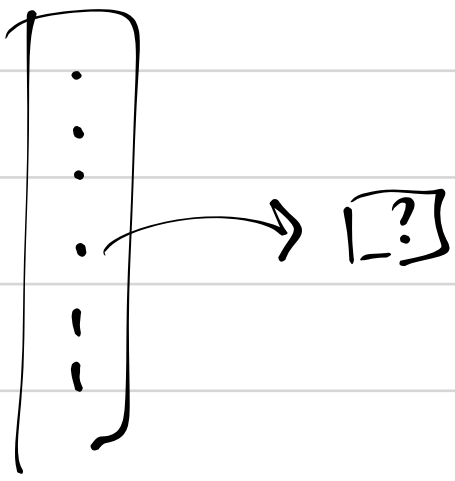


транспонировать!!



	ряд 1	ряд 2	ряд 3	кластеры
ряд 1				А
ряд 2				А
⋮				Б
ряд n				Б
				А
				⋮
				Б

Как создавать признаки?



→ опис. стат-ки для x_t

mean( ) для Δx_t

median( )

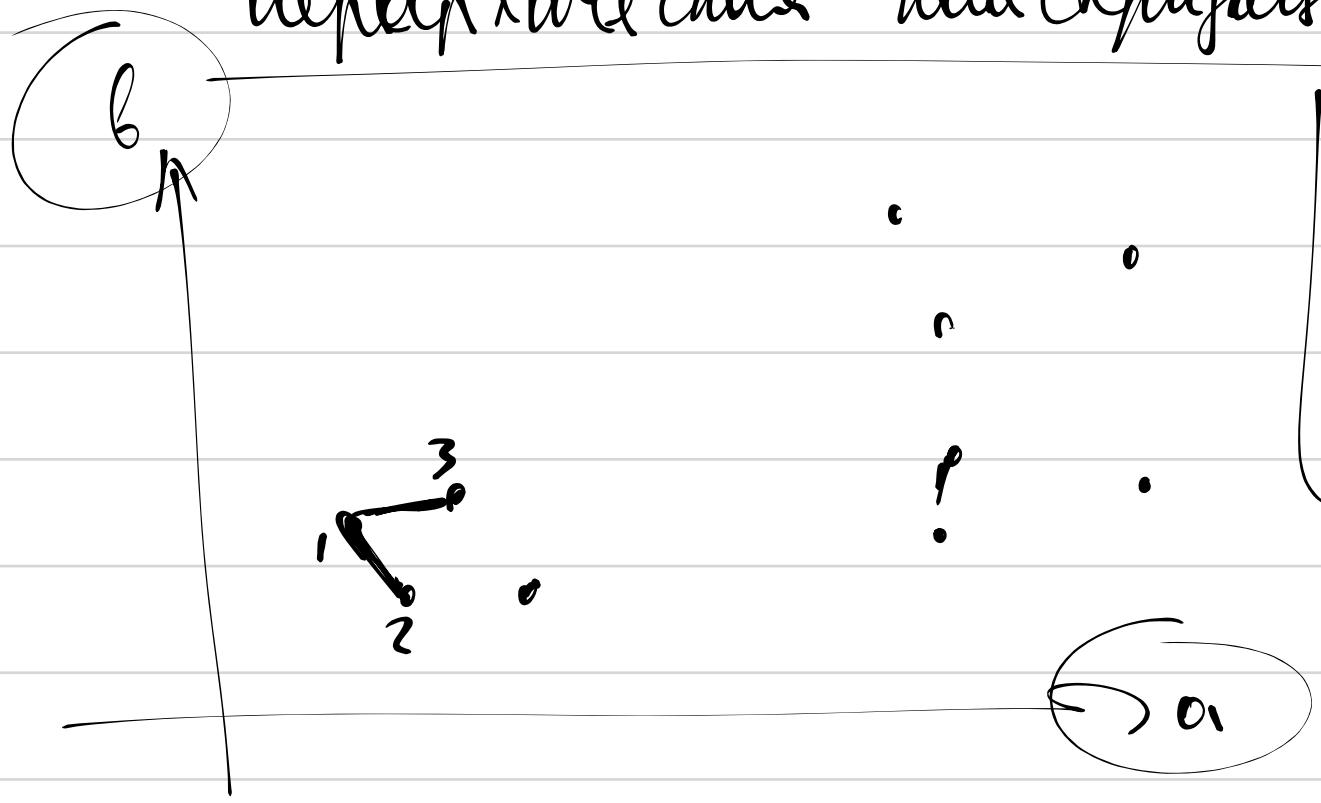
sd( )

автокорр / частные АС

интенс. сезонности

→ тесты (много)

иерархическая кластеризация.



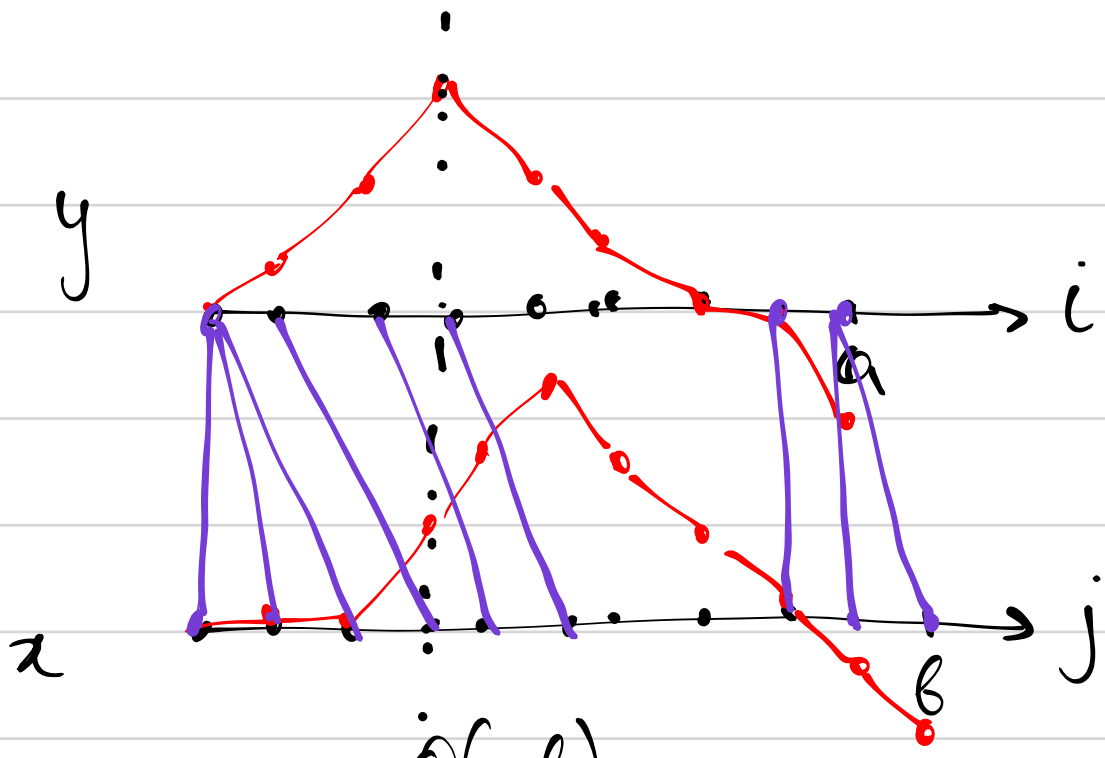
евклидово между ближ-ми представителями

Альтернатива

- посчитать расстояния между рядами.

Dynamic Time Warping.

Динам-ое масштабир-ие времени.

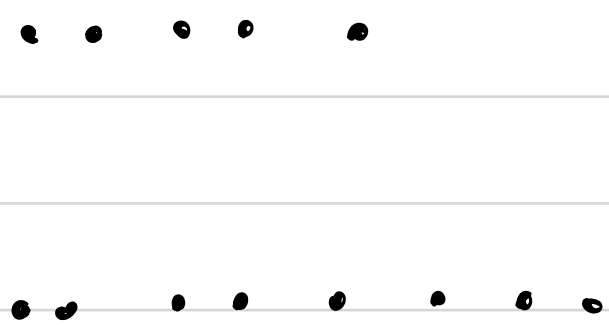


$$\sum_{i \leftrightarrow j} |y(i) - x(j)|$$

Есть алгоритм $O(a \cdot b)$

① $1 \bullet\!\!-\!\!\bullet 1$

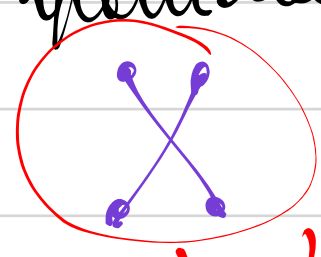
$a \bullet\!\!-\!\!\bullet b$



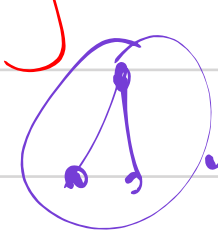
② каждый момент одного ряда сопост-н с одним или несколькими моментами другого.

③ нет возврата в прошлое

такое запрещено! →



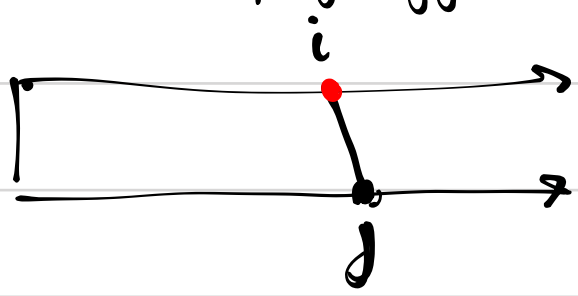
если $i_1 - j_1$, $i_2 - j_2$ и $i_2 > i_1$ $\Rightarrow j_2 \geq j_1$

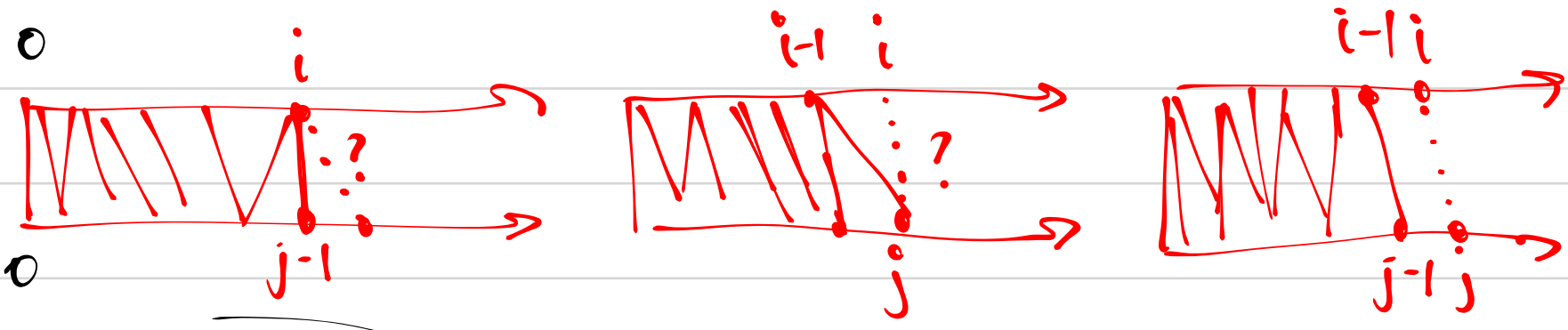


можно.

Алгоритм

допустим предыдущие задачи решены.





$D(i, j-1)$ $D(i-1, j)$ $D(i-1, j-1)$

$|y(i) - x(j)|$

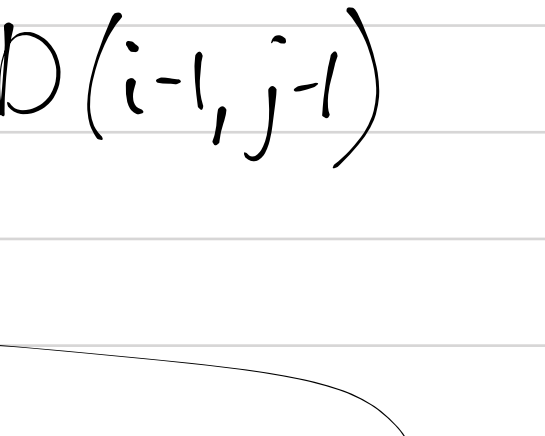
$$D(i,j) = |y(i) - x(j)| + \min(D(i,j-1), D(i-1,j-1), D(i,j-1))$$

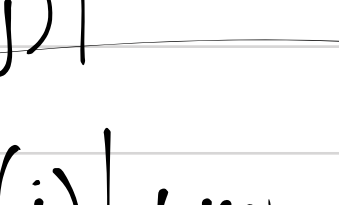
$i \in \{1, \dots a\}$ $j \in \{1, \dots b\}$

Init

$D(i,j) = +\infty$ где $(i>0 \quad j \geq 0)$ $(i \geq 0, j > 0)$

$D(0,0) = 0$





for i in 1:a

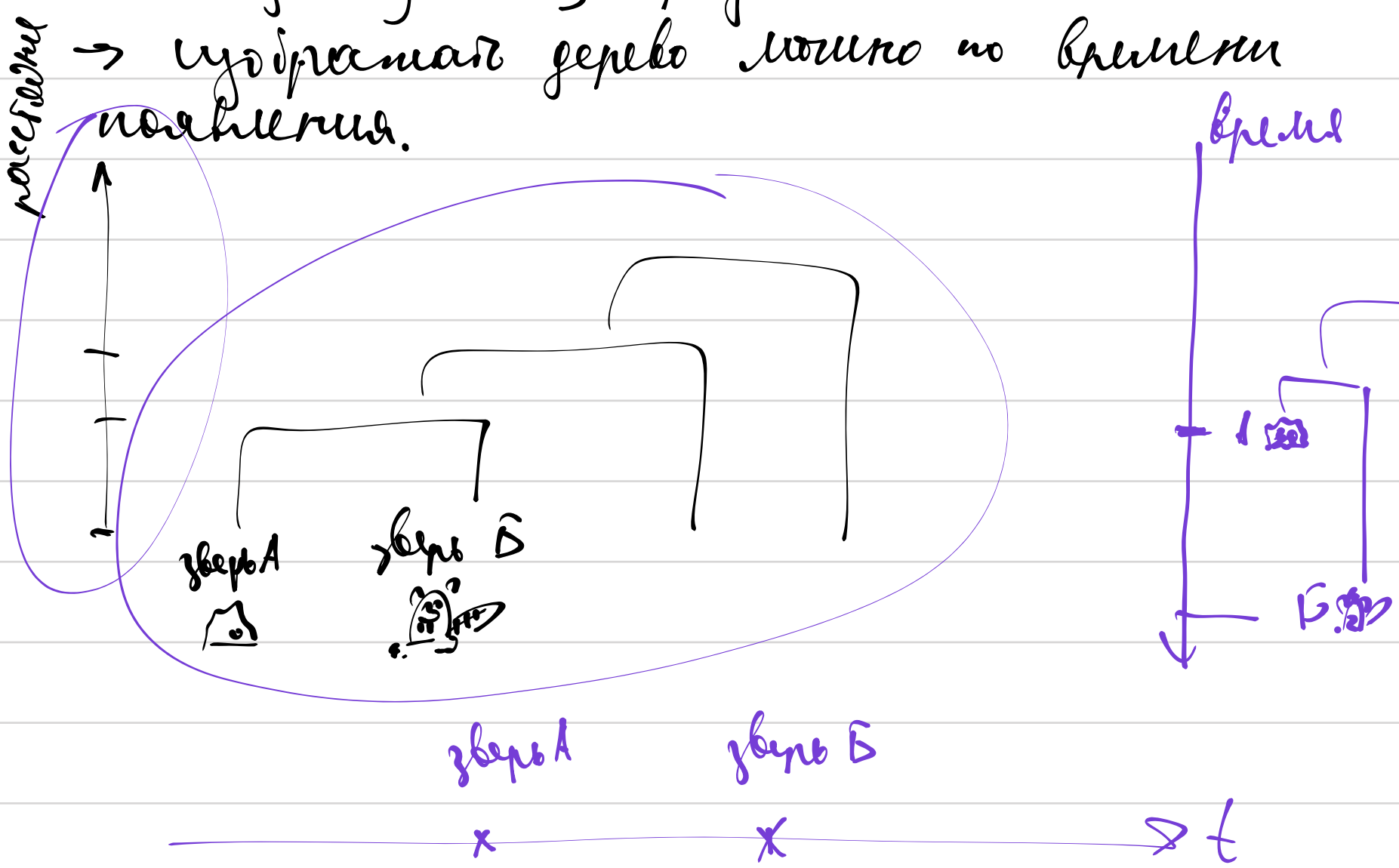
for j in 1:b

$$D(i,j) = |y(i) - x(j)| + \min(D(i,j-1), D(i-1,j-1), D(i,j-1))$$

end

end

Трюк от блокнов

→ строить иерарх-ую кластеризацию можно по одному набору признаков.

→ изображать дерево можно по времени появления.