Research Memo 2024-03-30.

Notations

1) Input Time Series.

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$ p-channels.

2) Time stamp → Non-uniform.

find the minimum time stamp size.

denoted by Δt

Then time index becomes $t = i\,\Delta t$.

$i = 0, 1, 2, \cdots$

Some values of $x_p(t)$, $1 \le p \le P$, are missing

Goal. provide. ~~interpolated~~ predicted

① $\hat{X}(t)$ for all P channels and all t time indices.

② Confidence of prediction.

$$Q(t) = \begin{bmatrix} q_1(t) \\ \vdots \\ q_p(t) \end{bmatrix} \qquad t = i \cdot \Delta t$$

$$0 \leq q_{p(t)} \leq 1$$

Initial Interpolation B-spline. $\rightarrow \hat{X}(t)$.

Reason. Suitable for non-uniform grids.

How about $Q(t)$.

x_1 o o o o o

x_2 o o o o o

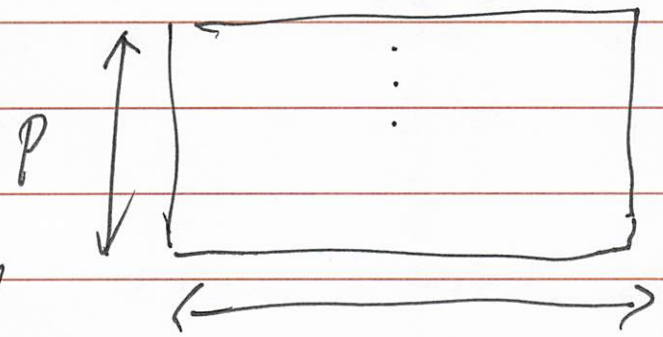
⋮ ⋮ ⋮ ⋮ ⋮ ⋮

x_2 o o o o o

~~c~~ $i-2$ $i-1$ i $i+1$ $i+2$ . . .

If $x_{p(i)}$ is available. $q_{p(i)} = 1$

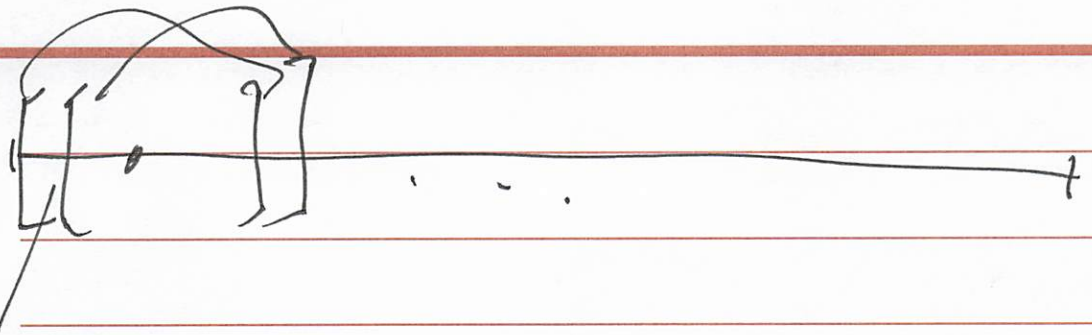
$x_{p(i)}$ is missing. $q_{p(i)} = \frac{\#\text{ of } 1\text{'s inside the window}}{P(2W+1)}$



$i-W$ .. i ... $i+W$

Temporal ~~Window~~ interval centered at i.

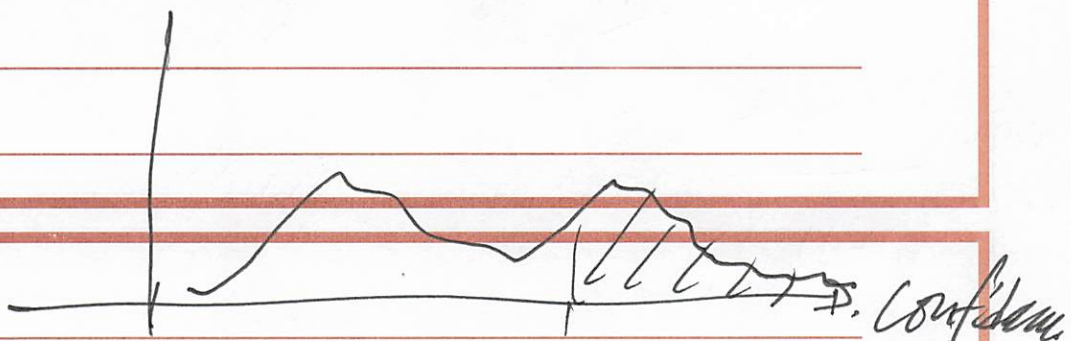
collect intervals with. stride ~~parameters~~ a certain decided by s.



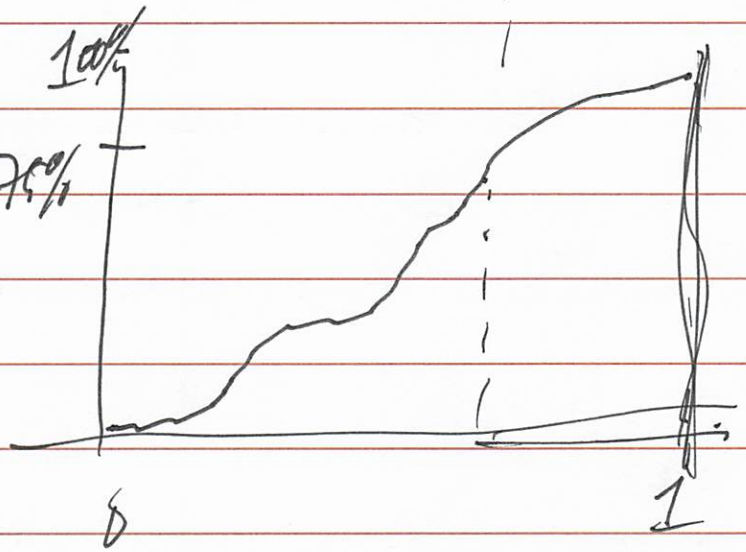
Stride.

We can classify intervals based on their confidence values.

histogram of confidence



cumulative histogram



We may choose the top 25% confidence

Conduct k-means clustering on these intervals.

- Each interval is a P(2W+1) vector of dimensions.
- We may apply the Saak transform to them

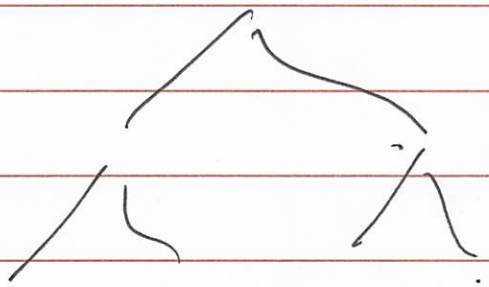
DC + AC

↑

PCA.

for dimension reduction to facilitate k-means clustering.

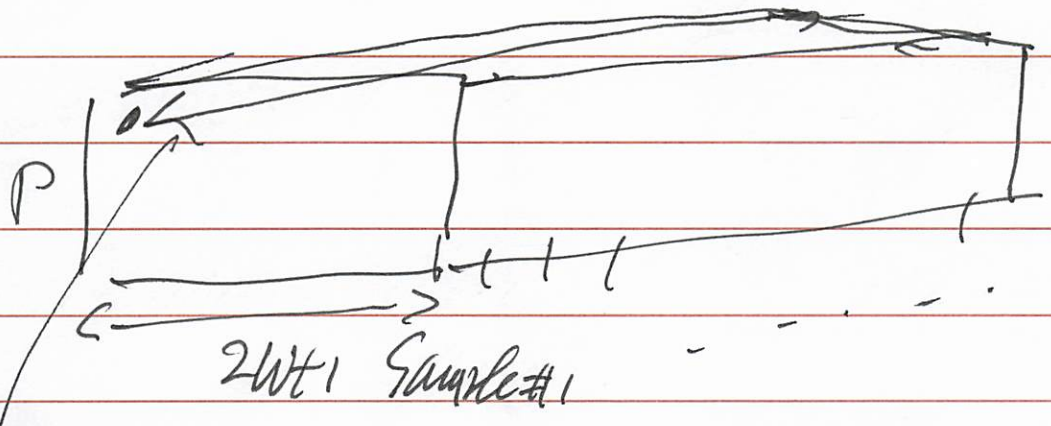
Hierarchical k-means



leaf nodes

For intervals in the same leaf nodes.

Compute element-wise mean.



confidence

- Weighted mean for each channel/time pair.

Final prediction

$$x_{p(i)} = g_{p(i)}\, x_{B,\,p(i)} + (1 - g_{p(i)})\, x_{M,\,p(i)}$$

$x_{B,\,p(i)}$: individual curve.

$x_{M,\,p(i)}$: cluster based prediction.

We can repeat this process for other confidence ranges.