Time Series Analysis

Several slides are borrowed from:

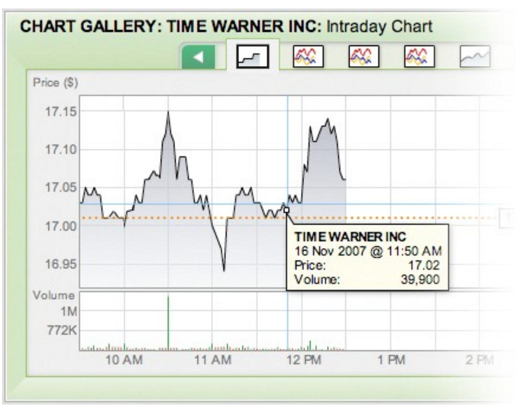
- *Han and Kamber, ”Data Mining: Concepts and Techniques – Mining time-series data”*
- *Lei Chen, ”Similarity Search Over Time-Series Data — Past, Present and Future”*

Contents

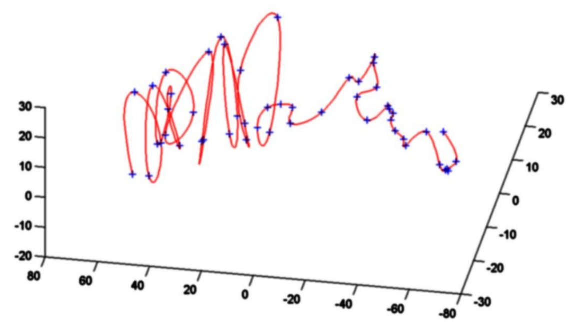
- **Basics**
  - **Time series & preprocessing methods**
- Simple trends models & cycles
  - Regression, seasonal and cyclic variations
- Autocorrelation models
  - ARIMA
- TS similarity
  - Euclidean, DWT, EDR, ERP
- Patterns (Motifs)

Time Series Data and Applications

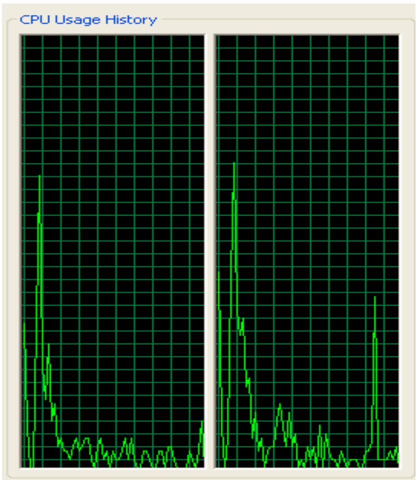
- A time series is an ordered sequence of data values at consecutive timestamps



financial stock data



trajectory of moving objects

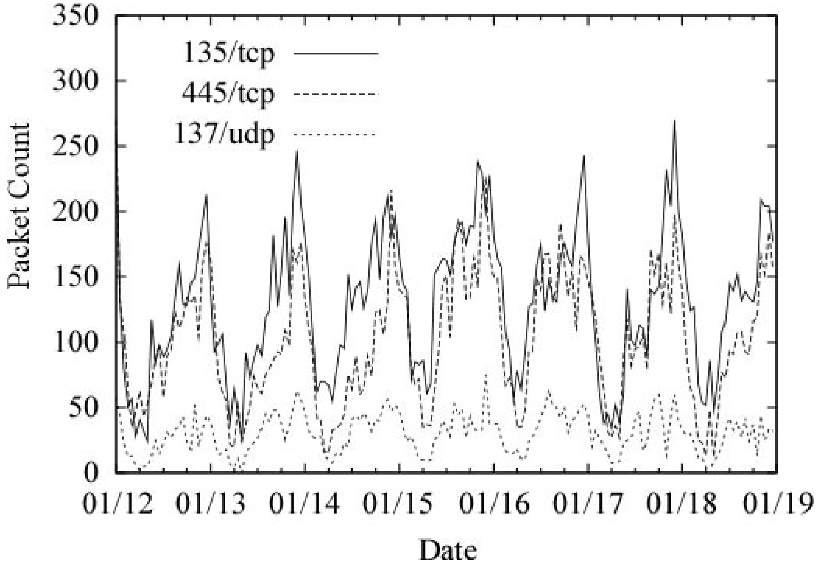


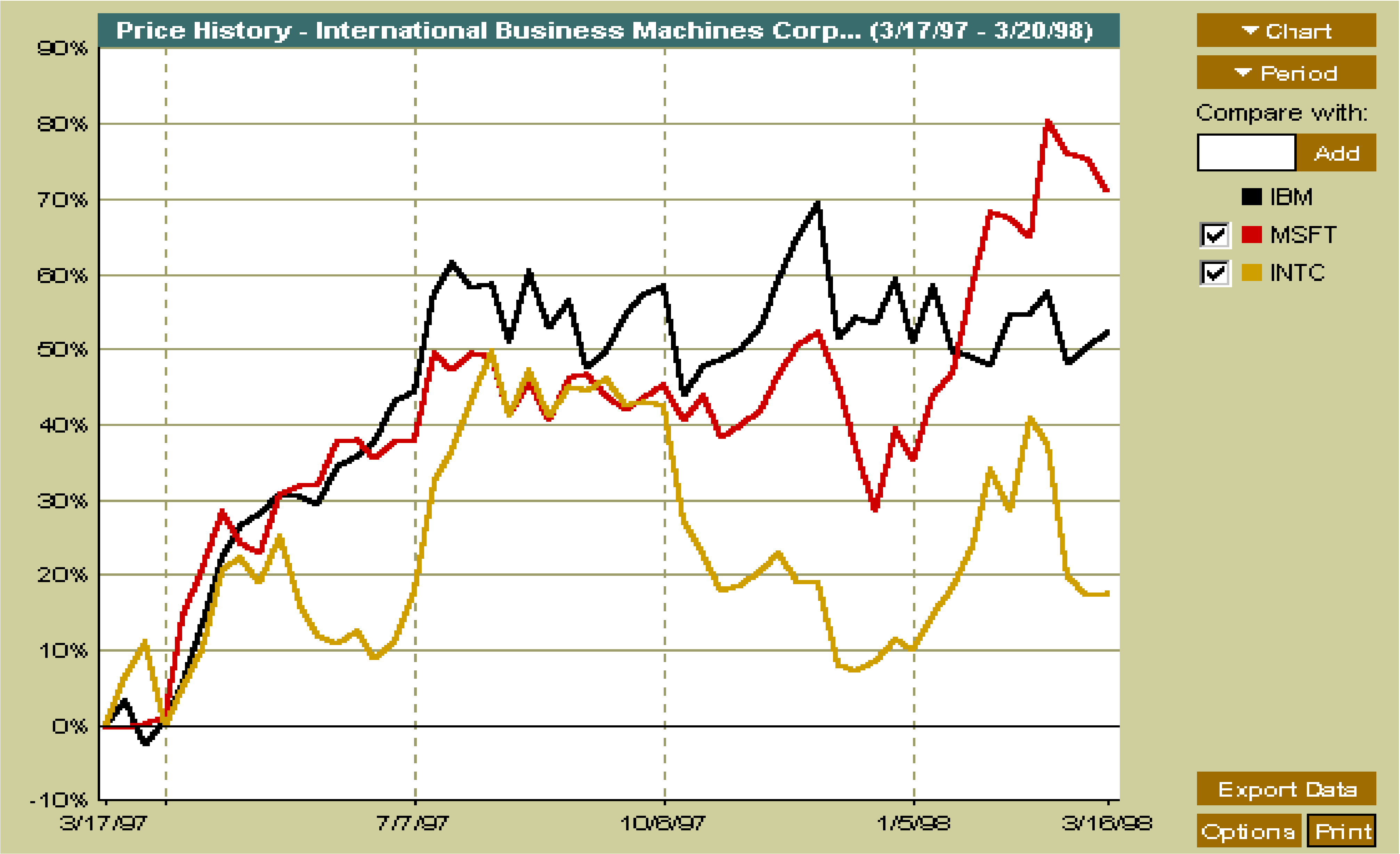
sensory data

video data

Properties of Time Series Data

- Time series data
  - Temporal data correlation
  - High dimensionality
  - Containing repeated patterns

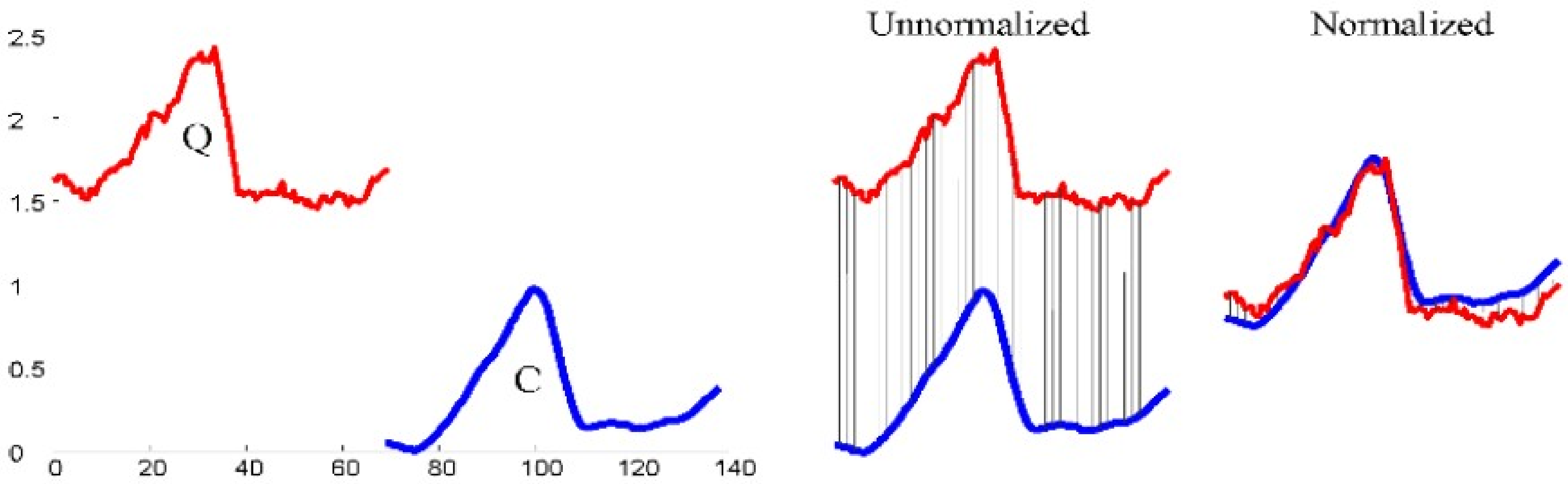




- A time series can be illustrated as a time-series graph which describes a point moving with the passage of time

Need for normalization

Values and times can be differently shifted in each TS



Normalize the time series before using them (e.g., measuring the distance)

$$x_i' = \frac{x_i - \sigma}{\sigma}$$

Other variations (e.g., acceleration and deceleration along the time axis) need other solutions – see later

(Goldin and Kanellakis, 1995)

Moving Average

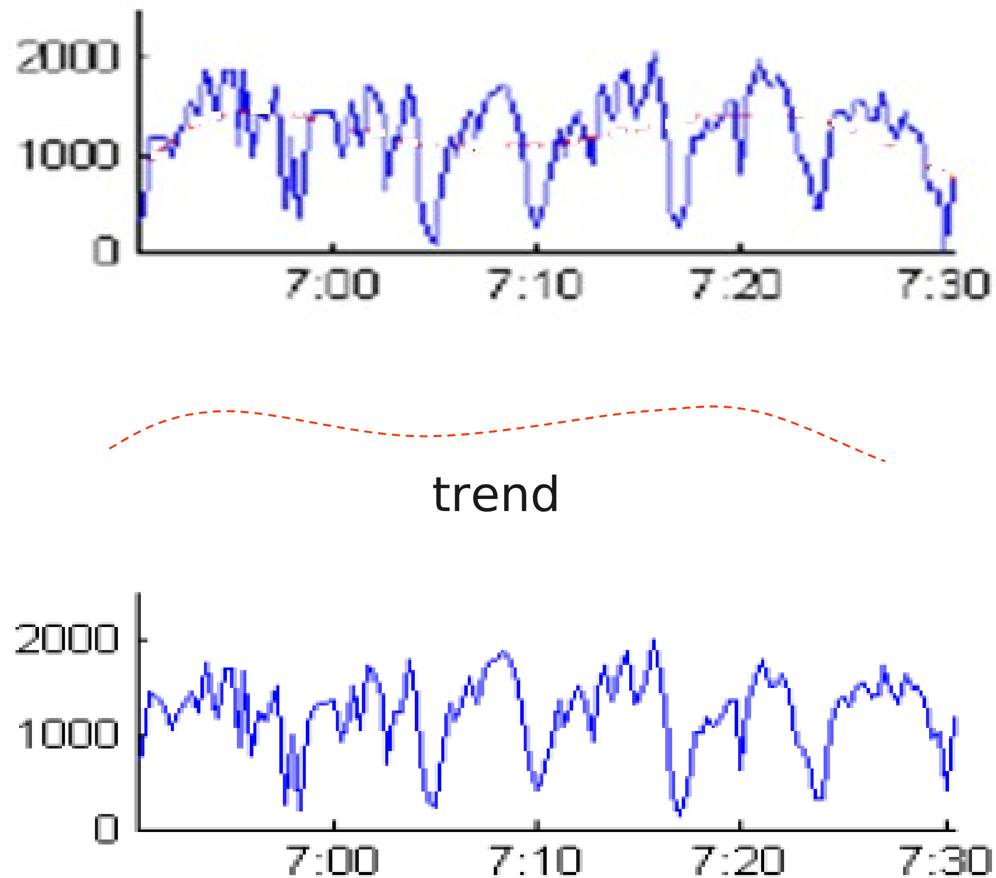
- Moving average of order n

$$\frac{y_1 + y_2 + \cdots + y_n}{n}, \frac{y_2 + y_3 + \cdots + y_{n+1}}{n}, \frac{y_3 + y_4 + \cdots + y_{n+2}}{n}, \cdots$$

  - Smoothes the data
  - Eliminates cyclic, seasonal and irregular movements (see later)
  - Loses the data at the beginning or end of a series
  - Sensitive to outliers (can be reduced by weighted moving average)

Application - Detrending

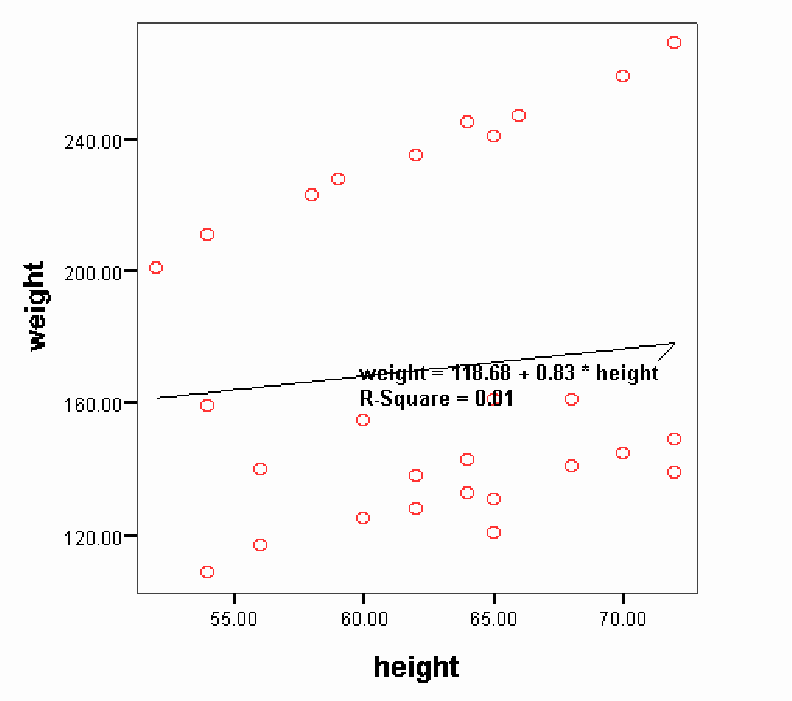
- Obtain the trend for irregular data series
- Subtract trend
- Reveal oscillations



Contents

What is Regression?

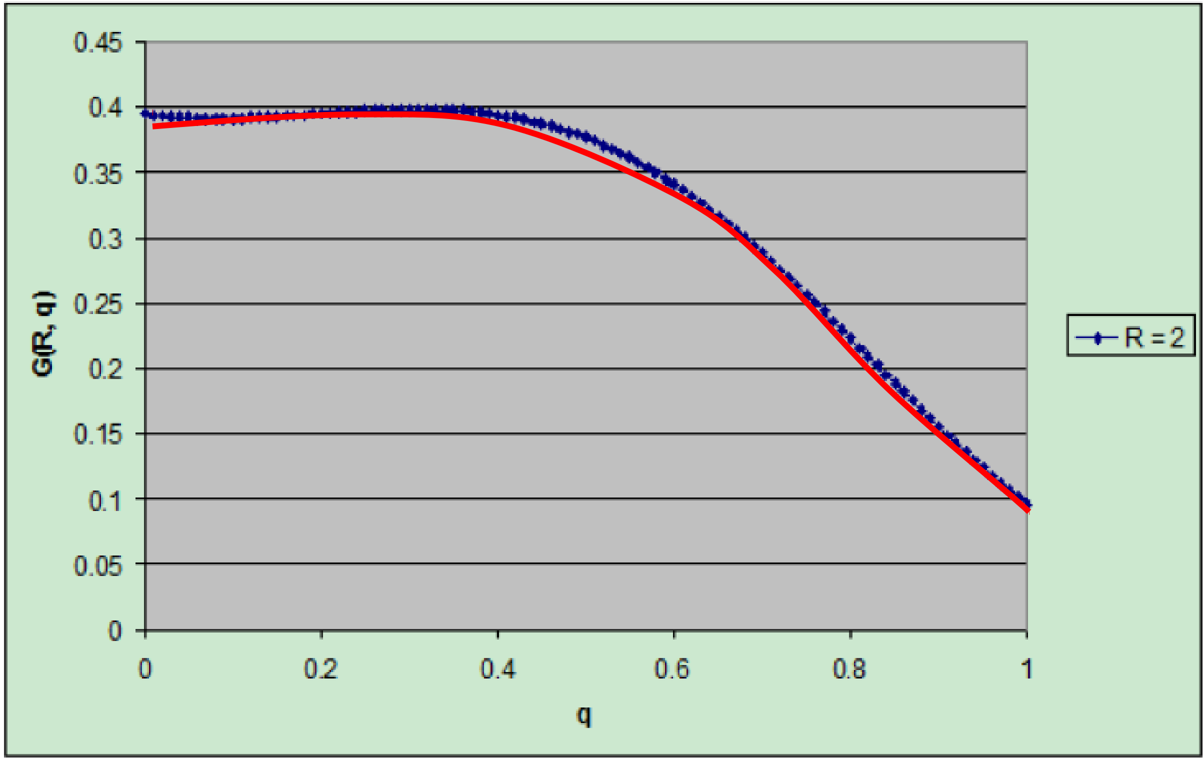
- Modeling the relationship between one **response** variable and one or more **predictor** variables
- Analyzing the confidence of the model
- E.g., height vs. weight



Regression Yields Analytical Model

- Discrete data points →Analytical model
  - General relationship
  - Easy calculation
  - Further analysis
- Application - Prediction

$$G(R,q) \approx \exp(-0.930 - 0.223q + 4.133q^2 - 2.906q^3 - 1.542\pi q^2/R).$$

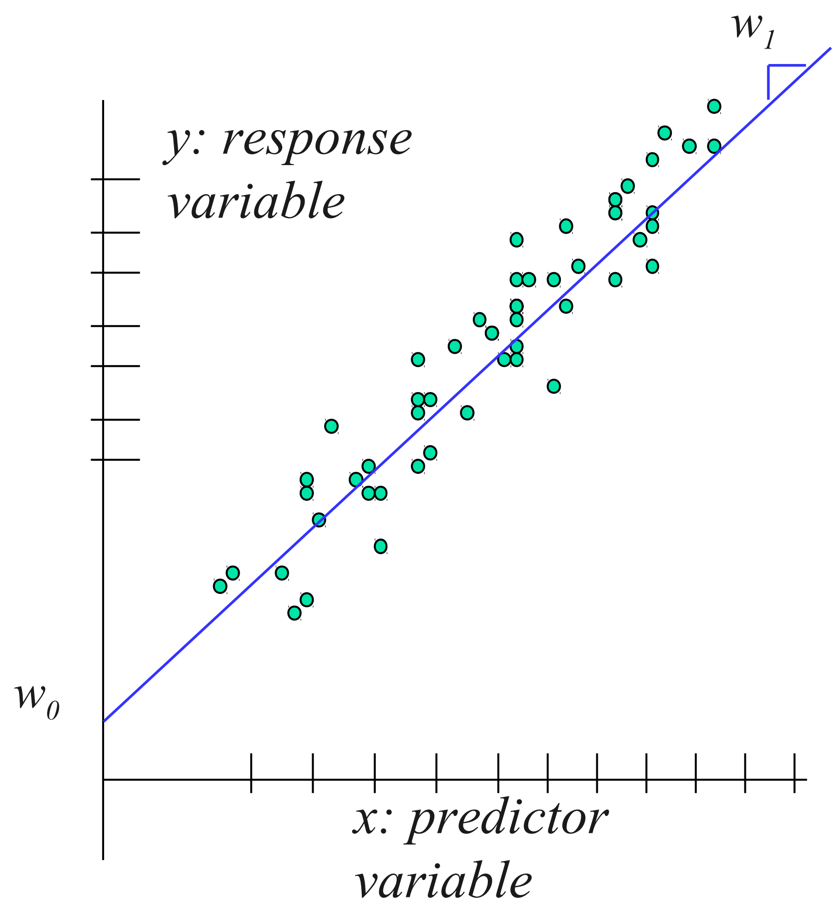


Linear Regression - Single Predictor

- Model is linear

$$y = w_0 + w_1 x$$

where w_0 (y-intercept) and w_1 (slope) are regression coefficients



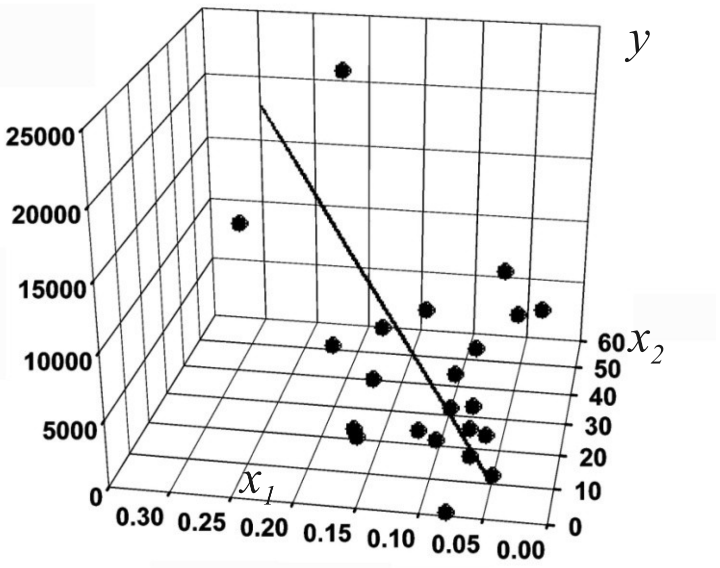
- Method of least squares:

$$w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2} \qquad w_0 = \bar{y} - w_1 \bar{x}$$

Linear Regression – Multiple Predictor

- Training data is of the form $(\boldsymbol{X_1}, y_1), (\boldsymbol{X_2}, y_2), \ldots, (\boldsymbol{X_{|D|}}, y_{|D|})$
- E.g., for 2-D data or

  $y = w_0 + w_1 x_1 + w_2 x_2$



- Solvable by
  - Extension of least square method

    $(X^T X)\, W = Y \rightarrow W = (X^T X)^{-1} Y$
  - Commercial software

    (SAS, S-Plus, etc.)

Nonlinear Regression with Linear Method

- Polynomial regression model
  - E.g., $y = w_0 + w_1 x + w_2 x^2 + w_3 x^3$

    Let $x_2 = x^2$, $x_3 = x^3$

    $y = w_0 + w_1 x + w_2 x_2 + w_3 x_3$
- Log-linear regression model
  - E. g., $y = exp(w_0 + w_1 x + w_2 x_2 + w_3 x_3)$

    Let $y' = log(y)$

    $y' = w_0 + w_1 x + w_2 x_2 + w_3 x_3$

Categories of Time-Series Movements

- Categories of Time-Series Movements
  - Long-term or trend movements (trend curve): general direction in which a time series is moving over a long interval of time
  - Cyclic movements or cycle variations: long term oscillations about a trend line or curve
    - e.g., business cycles, may or may not be periodic
  - Seasonal movements or seasonal variations
    - i.e, almost identical patterns that a time series appears to follow during corresponding months of successive years.
  - Irregular or random movements
- Time series analysis: decomposition of a time series into these four basic movements
  - Additive Modal: TS = T + C + S + I
  - Multiplicative Modal: TS = $T \times C \times S \times I$

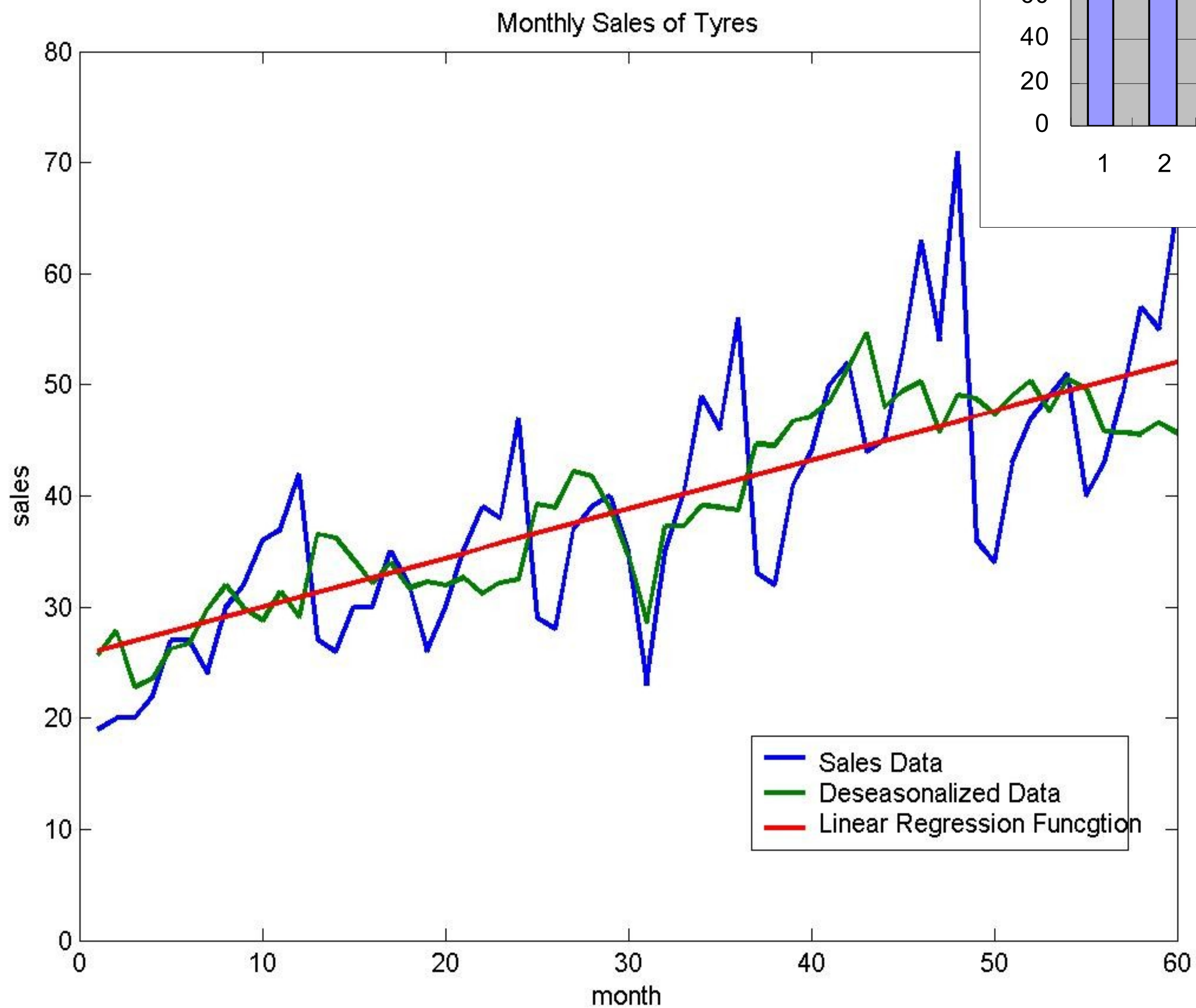
Estimation of Trend Curve

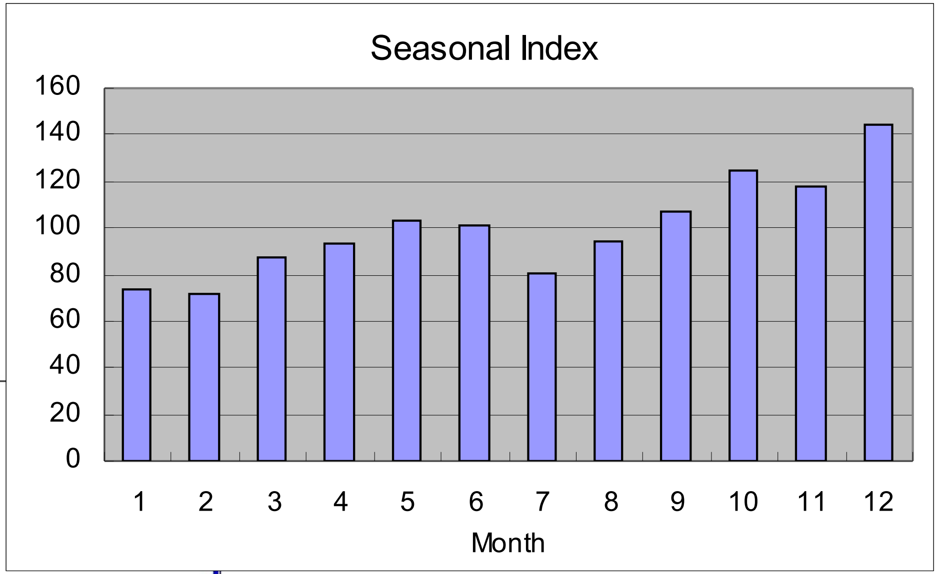
- The freehand method
  - Fit the curve by looking at the graph
  - Costly and barely reliable for large-scaled data mining
- The least-square method
  - Find the curve minimizing the sum of the squares of the deviation of points on the curve from the corresponding data points

Trend Discovery in Time-Series: Estimation of Seasonal Variations

- Seasonal index
  - Set of numbers showing the relative values of a variable during the months of the year
  - E.g., if the sales during October, November, and December are 80%, 120%, and 140% of the average monthly sales for the whole year, respectively, then 80, 120, and 140 are seasonal index numbers for these months
- Deseasonalized data
  - Data adjusted for seasonal variations for better trend and cyclic analysis
  - Divide the original monthly data by the seasonal index numbers for the corresponding months

Seasonal Index





Raw data from http://www.bbk.ac.uk/manop/man/docs/QII_2_2003%20Time%20series.pdf

Trend Discovery in Time-Series

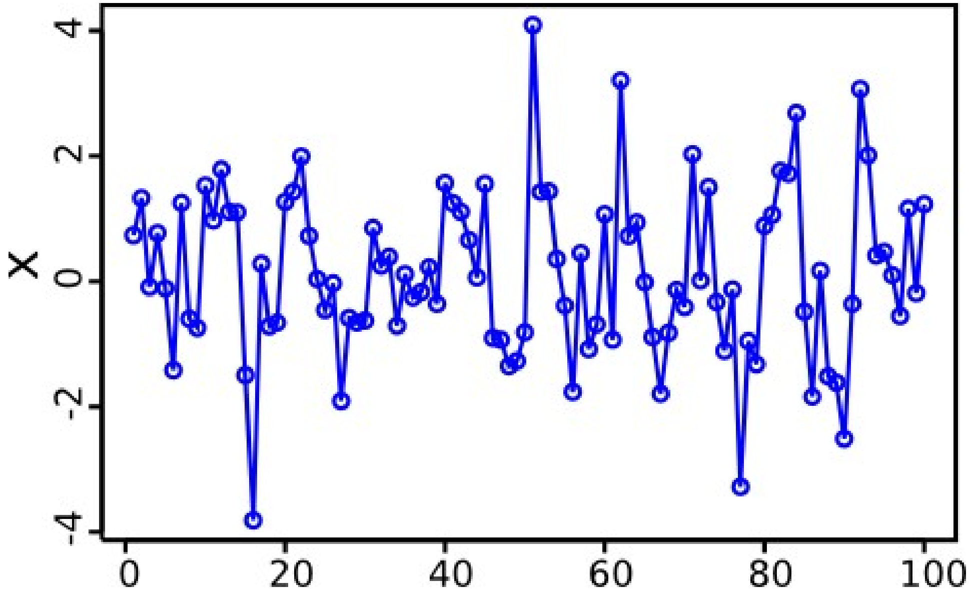
- Estimation of cyclic variations
  - If (approximate) periodicity of cycles occurs, cyclic index can be constructed in much the same manner as seasonal indexes
- Estimation of irregular variations
  - By adjusting the data for trend, seasonal and cyclic variations
- With the systematic analysis of the trend, cyclic, seasonal, and irregular components, it is possible to make long- or short-term predictions with reasonable quality

Contents

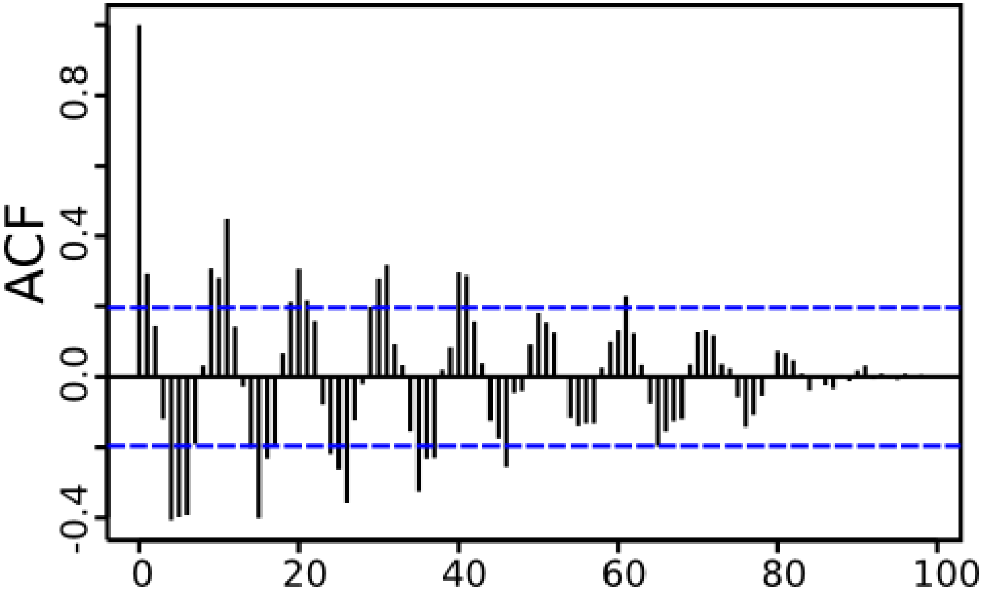
- Basics
  - Time series & preprocessing methods
- Simple trends models & cycles
  - Regression, seasonal and cyclic variations
- **Autocorrelation models**
  - **ARIMA**
- TS similarity
  - Euclidean, DWT, EDR, ERP
- Patterns (Motifs)

Autocorrelation

- Correlation of a time series with itself
  - Similarity between each pair of observations as a function of the time separation between them
- Not so easy to see from the data



Original time series



Correlogram (ACF vs. Time lag)

ARIMA [Box and Jenkins (1976)]

- Autoregressive integrated moving average
  - Complex model that includes two processes
- Element 1: Autoregressive process
  - You can estimate coefficients that describe consecutive elements of the series from previous elements:

    $x(t) = c + a_1*x(t-1) + a_2*x(t-2) + \ldots + \varepsilon$
  - observations made of a random error component and linear combination of prior observations

ARIMA / 2

- Element 2: ”Moving average” process
  - each element in the series can also be affected by the past error (beside autoregressive components):

    $x(t) = \mu + \varepsilon_t - b_1 * \varepsilon_{t-1} - b_2 * \varepsilon_{t-2} - b_3 * \varepsilon_{t-3} - \ldots$

  - observations made of a random error component and a linear combination of prior random shocks

Contents

- Basics
  - Time series & preprocessing methods
- Simple trends models & cycles
  - Regression, seasonal and cyclic variations
- Autocorrelation models
  - ARIMA
- **TS similarity**
  - **Euclidean, DWT, EDR, ERP**
- Patterns (Motifs)

Similarities for TS

- Several DM tasks are based on similarities
  - Clustering → compare the whole TS to form groups
  - Patterns → compare segments of TS
  - KNN Classification → compare whole TS to labeled data
- TS are generally different from vector data
  - Dimension is variable
  - Correspondance between elements can be flexible (S[i] might be compared against T[j], $i \neq j$)

Popular Distance Measures

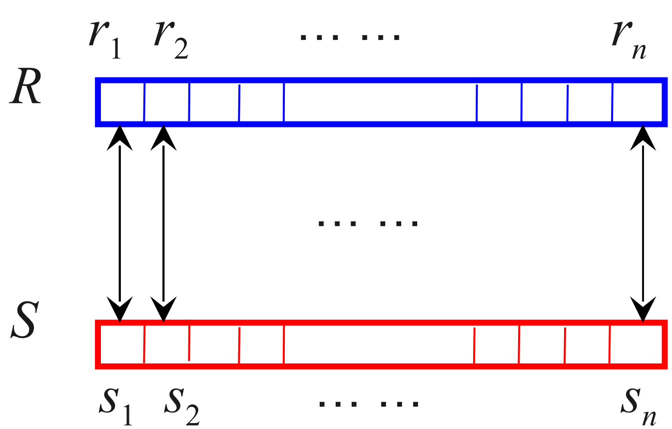
- **Lock-step Measure (one-to-one)**
  - Minkowski Distance
    - L_1 norm (Manhattan Distance)
    - L_2 norm (Euclidean Distance)
    - L_∞ norm (Supremum Distance)
- **Elastic Measure (one-to-many/one-to-none)**
  - Dynamic Time Warping (DTW)
  - Edit distance based measure
    - Longest Common SubSequence (LCSS)
    - Edit distance with Real Penalty (ERP)
    - Edit Sequence on Real Sequence (EDR)

(Ding et al., 2008)

Similarity Measures Without Warping

- Euclidean distance
  - Given $R = <r_1, r_2, \ldots, r_n>$ and $S = <s_1, s_2, \ldots, s_n>$

$$dist(R,S) = \sqrt{\sum_{i=1}^{n}(r_i - s_i)^2}$$



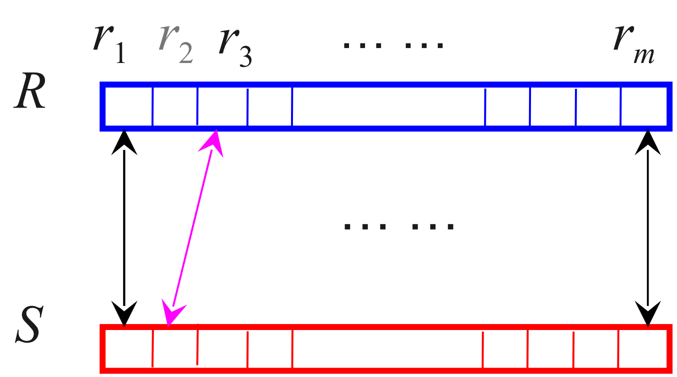
- L_p-norm distance

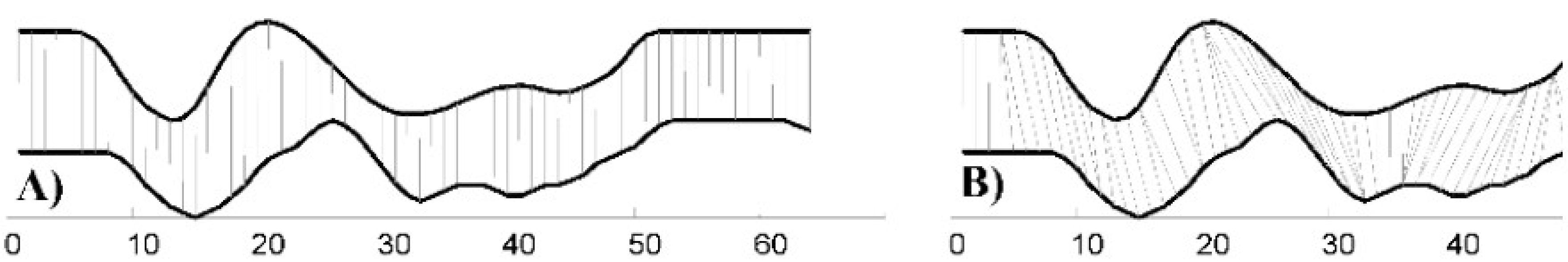
$$dist(R,S) = \sqrt[p]{\sum_{i=1}^{n}|r_i - s_i|^p} \quad (1 \le p < \infty)$$

$$dist(R,S) = max_{i=1}^{n}|r_i - s_i| \quad (p = \infty)$$

Similarity Measures With Warping

- The distance function allows the matching between two time series on warping positions

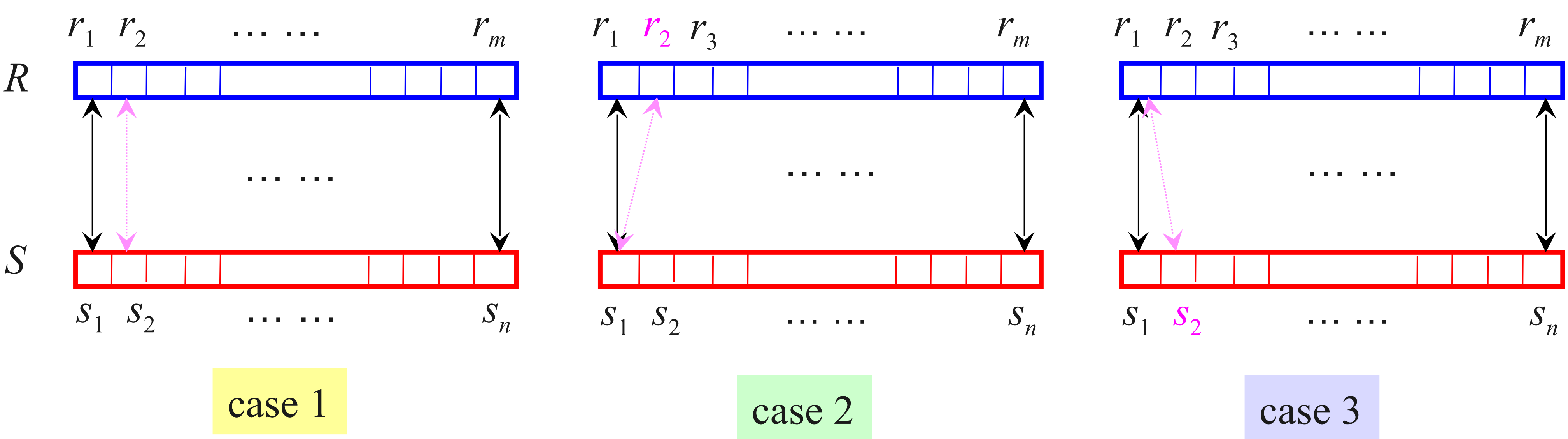




Similarity Measures With Warping (cont'd)

- Dynamic Time Warping (DTW)

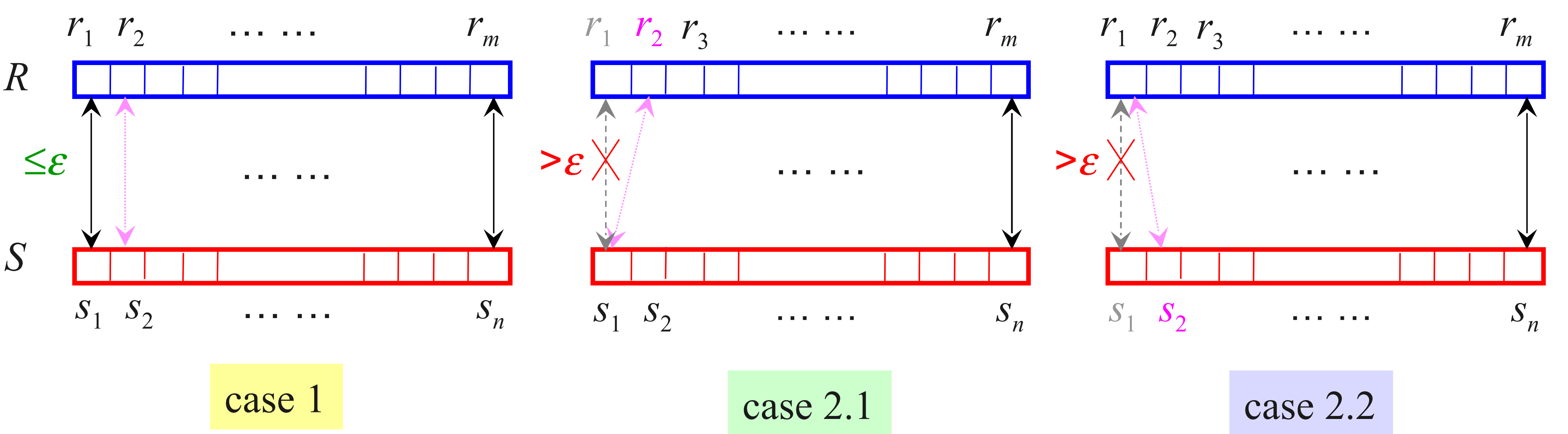
$$DTW(R,S) = \begin{cases} 0 & \text{if } m=n=0 \\ \infty & \text{if } m=0 \text{ or } n=0 \\ dist_{dtw}(r_1,s_1) + min\{DTW(Rest(R),Rest(S)), & \text{otherwise} \\ DTW(Rest(R),S), DTW(R,Rest(S))\} & \end{cases}$$



Similarity Measures With Warping (cont'd)

- Longest Common Subsequences (LCSS)

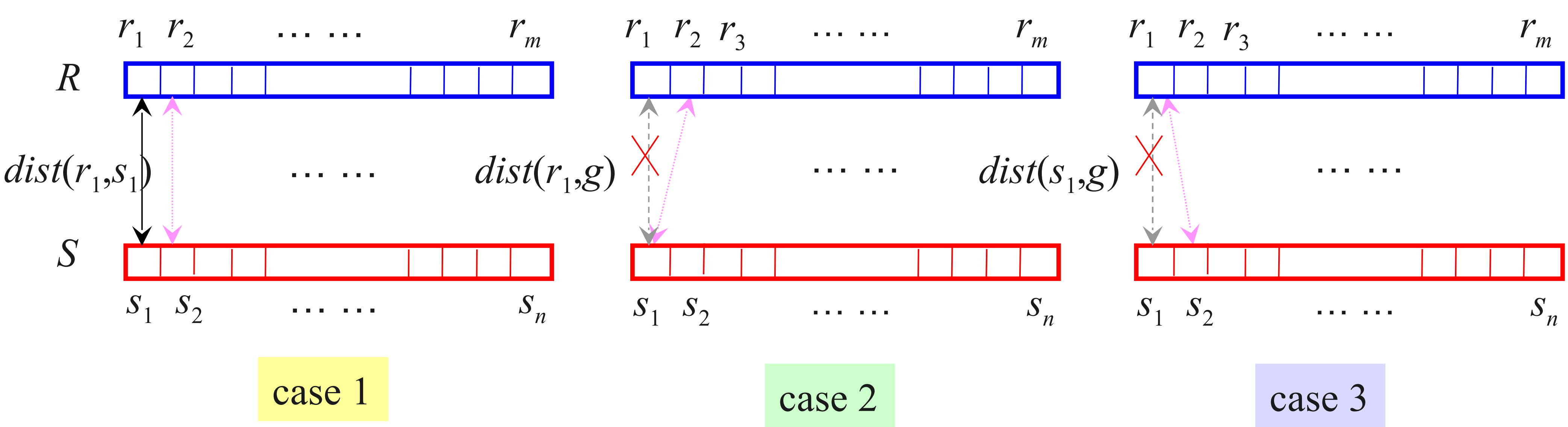
$$LCSS(R,S) = \begin{cases} 0 & \text{if } m = 0 \text{ or } n = 0 \\ LCSS(Rest(R), Rest(S)) + 1 & \text{if } |r_{1,x} - s_{1,x}| \le \epsilon \\ & \& |r_{1,y} - s_{1,y}| \le \epsilon \\ max\{LCSS(Rest(R), S), LCSS(R, Rest(S))\} & \text{otherwise} \end{cases}$$



Similarity Measures With Warping (cont'd)

- Edit distance with Real Penalty (ERP)

$$ERP(R,S) = \begin{cases} \sum_1^n dist(s_i, g) & \text{if } m = 0 \\ \sum_1^m dist(r_i, g) & \text{if } n = 0 \\ min\{ERP(Rest(R), Rest(S)) + dist(t_1, s_1), & \\ ERP(Rest(R), S) + dist(r_1, g), & \text{otherwise} \\ ERP(R, Rest(S)) + dist(s_1, g)\} & \end{cases}$$

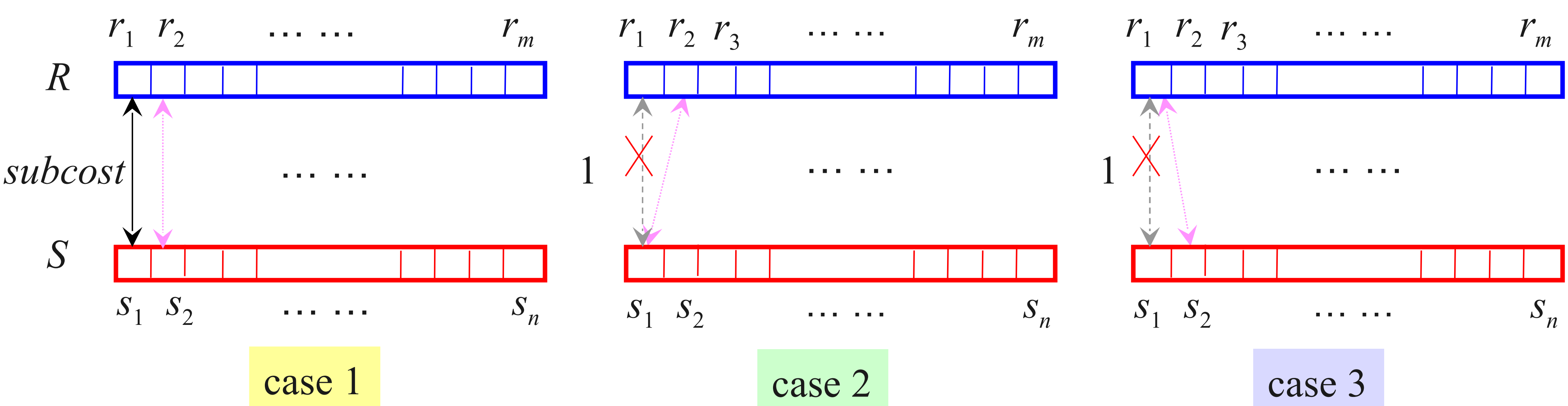


Similarity Measures With Warping (cont'd)

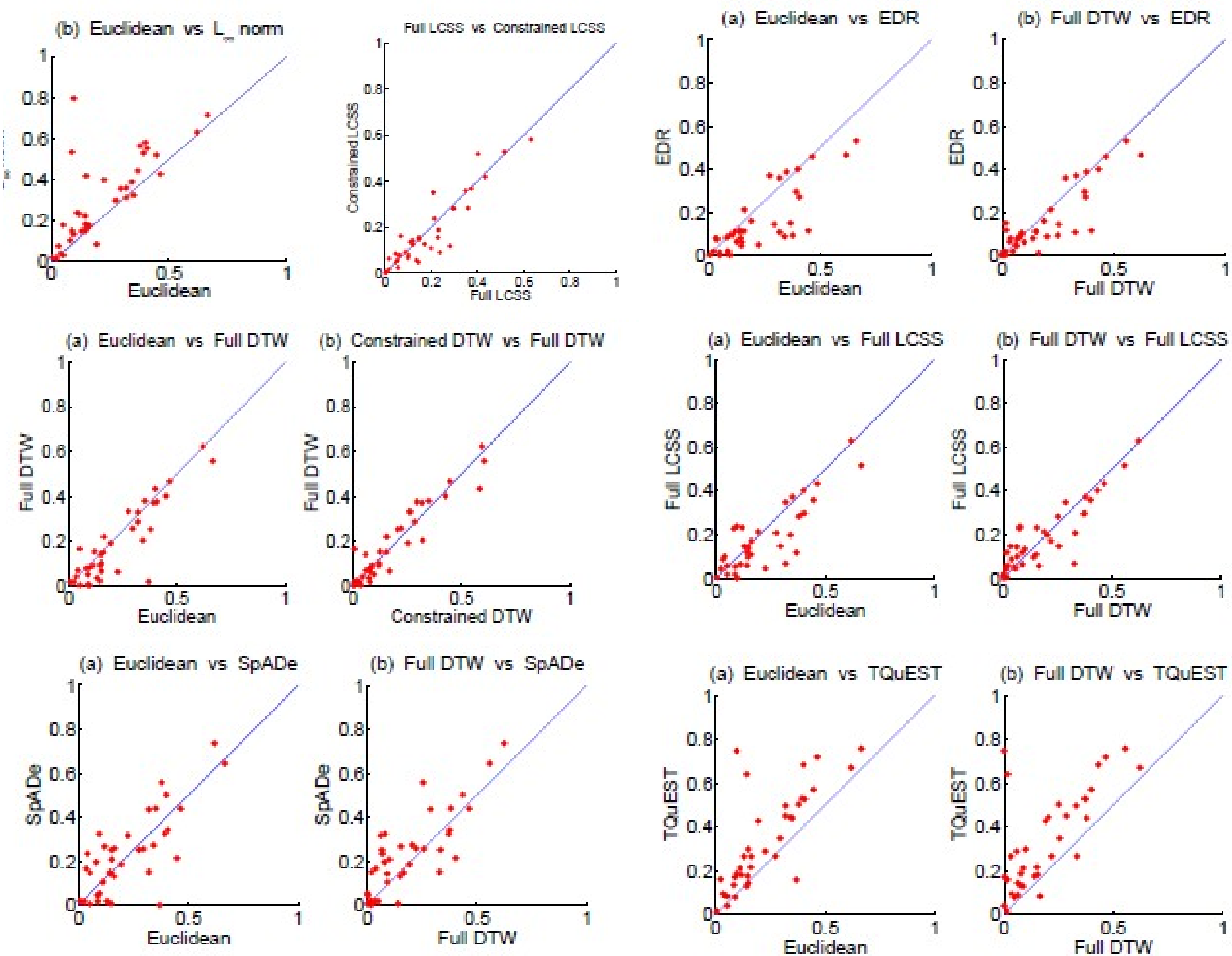
- Edit Distance on Real sequence (EDR)

$$EDR(R,S) = \begin{cases} n & \text{if } m = 0 \\ m & \text{if } n = 0 \\ min\{EDR(Rest(R), Rest(S)) + subcost, \\ \quad EDR(Rest(R), S) + 1, EDR(R, Rest(S)) + 1\} & \text{otherwise} \end{cases}$$

where $subcost = 0$ if $match(r_1, s_1) = true$ and $subcost = 1$ otherwise.



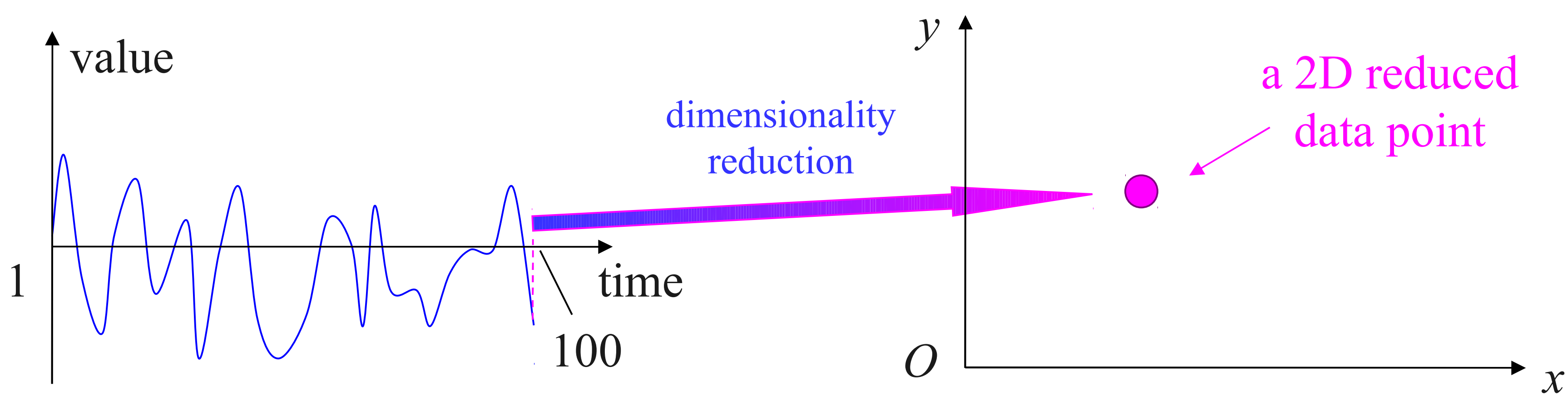
Comparison of Distance Measures



(Ding et al., 2008)

Dimensionality Reduction

- Since the dimensionality of time series is usually high, problems can arise
  - Similarity computations become expensive
  - *dimensionality curse*: query performances on multidimensional indexes degrades dramatically with the increasing dimensionality



a time series with dimensionality 100

a 2-dimensional reduced space

Dimensionality Reduction

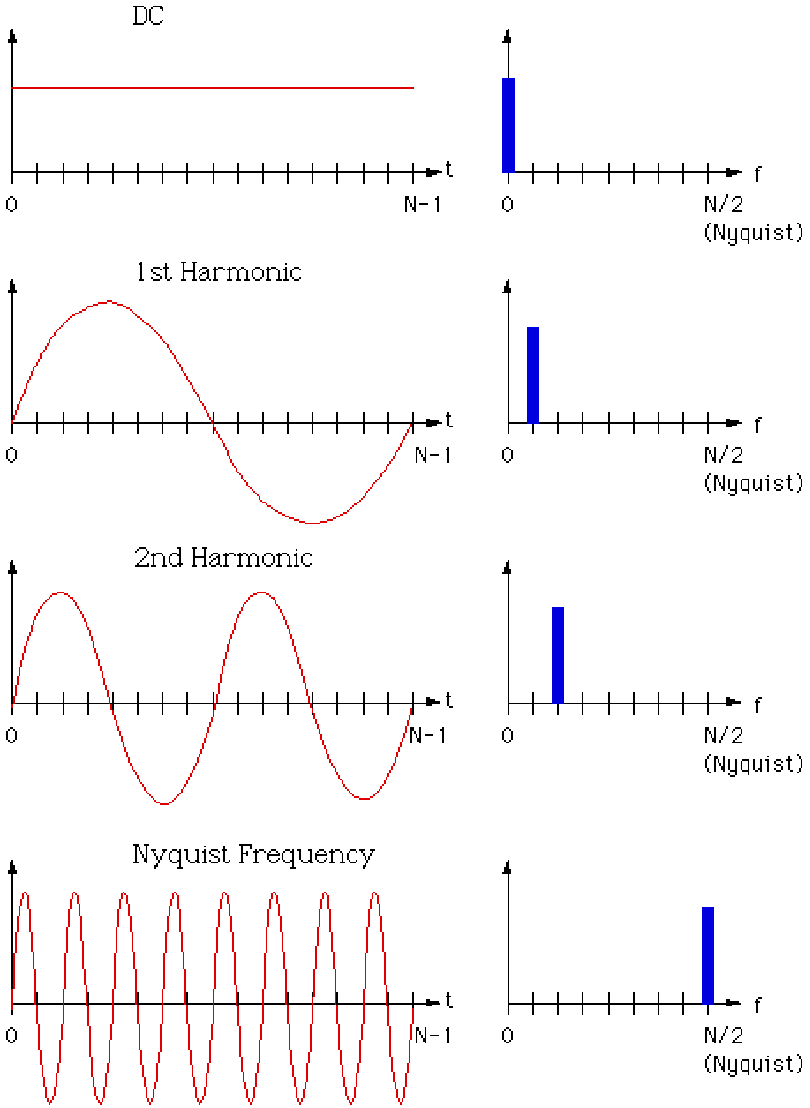
- Dimensionality reduction techniques:
  - *Singular Value Decomposition* (SVD)
  - *Discrete Fourier Transform* (DFT)
  - *Discrete Wavelet Transform* (DWT)
  - *Piecewise Aggregate Approximation* (PAA)
  - *Piecewise Linear Approximation* (PLA)
  - *Adaptive Piecewise Constant Approximation* (APCA)
  - *Chebyshev Polynomials* (CP)

Discrete Fourier Transform (DFT)

- Transform time series data from the time domain to frequency domain
  - x_t – time series data at timestamp t ($0 \leq t \leq n$ - 1)
  - X_f – transformed data
  - DFT:

$$X_f = 1/\sqrt{n} \sum_{t=0}^{n-1} x_t \exp\left(-j2\pi f t/n\right) \;\; f = 0, 1, \ldots, n-1$$

- Keep only the first k components



Contents

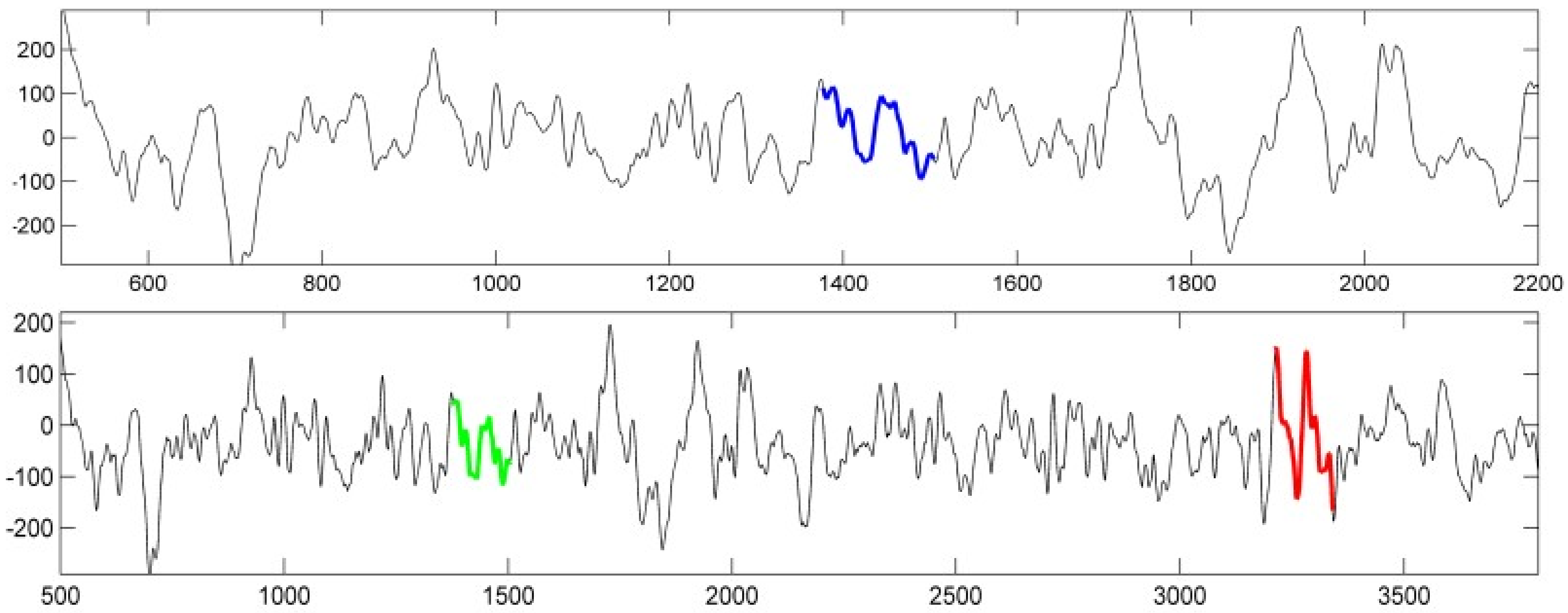
- Basics
  - Time series & preprocessing methods
- Simple trends models & cycles
  - Regression, seasonal and cyclic variations
- Autocorrelation models
  - ARIMA
- TS similarity
  - Euclidean, DWT, EDR, ERP
- **Patterns (Motifs)**

Motifs

- Patterns over a set of temporal sequences that for certain periods of time reflect a similar and/or a symmetric tendency
  - Corresponding intervals can be different in each TS
- Based on any time series similarity notion
- Given the similarity function between two subsequences X and Y, **sim(X,Y)**, X **matches** Y if sim(X; Y) > R where R is a user supplied positive real number

Motifs – example

- Sample pattern / motif



- Blue-Green & Blue-Red segments correspond
- Variants can consider also Green-Red segs.

Motifs

- Given
  - a database of time series ***D***
  - a minimum support ***minsupp*** and
  - a minimum value of similarity/correlation ***Rmin***

  a set ***S*** of subsequences (taken from D) form a **(approximate) motif**, if
  - |S| > minsupp
  - For all pairs (X,Y) from S, sim(a,b)> Rmin

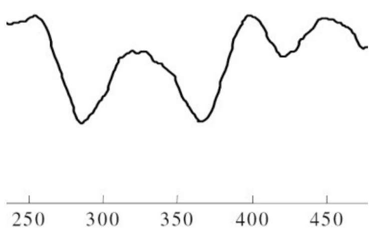
Motif extraction algorithms

- Often they are based on simplified representations of the data, in order to improve performances
  - Dimensionality reduction techniques
  - Discretization of the TS
    - SAX representation of time series

SAX: Symbolic Aggregate approXimation
Dim. Reduction/Compression

- "Symbolic Aggregate approXimation"

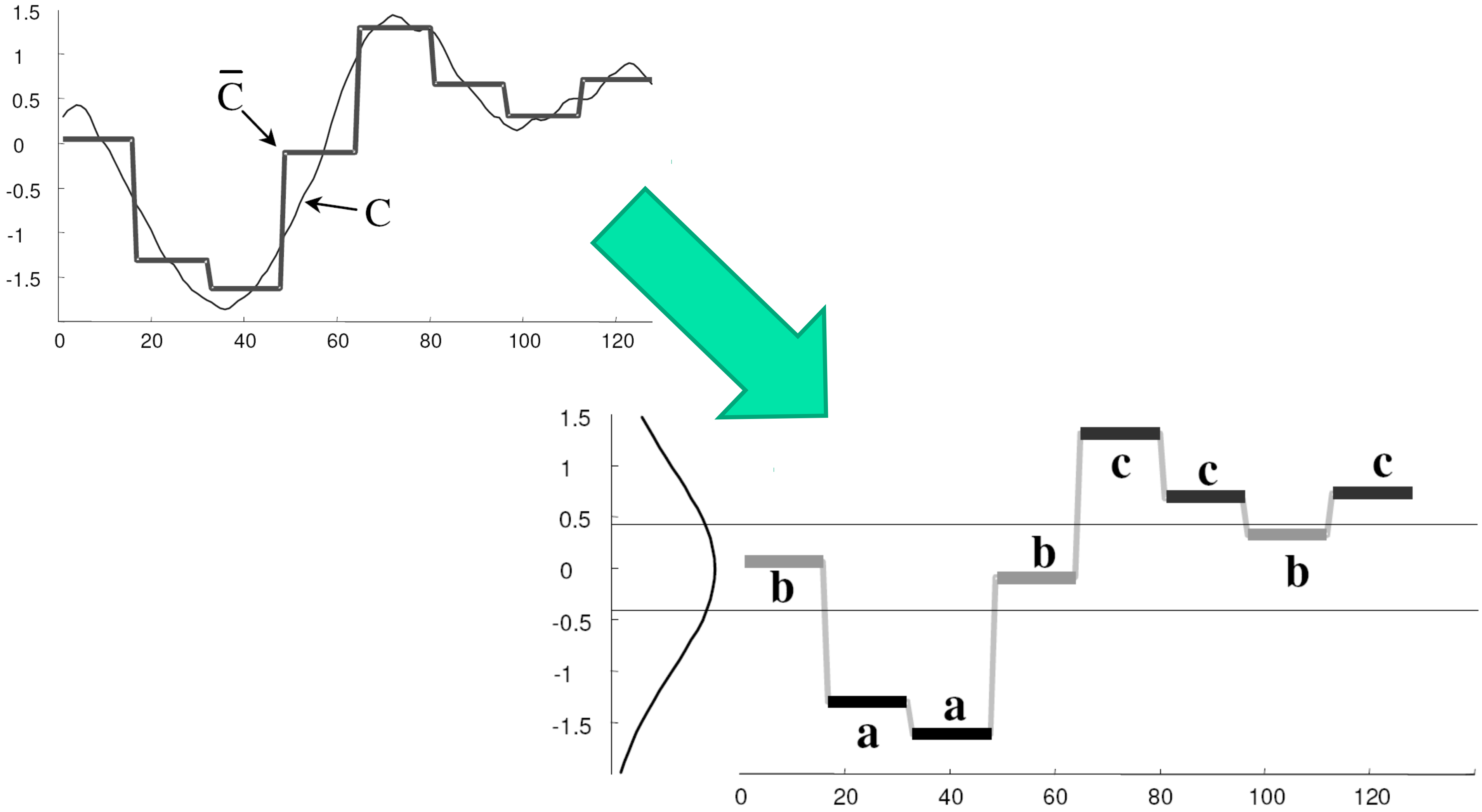
  SAX : $\mathbb{R} \rightarrow \Sigma$

  SAX : 

  $\mapsto$ ccbaabbbabcbcb

- Essentially an alphabet over the Piecewise Aggregate Approximation (PAA) rank
- Faster, simpler, more compression, yet on par with DFT, DWT and other dim. reductions

SAX Illustration



SAX Algorithm

Parameters: alphabet size, word (segment) length (or output rate)

1. Select probability distribution for TS
2. z-Normalize TS
3. PAA: Within each time interval, calculate aggregated value (mean) of the segment
4. Partition TS range by equal-area partitioning the PDF into n partitions (eq. freq. binning)
5. Label each segment with $a_{rank} \in \sum$ for aggregate's corresponding partition rank

Finding Motifs in a Time Series

EMMA Algorithm: Finds motifs of fixed length n

- SAX Compression (Dim. Reduction)
  - Possible to store $D(i,j)\ \forall (i,j) \in \Sigma \times \Sigma$
  - Allows use of various distance measures (Minkowski, Dynamic Time Warping)
- Multiple Tiers
  - Tier 1: Uses **sliding window** to **hash** length-w SAX subsequences (a^w addresses, total size O(m)).
    Bucket B with most **collisions** & buckets with **MINDIST**(B) < R form **neighborhood** of B.
  - Tier 2: Neighborhood is **pruned** using more precise **ADM algorithm**. N_i with max. matches is 1-motif. **Early stop** if |ADM matches| > $\max_{k>i}$ ($|\text{neighborhood}_k|$)