# Anomaly Detection
# Data adaptation: time series

Paul Irofti
Cristian Rusu
Andrei Pătraşcu

Computer Science Department
University of Bucharest

# Class structure

Topics for today:

- discuss characteristics of time series
- change-point model for time series
- average and linear models for time series
- anomaly detection

What is a time series?
For us, in this class: time series = vector of real values + time stamp

They appear everywhere where a phenomenon is monitored:

- finance (performance indicator measurements)
- healthcare (vital sign measurements)
- industry (sensor measuments)
- ...

Two questions that will interest us today:

- do time series suffer significant changes over time?
- is there something anomalous in the time series?

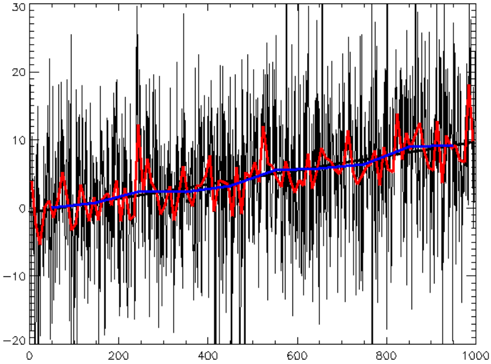
What is the first thing that comes to mind when trying to work with time series?

What is the first thing that comes to mind when trying to work with time series?

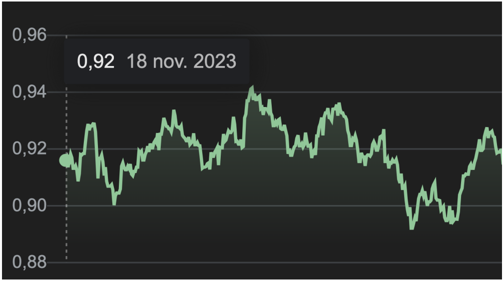
- seems to be a 1D regression problem (if we ignore the time information)
  - how can we turn in into a regression problem and keep the time information?
- the regression problem can be extended into multiple dimensions (feature engineering)
- average a couple of values from the past to try to predict new values (in the style of K-NN)
- try to find seasonal components in the data (peridicity analysis - Fourier)

# Time series characteristics

A typical time series composed of random data plus a linear trend (source: Wikipedia)



Another typical time series Dollar vs. Euro exchange (source: Google)



Another typical time series Dollar vs. Euro exchange, with an orange change point (source: Google)

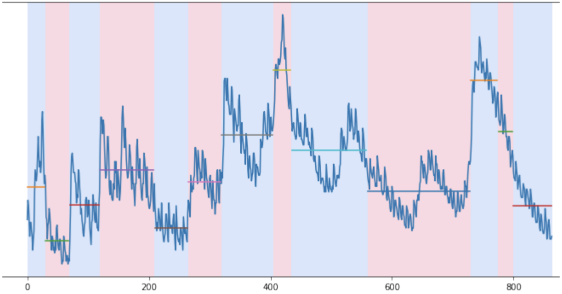

time series like this are not non-stationary

what would linear regression look like on this data?

# The end.

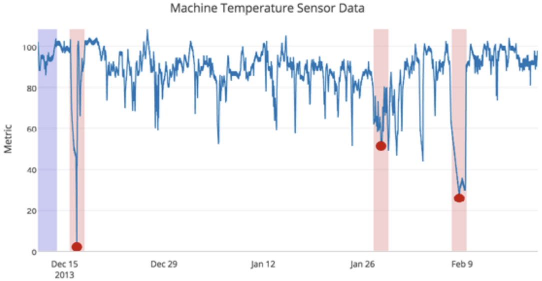
Problem: given a time series, retrieve $K$ points in the time series where a significant change occurs (source: US unemployment data)

- find how many change points there are
- tell us where these change points are



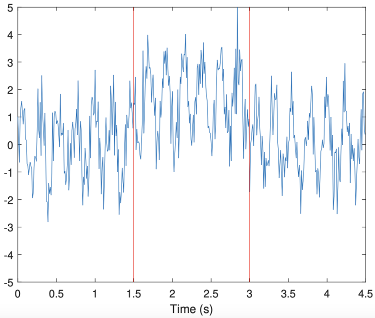
Problem: given a time series, retrieve the points where something unusual happens

- what is the definition of unusual?
- how many unusual points are you looking for?



Goals:

- find the abrupt changes in the time series $x[n]$
- find the time at which these happen
- find how many there are
- we call the set of times where an abrupt change happens $\mathcal{T}^\star$



Problem statement

$$(\hat{t}_1, \ldots, \hat{t}_K) = \underset{t_1, \ldots, t_K}{\arg\min} \sum_{k=1}^{K} c(x[t_k : t_{k+1}]) \tag{1}$$

We have made the following notation:

- $t_k : t_{k+1}$ is Matlab notation for the set $\{t_k, t_k + 1, \ldots, t_{k+1} - 1\}$
- $c$ is a cost function that measures homogeneity
- what are some good picks for the cost function $c$?

Problem statement

$$(\hat{t}_1, \ldots, \hat{t}_K) = \underset{t_1, \ldots, t_K}{\arg\min} \sum_{k=1}^{K} c(x[t_k : t_{k+1}]) \tag{1}$$

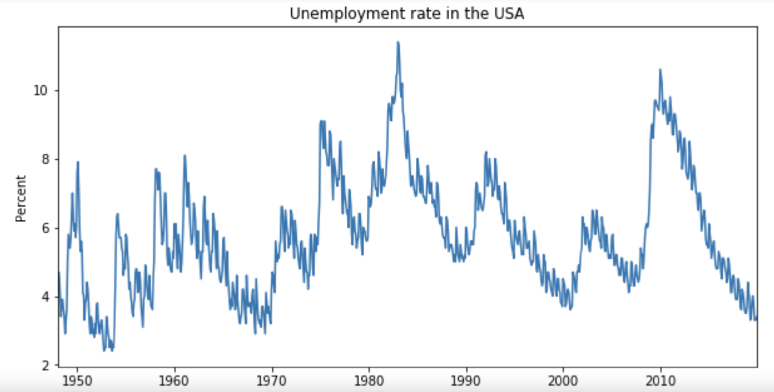
We have made the following notation:

- $t_k : t_{k+1}$ is Matlab notation for the set $\{t_k, t_k + 1, \ldots, t_{k+1} - 1\}$
- $c$ is a cost function that measures homogeneity
- what are some good picks for the cost function $c$?
  - log likelihood term
  - the mean
  - the median
  - the error (RMSE) of a linear model
  - the error (RMSE) of a more sophisticated linear model

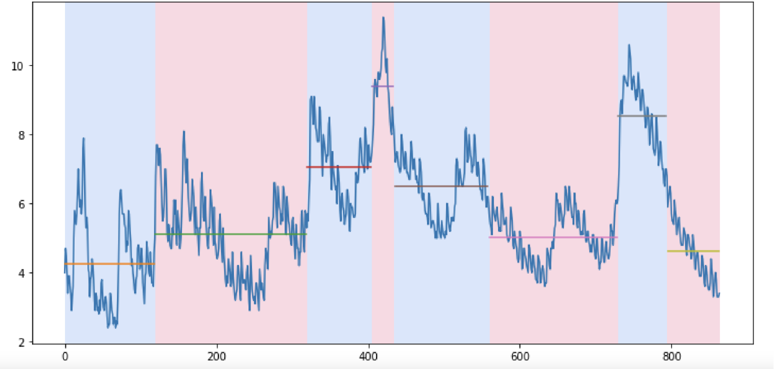
Choices for the cost function:

1. $c_{\text{prob}}(t_k : t_{k+1}) = -\max_{\theta} \sum_{n=t_k}^{t_{k+1}-1} \log p(x[n]|\theta)$
2. $c_{\text{L2}}(t_k : t_{k+1}) = \sum_{n=t_k}^{t_{k+1}-1} \|x[n] - \mu_{t_k:t_{k+1}}\|_2^2$
3. $c_{\Sigma}(t_k : t_{k+1}) = (b-a)\log \sigma_{t_k:t_{k+1}}^2 + \frac{1}{\sigma_{t_k:t_{k+1}}^2} \sum_{n=t_k}^{t_{k+1}-1} \|x[n] - \mu_{t_k:t_{k+1}}\|_2^2$
4. $c_{\text{lin}}(t_k : t_{k+1}) = \min_{\alpha} \sum_{n=t_k}^{t_{k+1}-1} \|x[n] - \sum_{i=1}^{M} \alpha_i \beta_i[n]\|_2^2$

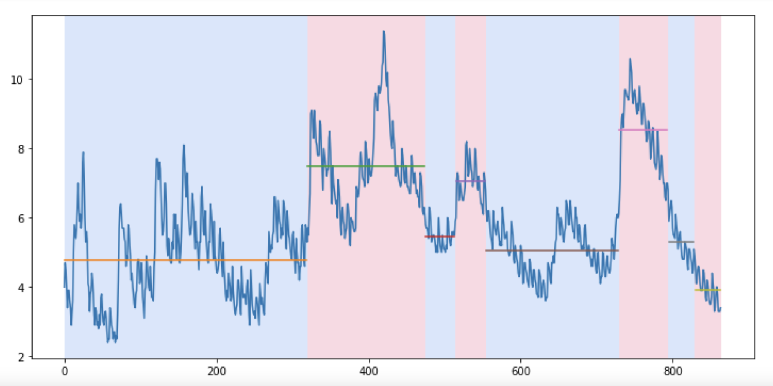
## US Unemployment



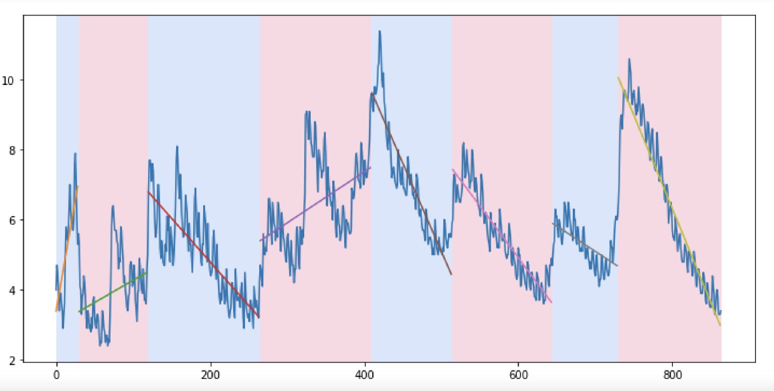
US Unemployment - $c_{L2}, K = 7$



US Unemployment - $c_\Sigma, K = 7$



US Unemployment - $c_{\text{lin}}, K = 7$



Solve optimally by dynamic programming

For

$$\mathcal{V}(\mathcal{T}, \mathbf{x}) = \sum_{k=0}^{K} c(x[t_k : t_{k+1}])$$

we have that

$$\begin{aligned}
\min_{|\mathcal{T}|=K} V(\mathcal{T}, \mathbf{x}) &= \min_{0=t_0<t_1<\cdots<t_K<t_{K+1}=N} \sum_{k=0}^{K} c(x[t_k : t_{k+1}]) \\
&= \min_{t \leq N-K} \left[ c(x[0 : t]) + \min_{t_0=t<t_1<\cdots<t_{K-1}<t_K=N} \sum_{k=0}^{K-1} c(x[t_k : t_{k+1}]) \right] \\
&= \min_{t \leq N-K} \left[ c(x[0 : t]) + \min_{|\mathcal{T}|=K-1} V(\mathcal{T}, x[t : N]) \right]
\end{aligned}$$

Complexity?

Solve optimally by dynamic programming

For

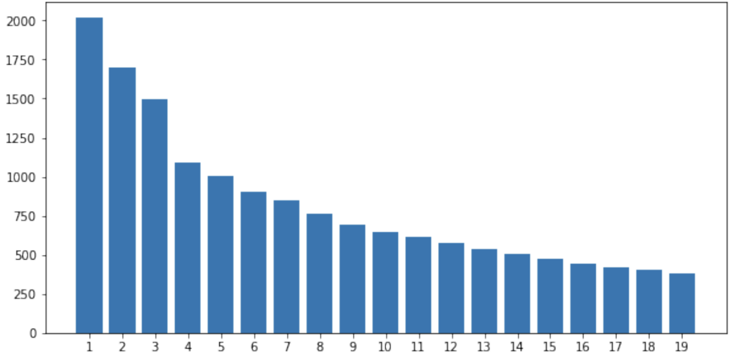
$$\mathcal{V}(\mathcal{T}, \mathbf{x}) = \sum_{k=0}^{K} c(x[t_k : t_{k+1}])$$

we have that

$$\begin{aligned}
\min_{|\mathcal{T}|=K} V(\mathcal{T}, \mathbf{x}) &= \min_{0=t_0<t_1<\cdots<t_K<t_{K+1}=N} \sum_{k=0}^{K} c(x[t_k : t_{k+1}]) \\
&= \min_{t \leq N-K} \left[ c(x[0 : t]) + \min_{t_0=t<t_1<\cdots<t_{K-1}<t_K=N} \sum_{k=0}^{K-1} c(x[t_k : t_{k+1}]) \right] \\
&= \min_{t \leq N-K} \left[ c(x[0 : t]) + \min_{|\mathcal{T}|=K-1} V(\mathcal{T}, x[t : N]) \right]
\end{aligned}$$

Complexity? $O(KN^2)$

- run for all values $K$ from 1 to $K_{max}$
- find an “elbow” in the resulting curve




how would you integrate the $K$ into the problem itself?

New, regularized problem statement (penalized change point detection)

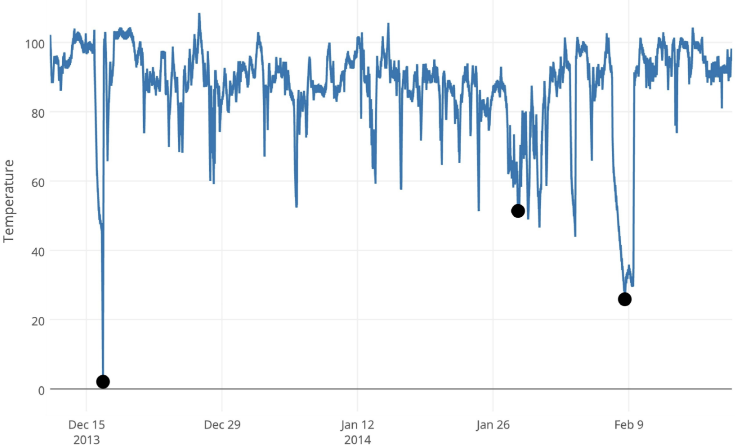
$$(\hat{t}_1, \ldots, \hat{t}_K) = \underset{t_1, \ldots, t_K}{\arg\min} \sum_{k=1}^{K} c(x[t_k : t_{k+1}]) + \lambda K \quad (2)$$

- this type of regularization is typical in machine learning
- the size of the solution set $\mathcal{T}$ is taken into account at each step of the algorithm
- many algorithms have been proposed for this task
- new problem: find $\lambda \in \mathbb{R}_+$ (in general there is no clear formula between $\lambda$ and $K$)

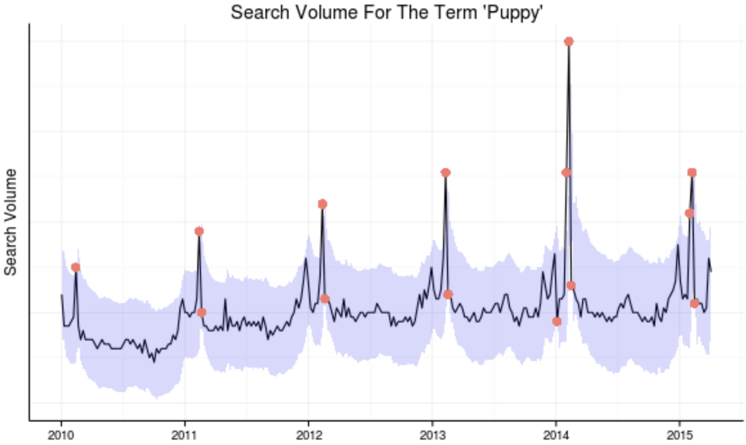
After we have the change point splitting done we can do several things:

- for the mean and statistical cost functions $c_{\mathrm{L2}}$ and $c_{\Sigma}$ we can use methods developed in Lecture 5 Statistical algorithms: truncation, LODA
- for the regression (linear) statistical cost function $c_{\mathrm{lin}}$ we can use the leverage scores developed in Lecture 2 Leverage scores for linear regression
- the third option is to use an adaptive model that changes with the time series

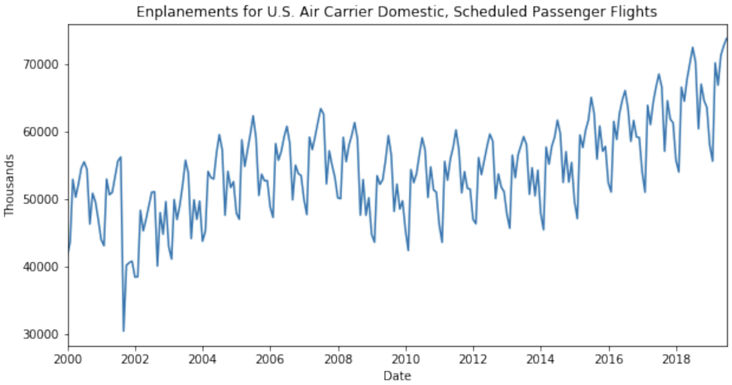
## Easy example



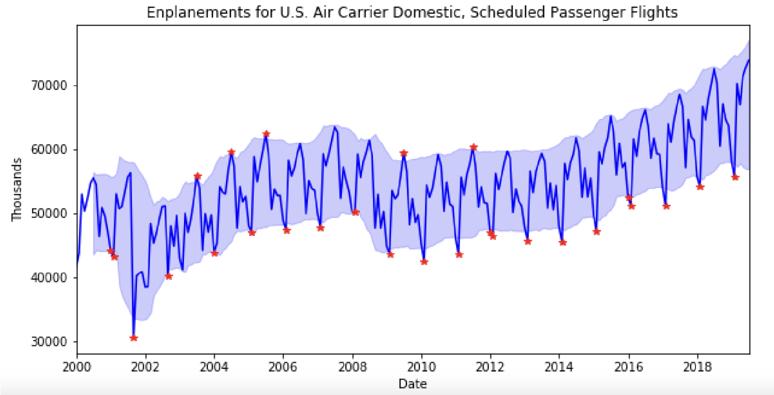
## Easy example



## Easy example

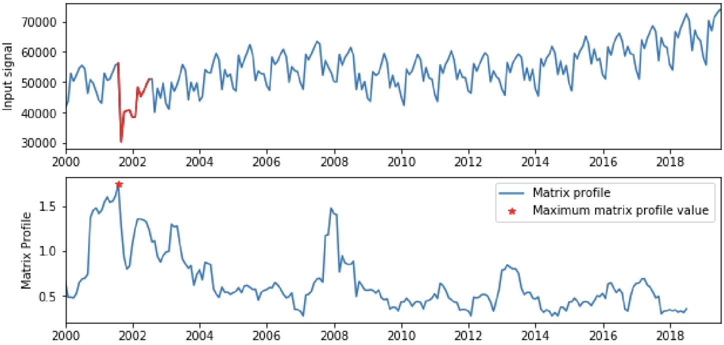


Mu-sigma model: $|x[n] - \mu| > \lambda\sigma$ ($\lambda = 1.5$ window of size 12)

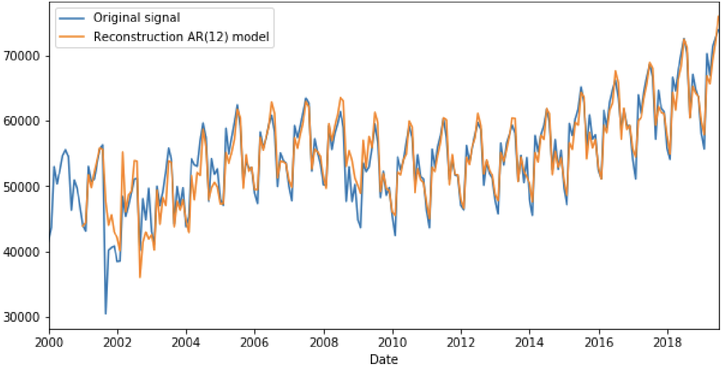


GIven a chunk of size $L$ try to find in the time series a similar pattern

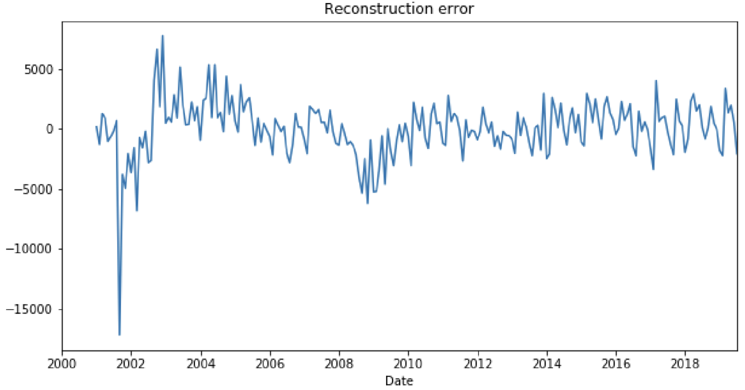
$$\text{pattern}[n] = \min d(x[n : n+L-1], x[i : i+L-1]) \text{ for } i < n-L \text{ and } i > n+L \quad (3)$$



Model based: AR model reconstruction



Model based: AR model error



# The end.