“ONLY TIME WILL TELL”

Methods and Tools for Time Series

Bl’ast!

Luca Pozzi
Radius Intelligence

ABOUT ME

- Escaped from Alcatraz multiple times
- MS Applied Mathematics
- PhD Biostatistics U.C. Berkeley
- Independent Consultant (European Union, Novartis Pharma, FAO, etc.)
- Radius Intelligence

RADIUS

- **Database** of small & medium business data.
- Create **Intuitive Grouping** of marketing targets.
- Deal with **Incomplete Records**.

TODAY

Some more time series

OUTLINE

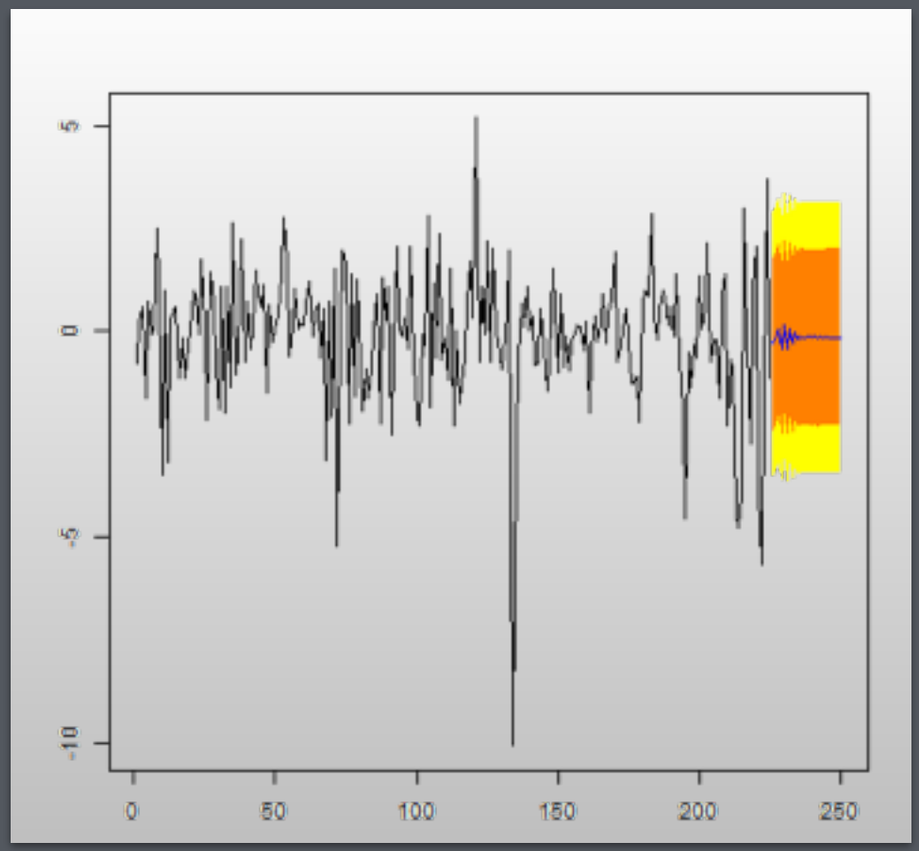
From Statistics to ML:

1. **Forecasting**
2. **Model Selection**
3. **Segmentation**
4. **Featurization, Grouping** and **Anomalies**

FORECASTING

What comes Next?

Solution: Project along a regression curve.



ARMA MODELS

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

$$X_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

SOFTWARE

Mainly **R** packages (but also DIY…):

- **STL** Seasonal-Trend Decomposition.
- **forecast** contains lots of useful methods.
- Base **R!**

MODEL SELECTION

How do we pick the Best Model?

A Pessimist Approach to Complex Data:

"The Bane of Statistical Learning"

Non iid by definition: Standard Cross Validation doesn't work…

CROSS VALIDATION

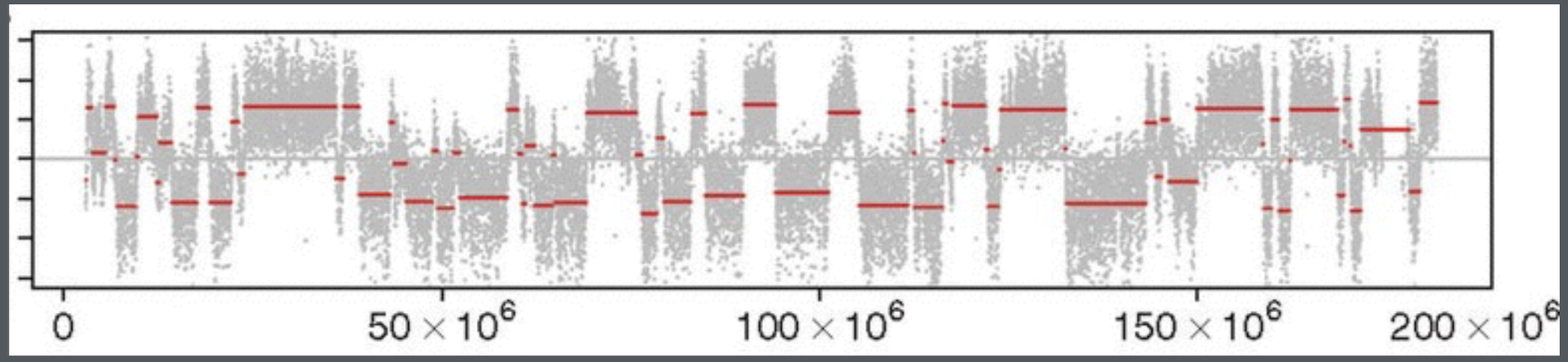
5 Fold CV Example:

1. **Train on Fold 1 Test on Fold 2**
2. **Train on Folds 1,2 Test on Fold 3**
3. **Train on Folds 1,2,3 Test on Fold 4**
4. **Train on Folds 1:4 Test on Fold 5**

Preserve the time dependency…

SEGMENTATION

Identify meaningful intervals…

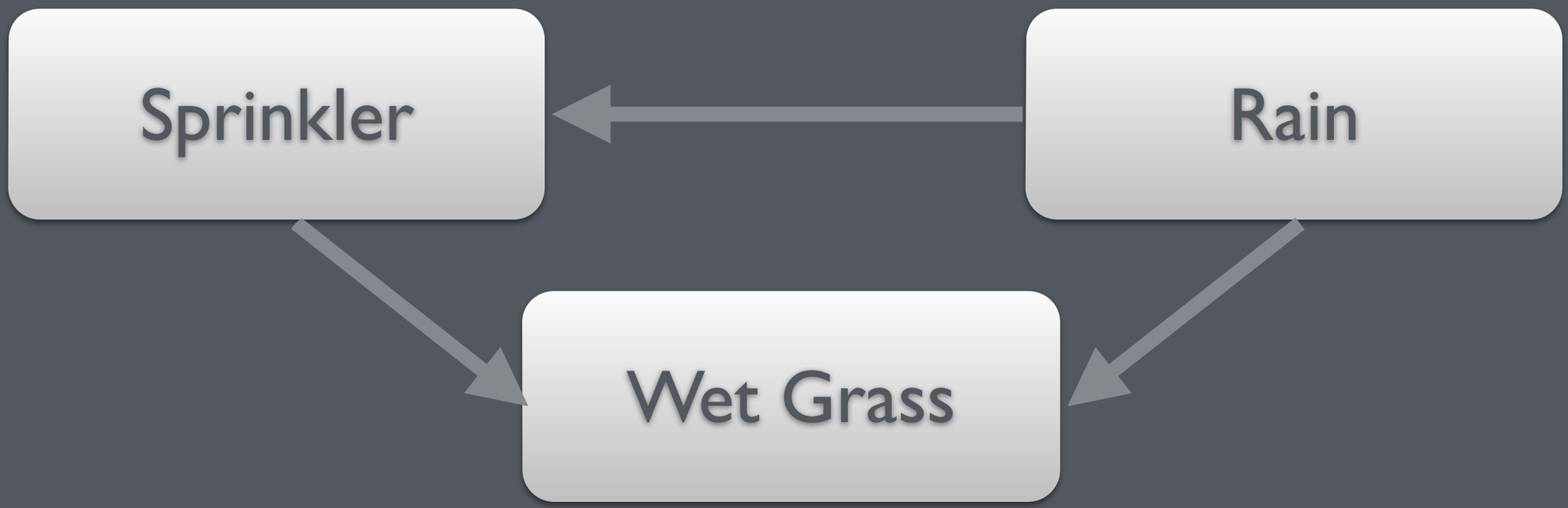


Solution: Modeling the process…

GRAPHICAL MODELS

Modeling the **joint distribution** with a graph:

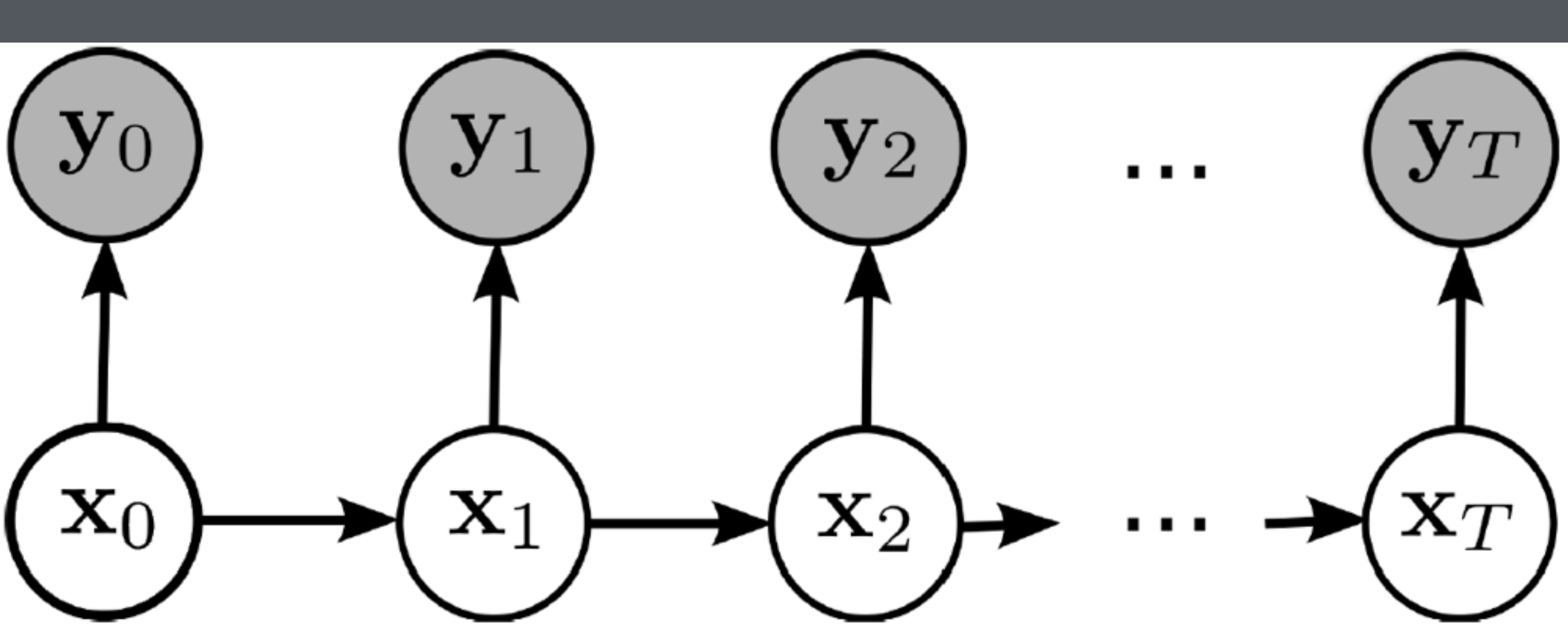
Edges = Dependencies



HIDDEN MARKOV MODELS

Markov Assumption (Chain Graph):

"Future depends on the past only through the present"



HMM CHARACTERISTICS

Generative Model

Models the joint distribution.

Latent Variable:

time dependent mixture model.

Unsupervised:

we don't observe the latent state.

HMM LEARNING

What are we trying to Learn?

- **Sequence of Hidden States:**
  (Segmentation) Baum Welch - Viterbi

- **Transitions Probabilities:** (Forecasting)
  Baum Welch (EM)

HMM EXAMPLES

- Decoding transmission over a channel.
- Anomaly Detection
- Draughts analysis
- Behavioral metrics (e.g. Addiction)
- Personal Analytics
- Genetics

Also NHMM, PHMM, etc.

HMM SOFTWARE

- **hmmlearn** formerly in **sklearn**
- **seqlearn** also familiar interface..
- **GHMM** and lots of others…

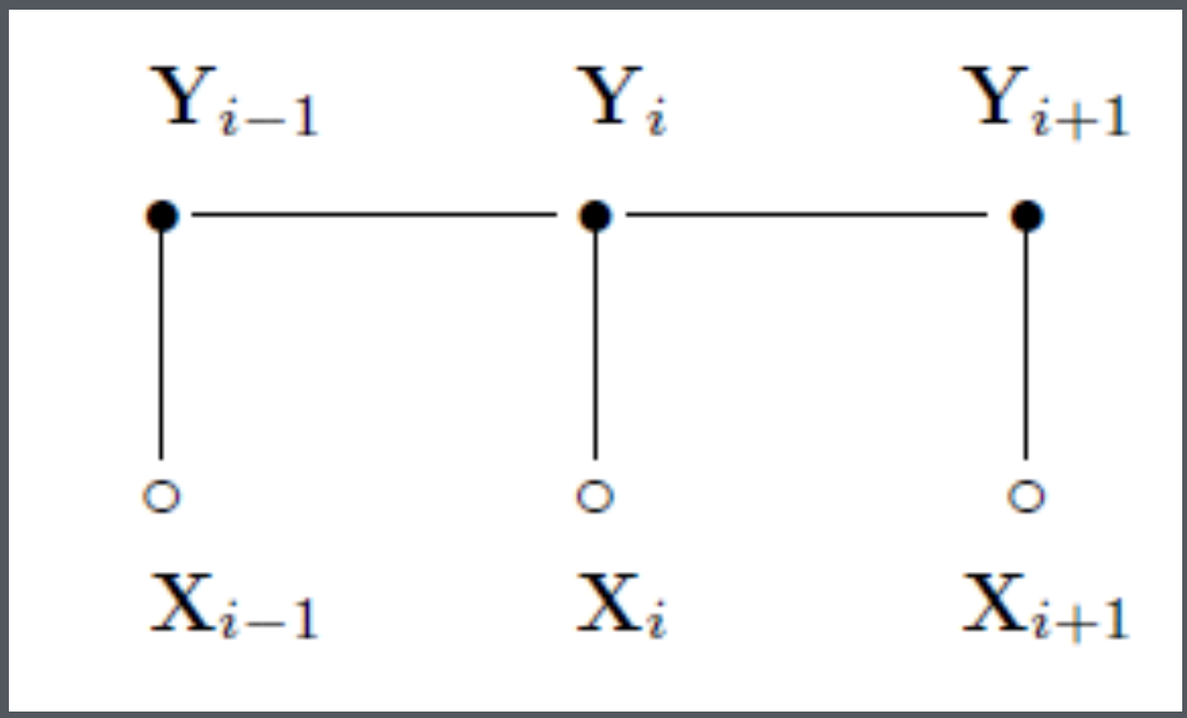
CONDITIONAL RANDOM FIELDS

Discriminative Model

Models the posterior distribution.

Supervised:

need to know the label (but this can be relaxed…)



CRF LEARNING

The Joint Likelihood is factorized along the edges and Gradient Descent is used to estimate the coefficients for each factor.

$$p_\theta(\mathbf{y} \mid \mathbf{x}) \propto \qquad (1)$$
$$\exp\left(\sum_{e \in E,k} \lambda_k \, f_k(e, \mathbf{y}|_e, \mathbf{x}) + \sum_{v \in V,k} \mu_k \, g_k(v, \mathbf{y}|_v, \mathbf{x})\right) \ ,$$

CRF: EXAMPLES

- NLP Part of Speech Tagging
- Image Segmentation
- Anomaly Detection

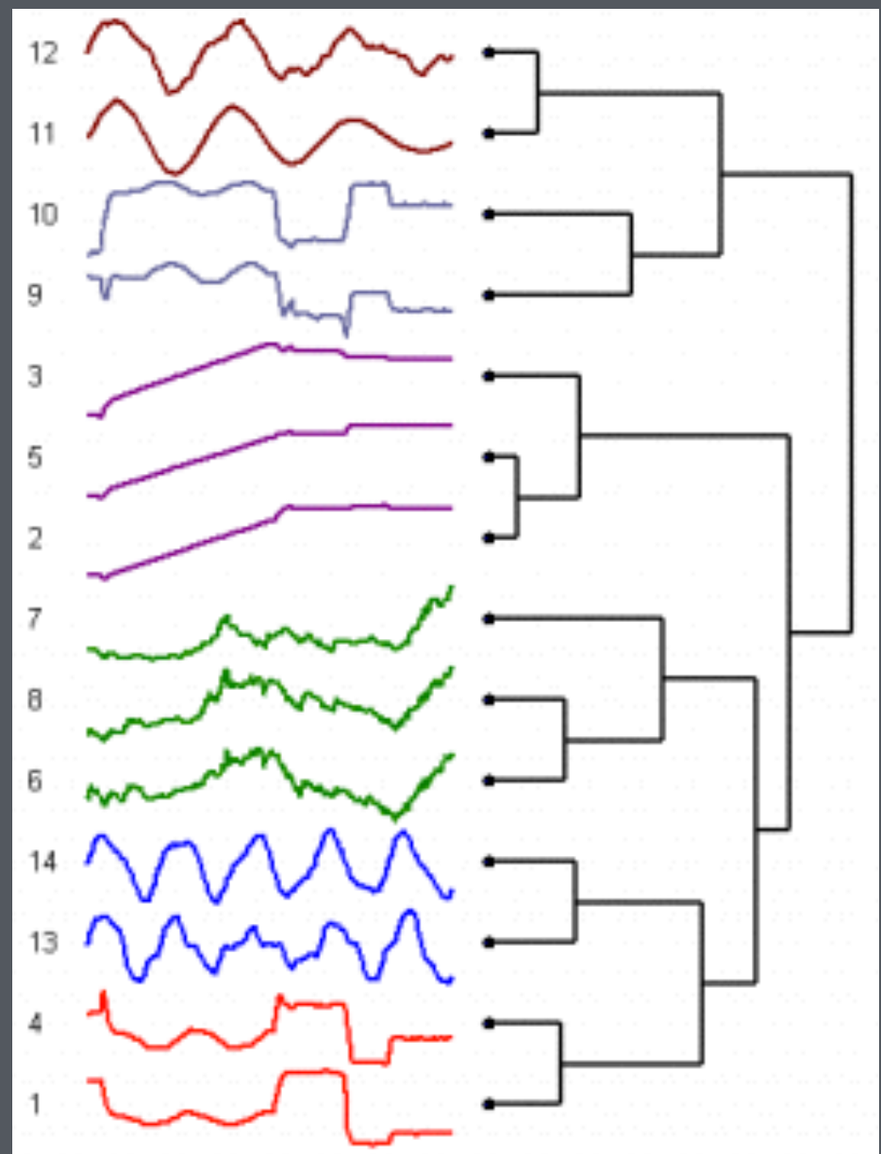
CRF SOFTWARE

- **Factorie** Scala library by Mc Callum et al.
- **Mallet** Java library, predecessor of Factorie
- **NLTK** has a CRF tagger…

CHARACTERIZATION

How to Compare and Group Series?

Solution: treat them as general objects to do Machine Learning on…



KERNEL MACHINE

The Power of SVM is to define a common framework for very different data types:

Kernel for Time Series!

Defined using the Cross Correlation
O(N log(N)) using FFT!

SIMILARITIES

- Based on the Kernel:
  $d(x,y) = K(x,x) + K(y,y) - 2\,K(x,y)$

- DTW
- Longest Common Subsequence
- Early Abandon Euclidean

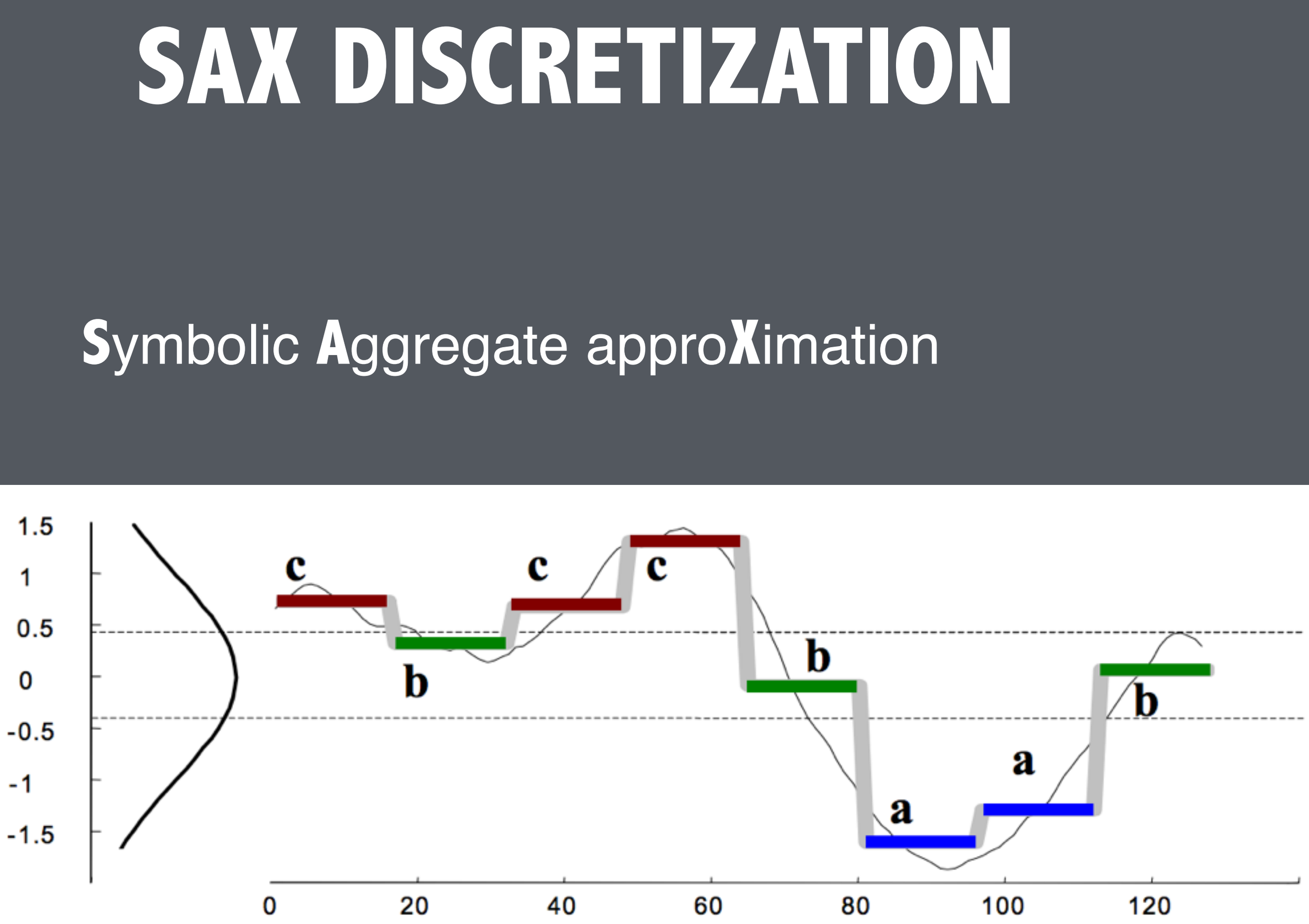
FEATURES

- Trend
- Seasonality
- Non-Linearity
- Moving Window Summaries
- etc.

Can be fed to Clustering Algorithms!

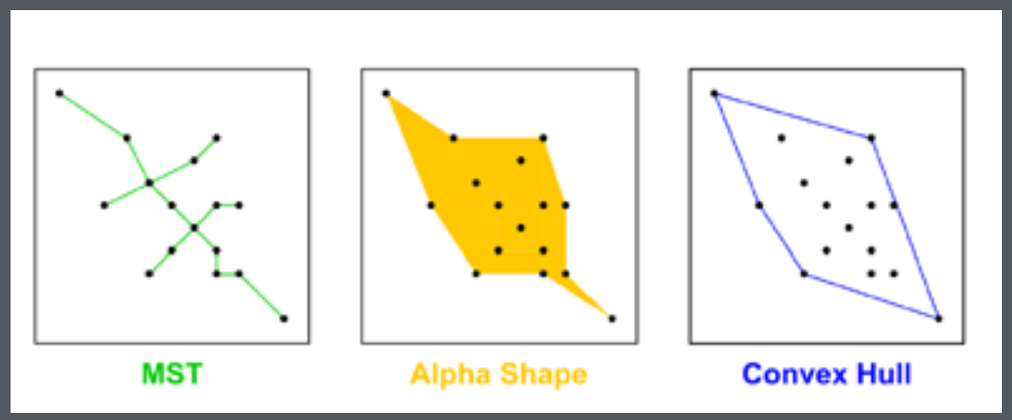
SAX DISCRETIZATION

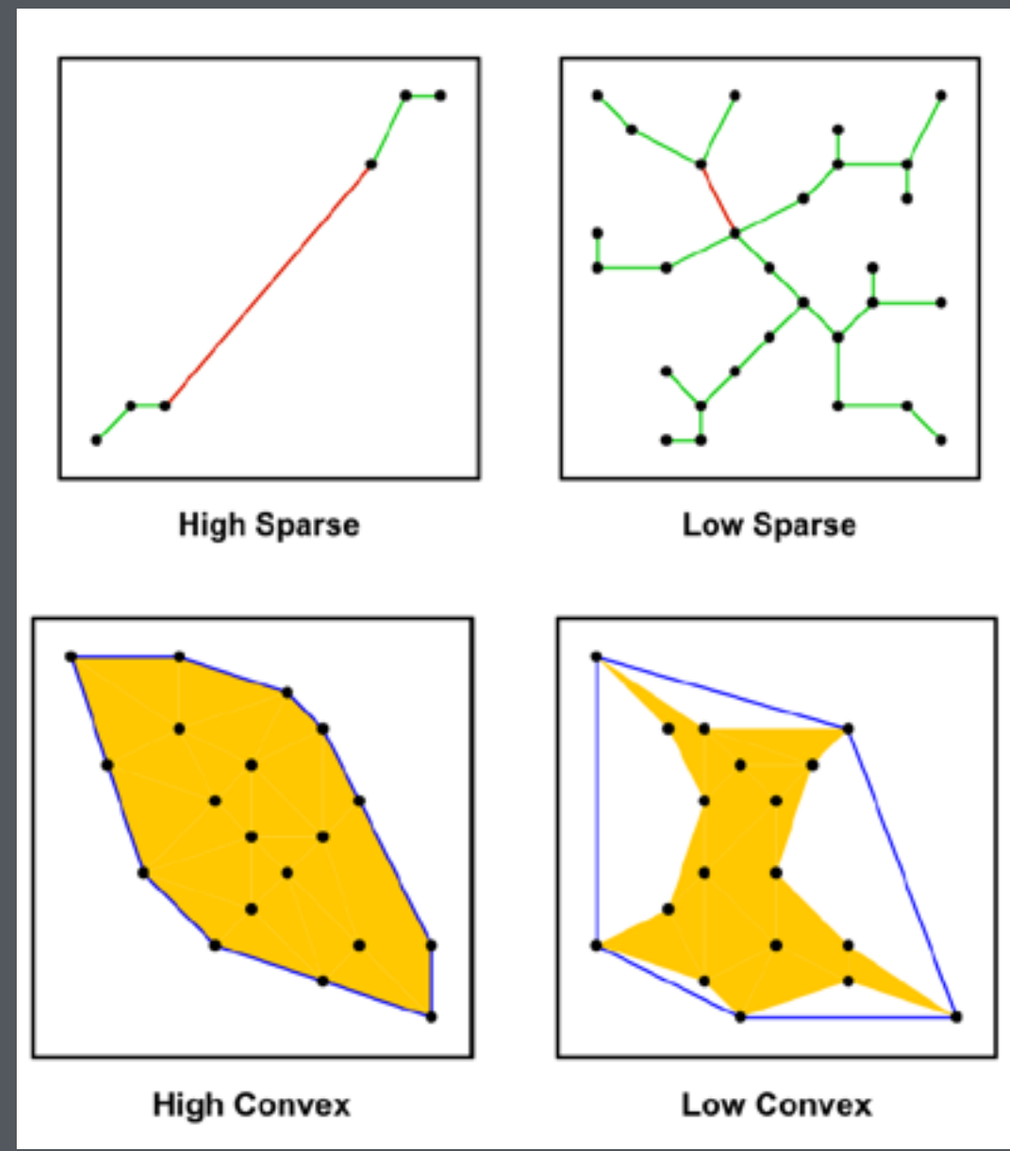
Symbolic **A**ggregate appro**X**imation



SCAGNOSTIC

Create **Convex Hull, Alpha Shape** and **Minimum Spanning Tree**:





FINAL REMARKS

- **DO NOT** treat time Series like any other Data: Be Careful…
- **DO** treat time Series like any other Data: Plenty of sophisticated tools…
- Time Series are not just about Time…

SOME REFERENCES

- E. Keogh's website http://www.cs.ucr.edu/~eamonn
- R. Hyndman (2006) Characteristic-based clustering for time series data
- L. Wilkinson (2012) Timeseer: Detecting interesting distributions in multiple time series data
- A. McCallum (2001) Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data
- G. Wachman (2009) Kernels for Periodic Time Series Arising in Astronomy

THANK YOU

Luca Pozzi

poz.luc@gmail.com