DS-GA 1018.001
Probabilistic time series analysis
Lecture 1
Logistics. Introduction to probabilistic time series analysis

Instructor: Cristina Savin

NYU, CNS & CDS

Course logistics

Instructor
Cristina Savin, csavin@nyu.edu
Office hours: TBD, in person

TAs
- Section 002 - Haresh Rengaraj Rajamohan (hrr288@nyu.edu)
- Section 003 - Ying Wang (yw3076@nyu.edu)

Office hours: TBD via zoom

Couse page: https://github.com/savinteachingorg/pTSAfall2021.git
Piazza: piazza.com/nyu/fall2021/dsga3001001
!!! **access code** pTSA21

[Quick feedback much appreciated. (anonymous)]

Course overview

This graduate level course covers fundamental probabilistic models for characterizing data with dependencies over time, and their use for predicting future outcomes. The methods covered have broad applications from econometrics to neuroscience.

Course aims:

- Understanding statistical assumptions made by different models so as to be able to reason about the applicability of different tools to a given dataset
- Being able to derive from scratch fundamental algorithms for inference and learning, including Kalman filtering/smoothing, alpha-beta for HMMs, EM, basic back propagation through time
- Being able to implement in python all these fundamental algorithms
- Developing a clear overview of how different models relate to one another and how the fundamental models can be generalized to capture more complex statistical dependencies in the data

Prerequisites: probability and **linear algebra**

If haven's seen any recently, a quick refresher is strongly encouraged

Course logistics: grading

5 problem sets 25%
Primarily derivations, 2w for each, top scoring 4

Lab work 20%
Coding: **python**, weekly, ideally finished during lab
Best scoring 7

Midterm 20%
Nov 8th, ARIMA+ LDS+HMMs

Project 25%
Groups of 2-3, topic of choice
Project proposal due **Oct. 22nd**

Replication challenge RC2021!

Participation 10%
Class discussions, office hours, piazza

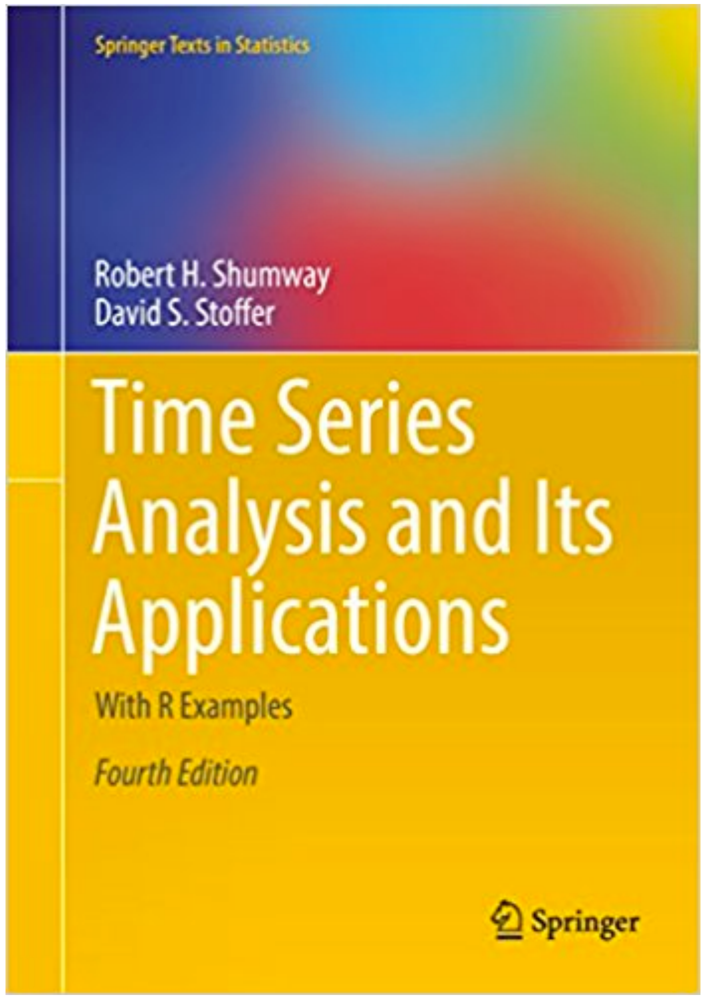
Academic honesty

We expect you to try solving each problem set on your own. However, if stuck you should discuss things with other students in the class, subject to the following rules:

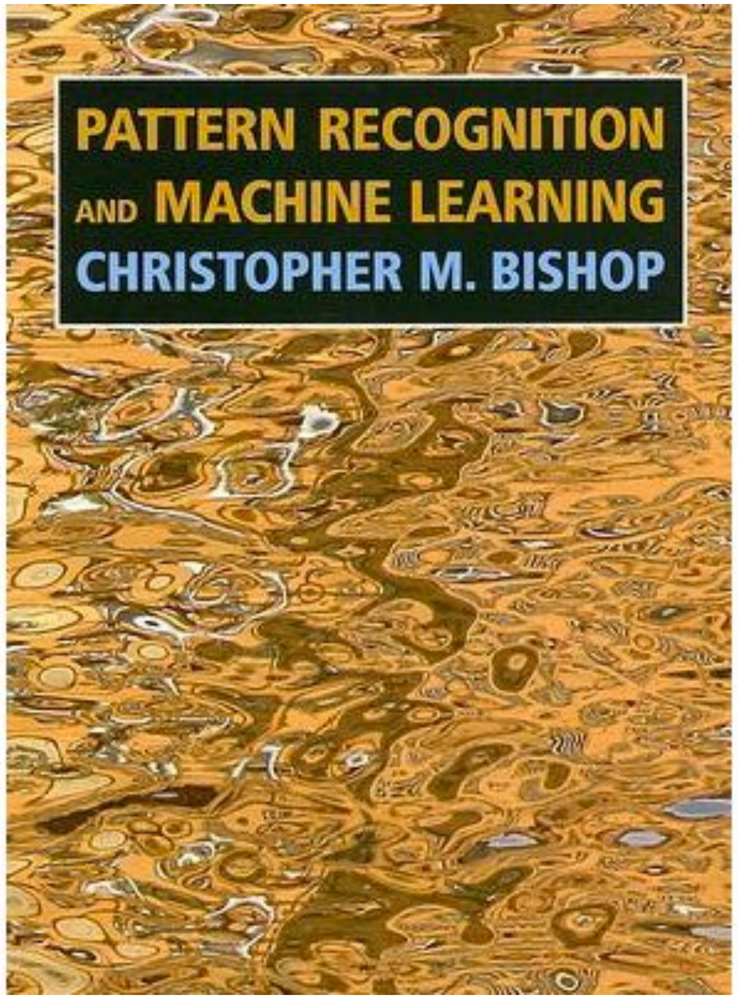
- Brainstorming and verbally discussing the problem with other colleagues ok, going together through possible solutions, but should not involve one student telling another a complete solution.
- Once you solve the homework, you must write up your solutions on your own.
- You must write down the names of any person with whom you discussed it. This will not affect your grade.
- Do not consult other people's solutions from similar courses.
- Credit should be explicitly given for any code you use that you did not write yourself.
- Violations result in a zero score on that assignment, and a notice to the DGS.

Bibliography

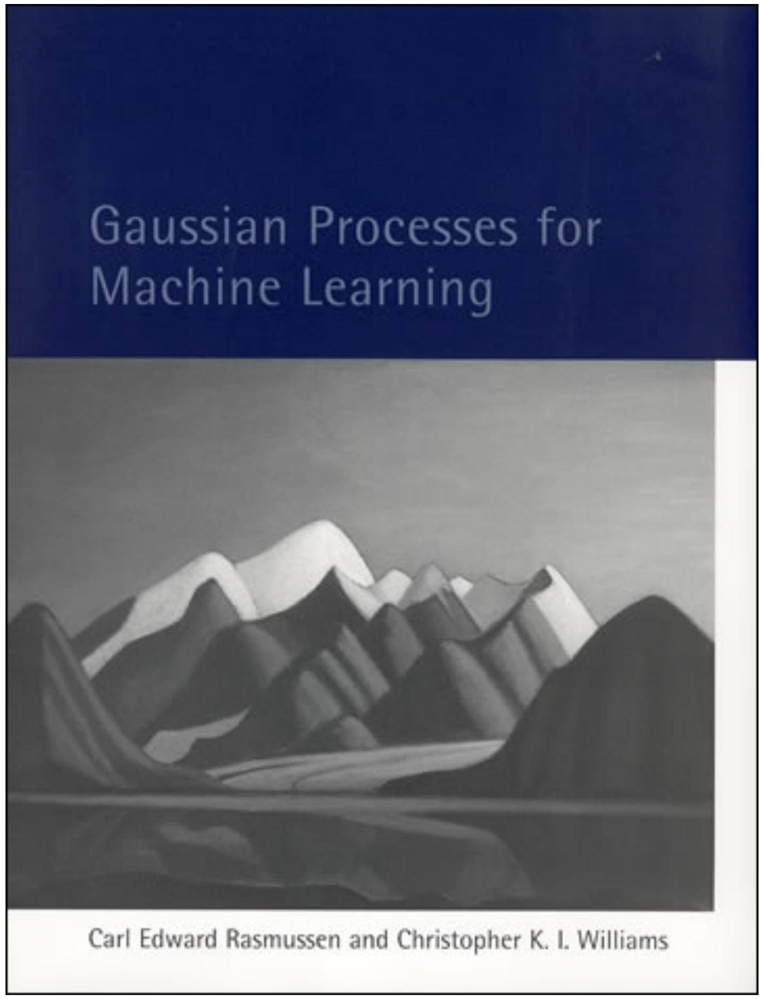
No course textbook, handouts for each section. Lectures based on:



Basics, AR(I)MA, Spectral methods



Latent space models

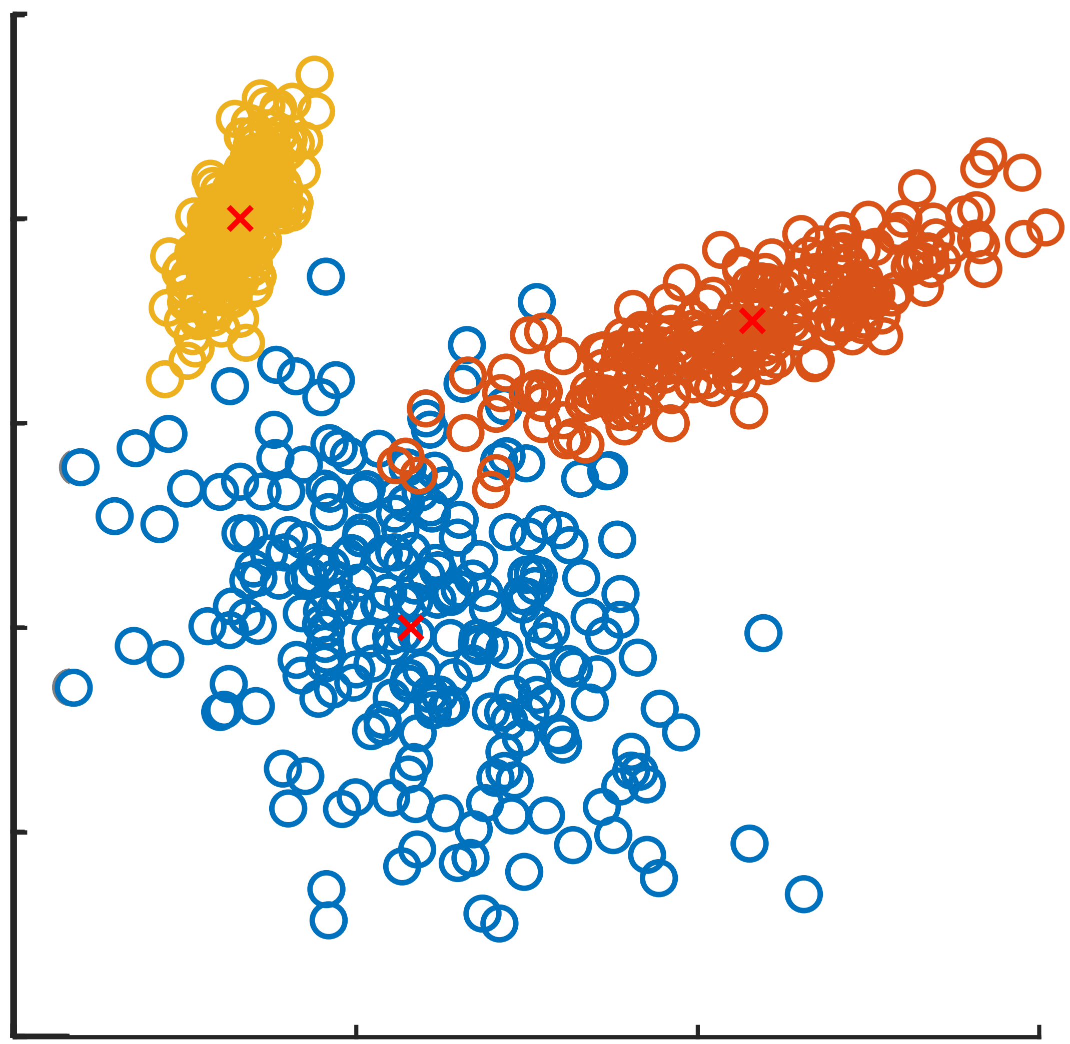


GP

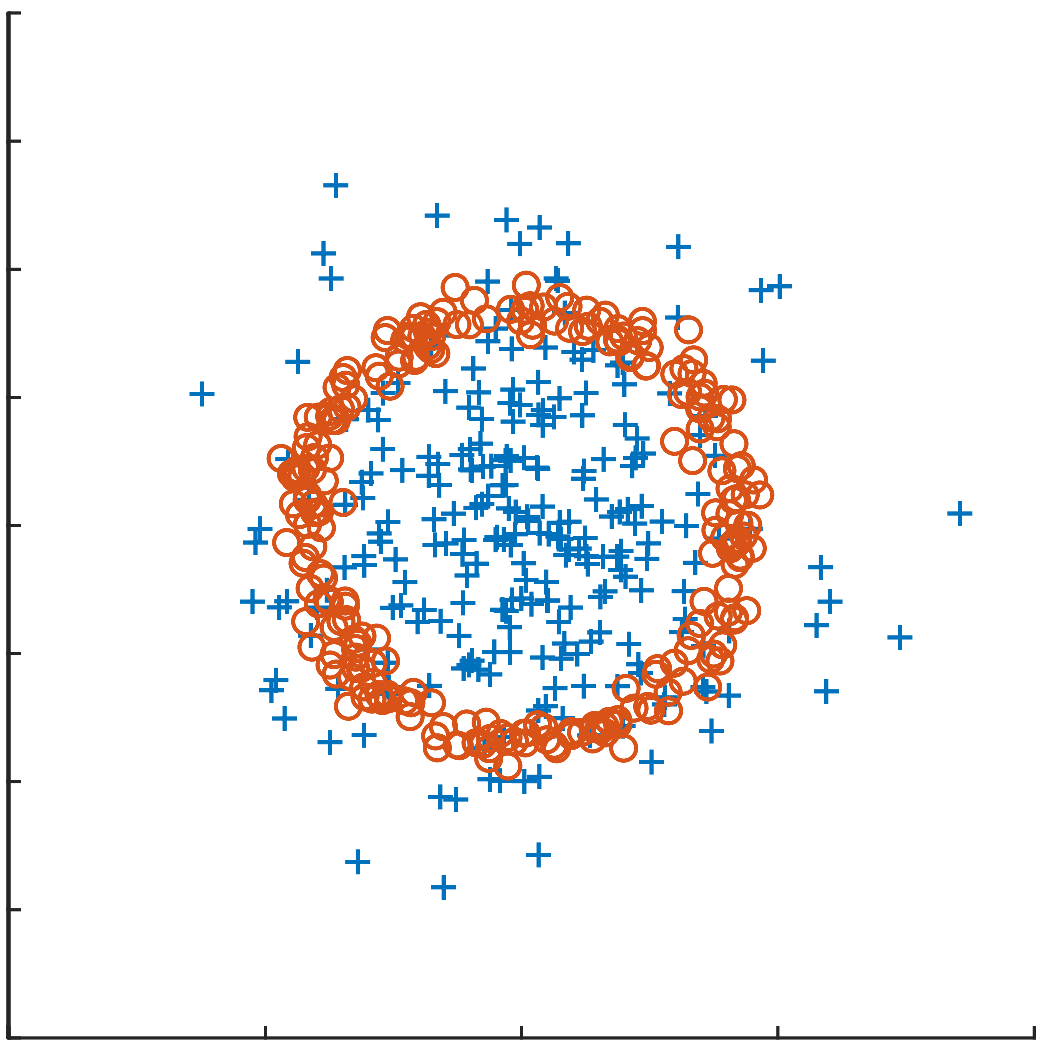
What is probabilistic time series analysis?

All ML: finding structure in data and using it to make predictions

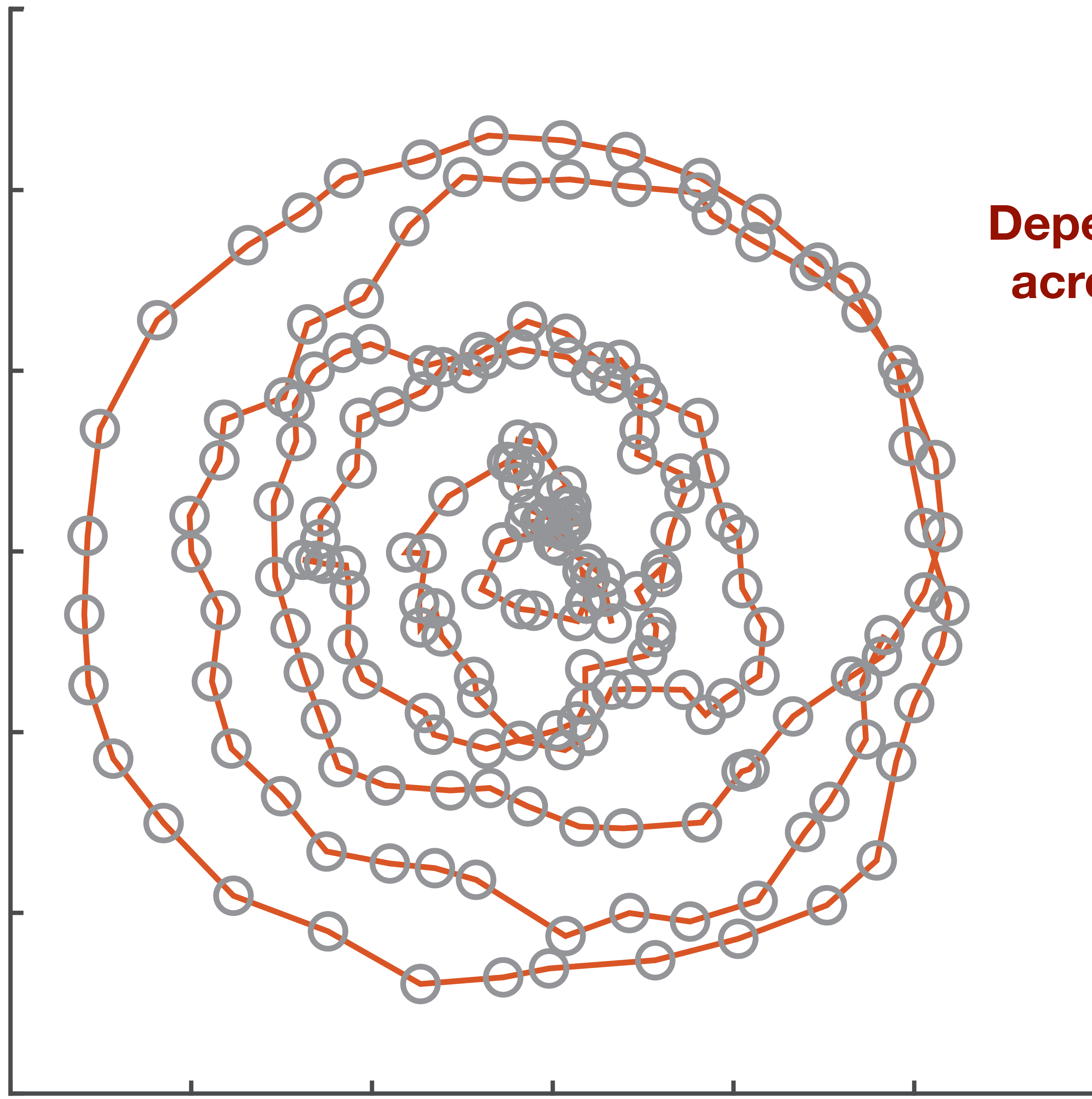
unsupervised



supervised



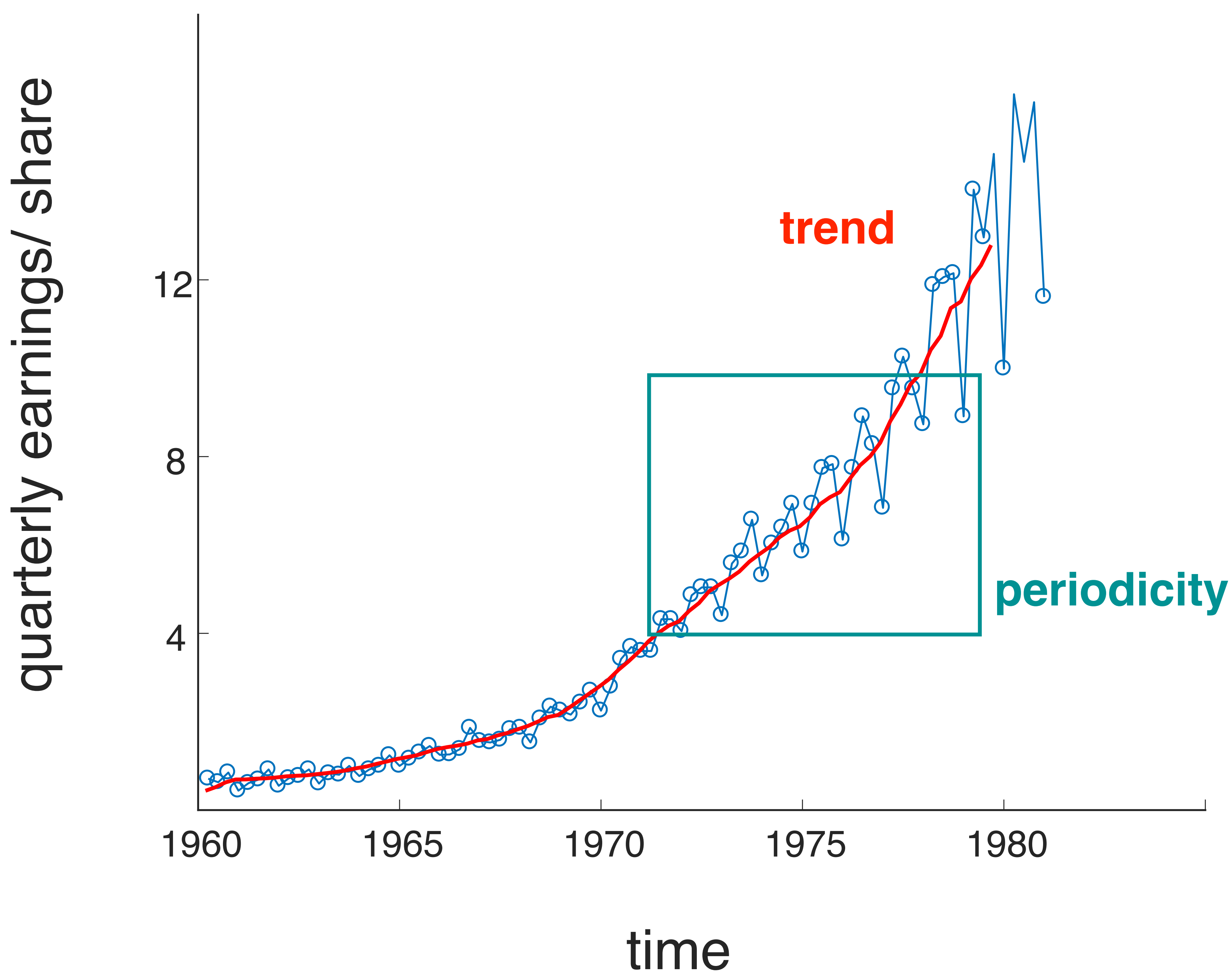
TSA: finding structure and using it to make predictions in <u>sequential data</u>



Dependencies across time!

Some concrete examples

Task: predict future earnings, interpret data structure

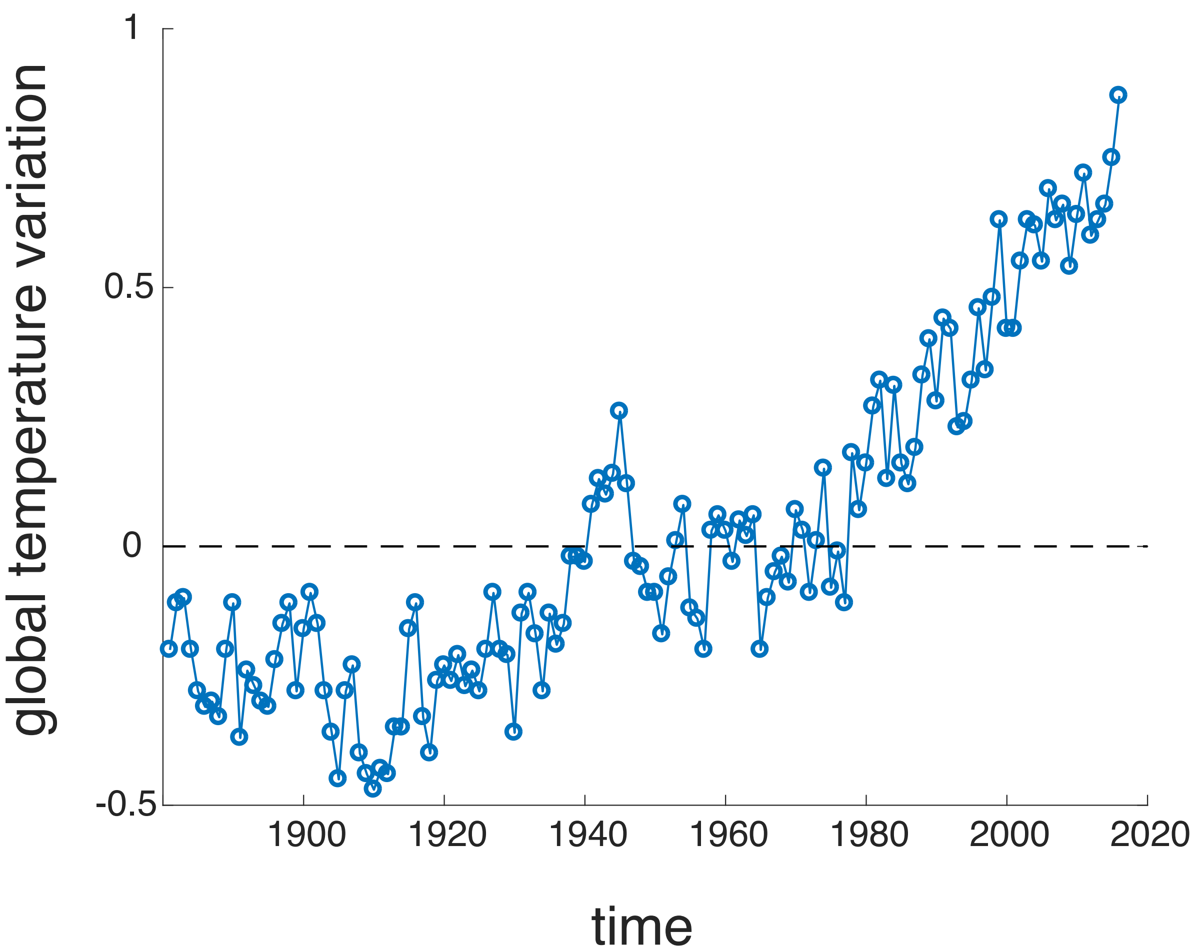


Johnson & Johnson

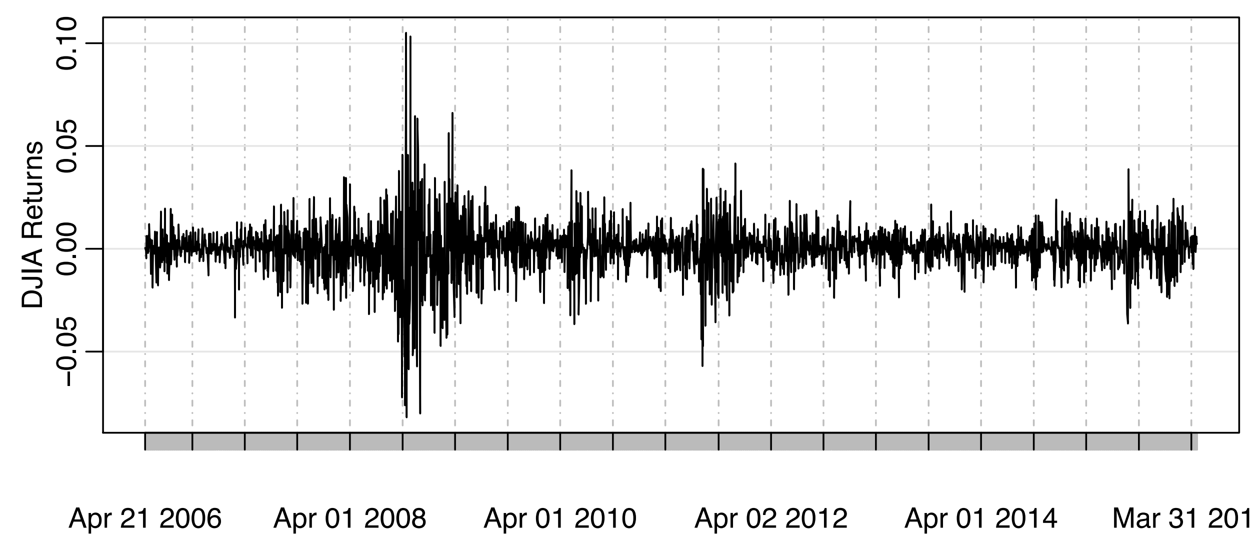
Task: identifying trends, do things change systematically over time?

Task: detect high volatility periods

Global warming

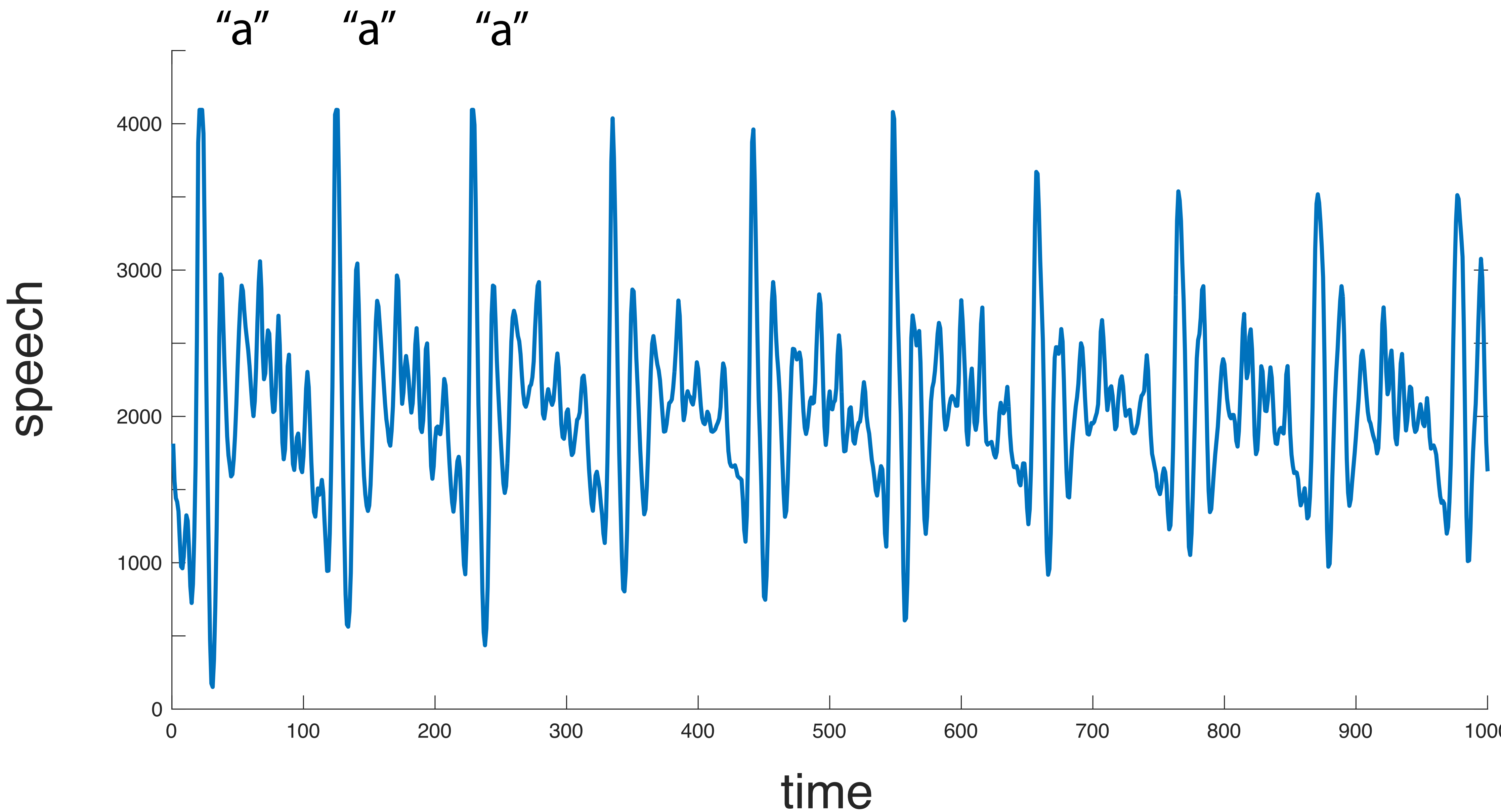


Financial crisis

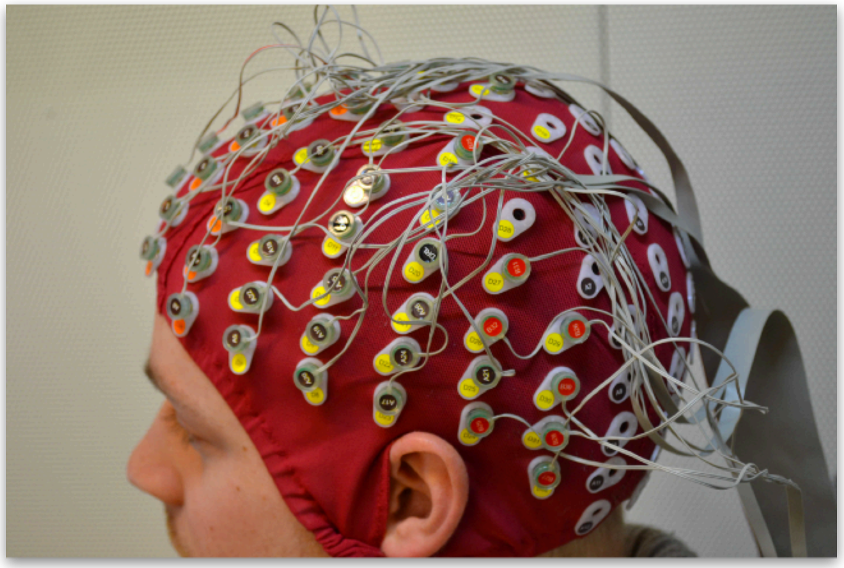


DJIA: *Dow Jones Industrial Average*

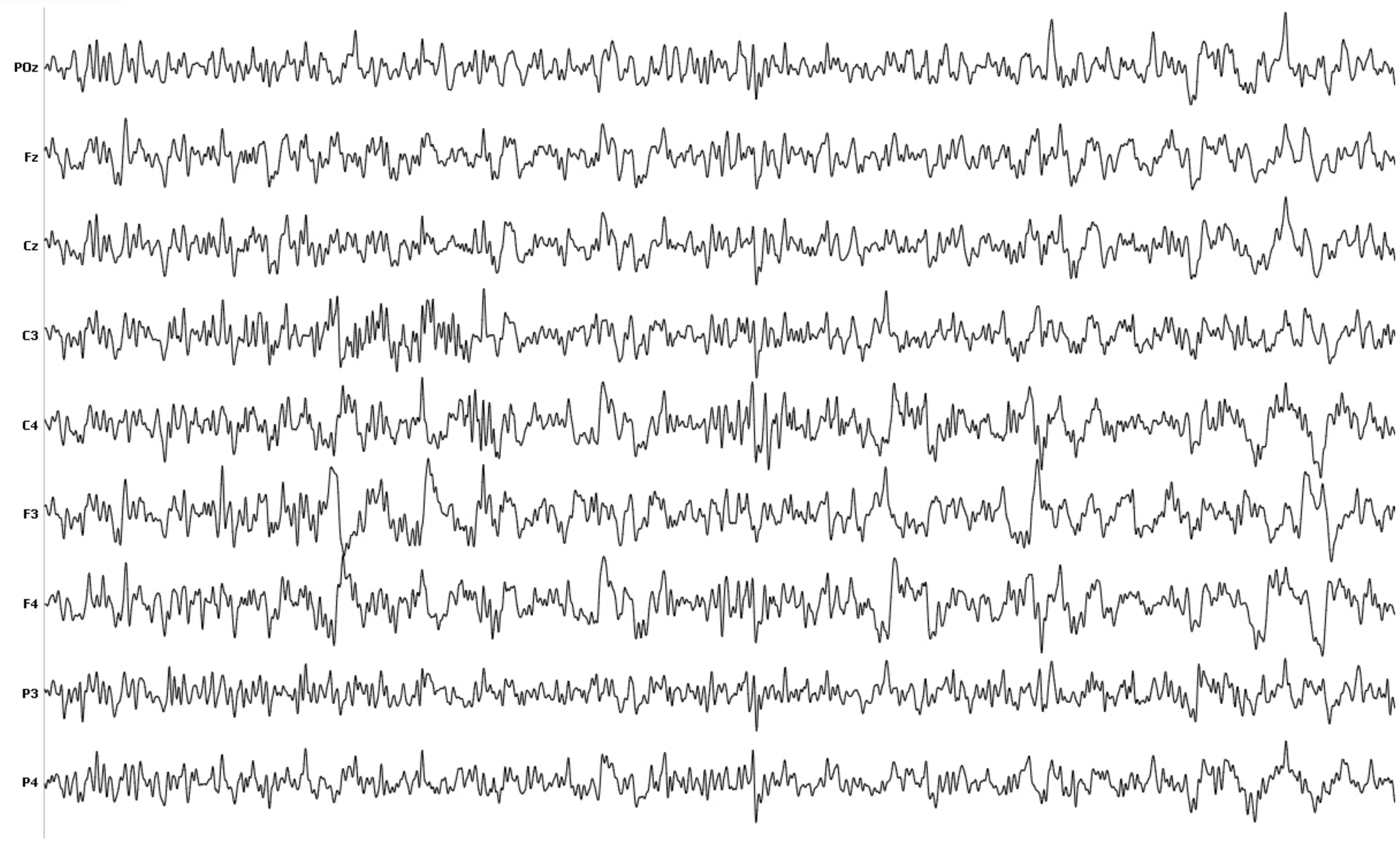
Task: infer discrete latent structure from analog signals



Voice recognition

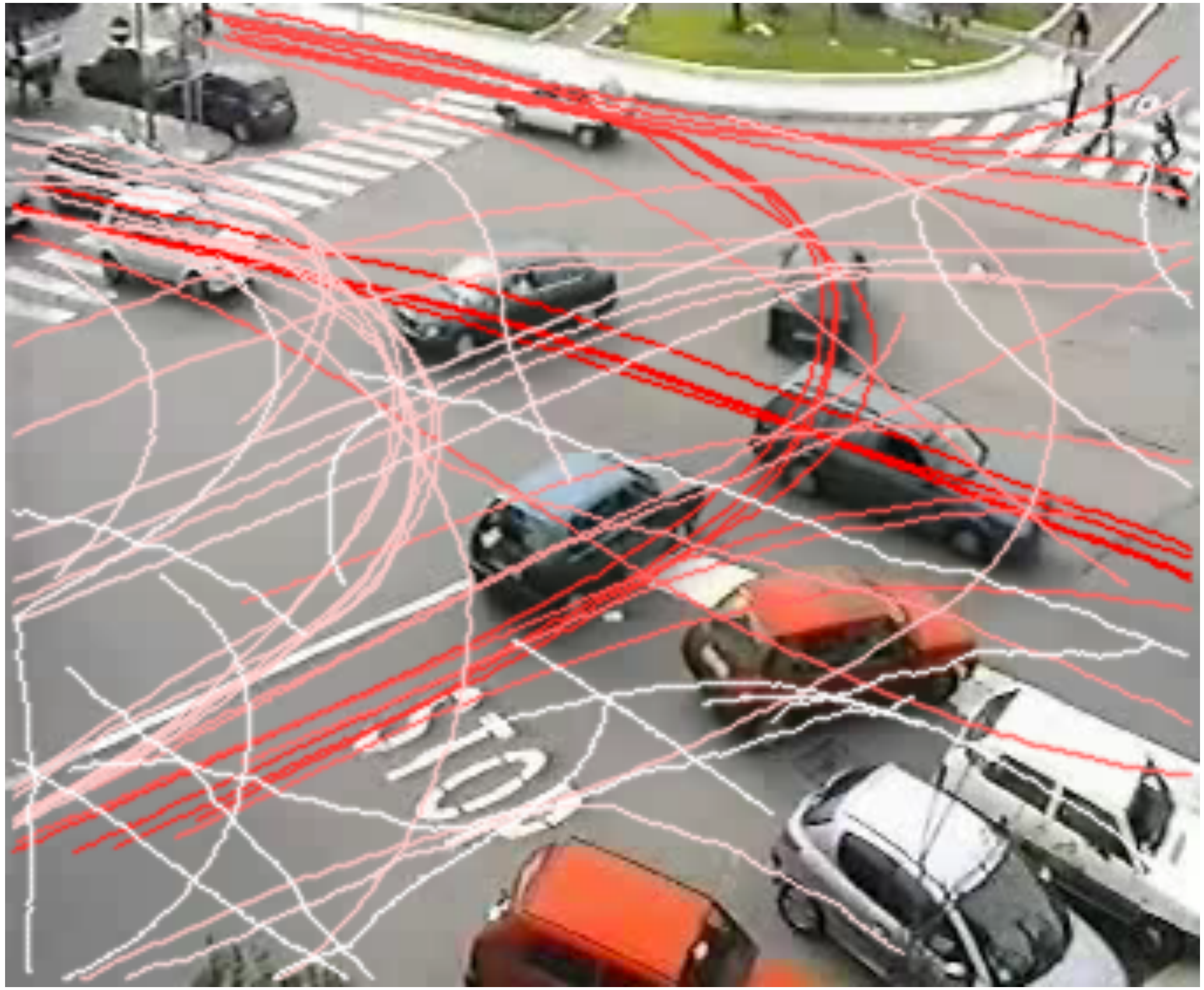


Multivariate time series: EEG



Task: decoding
Brain-computer interfaces

Self driving cars

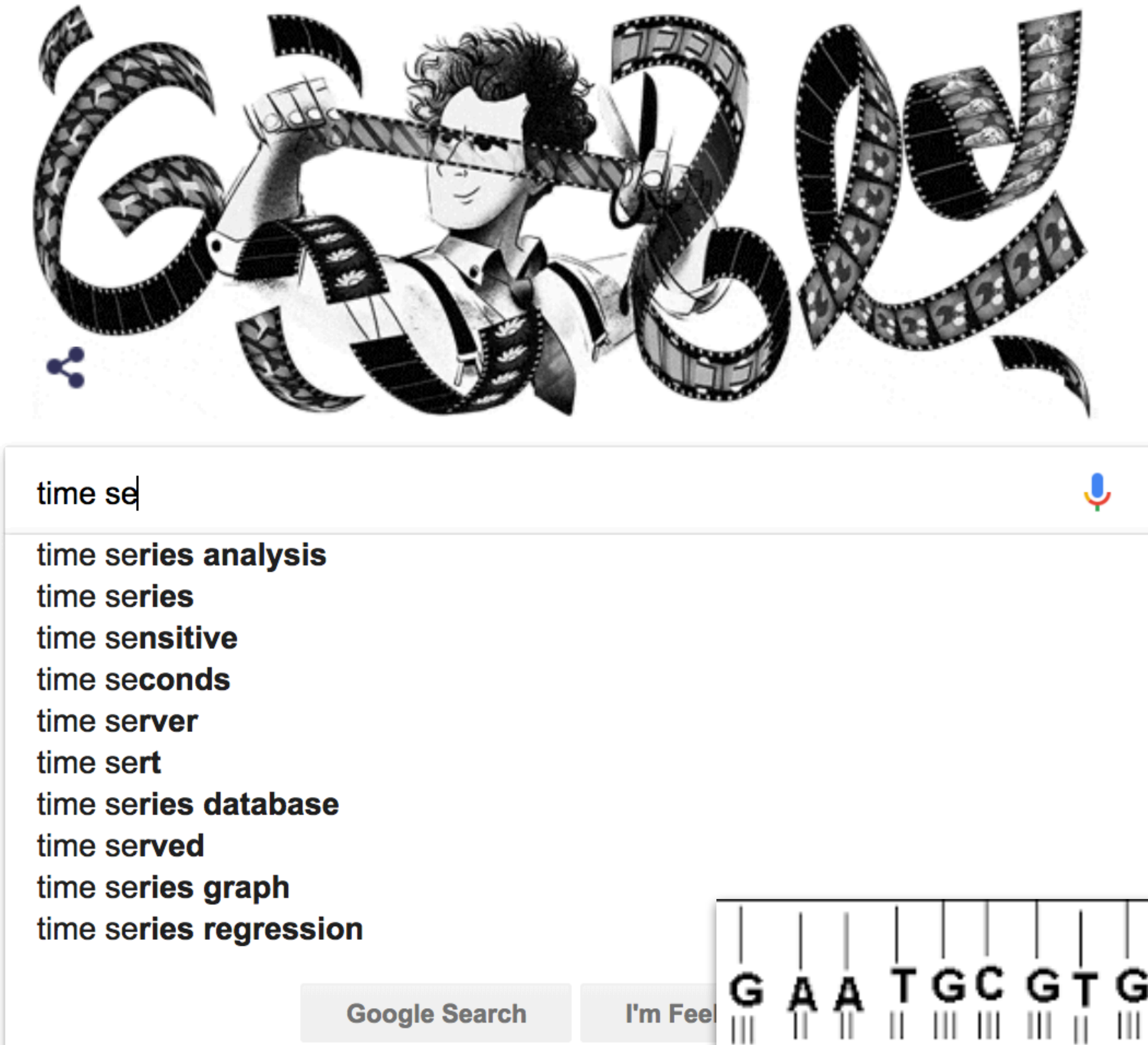


Tasks: simulation, control

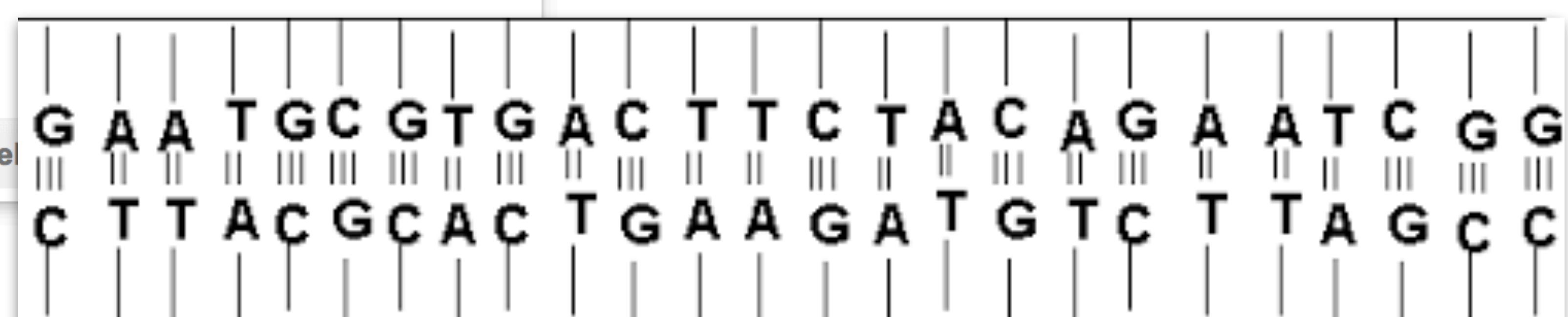
We need uncertainty representation for
optimal decision making, risk minimization

From Saunier & Sayed 2008

Intro: what is a time series?

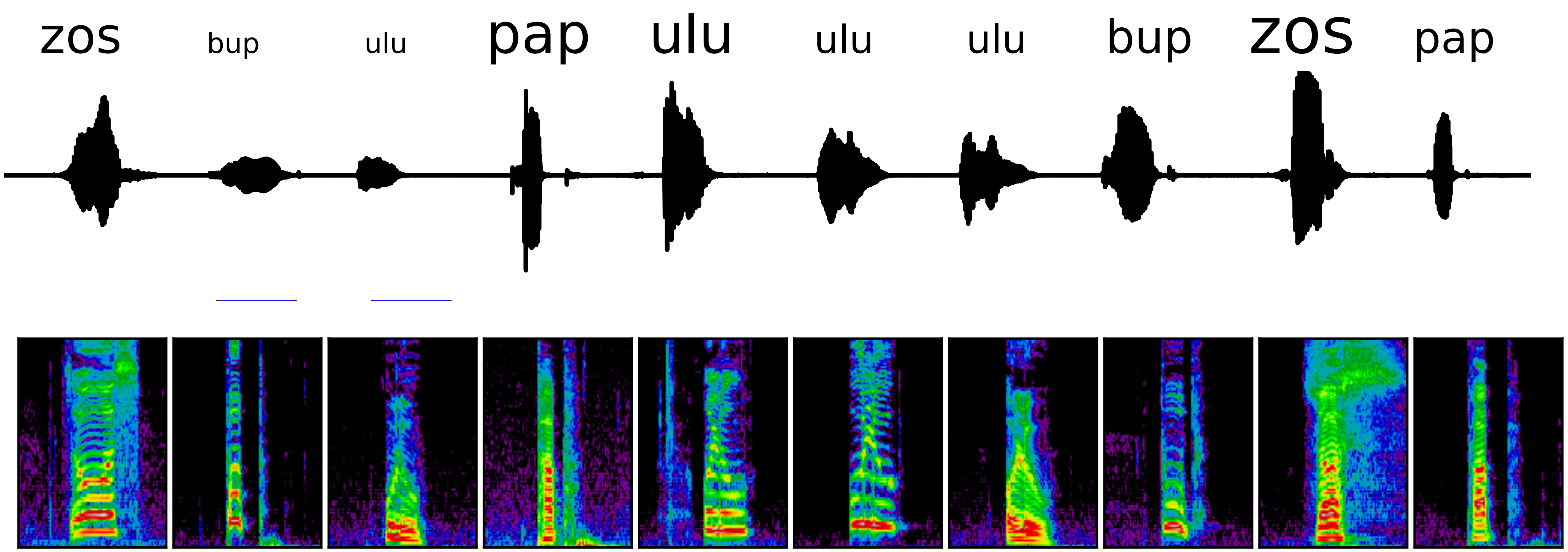


Language!
NLP
Machine translation



More general sequential structure:
e.g. sequence of nucleotide base pairs in DNA

Intro: what is a time series?

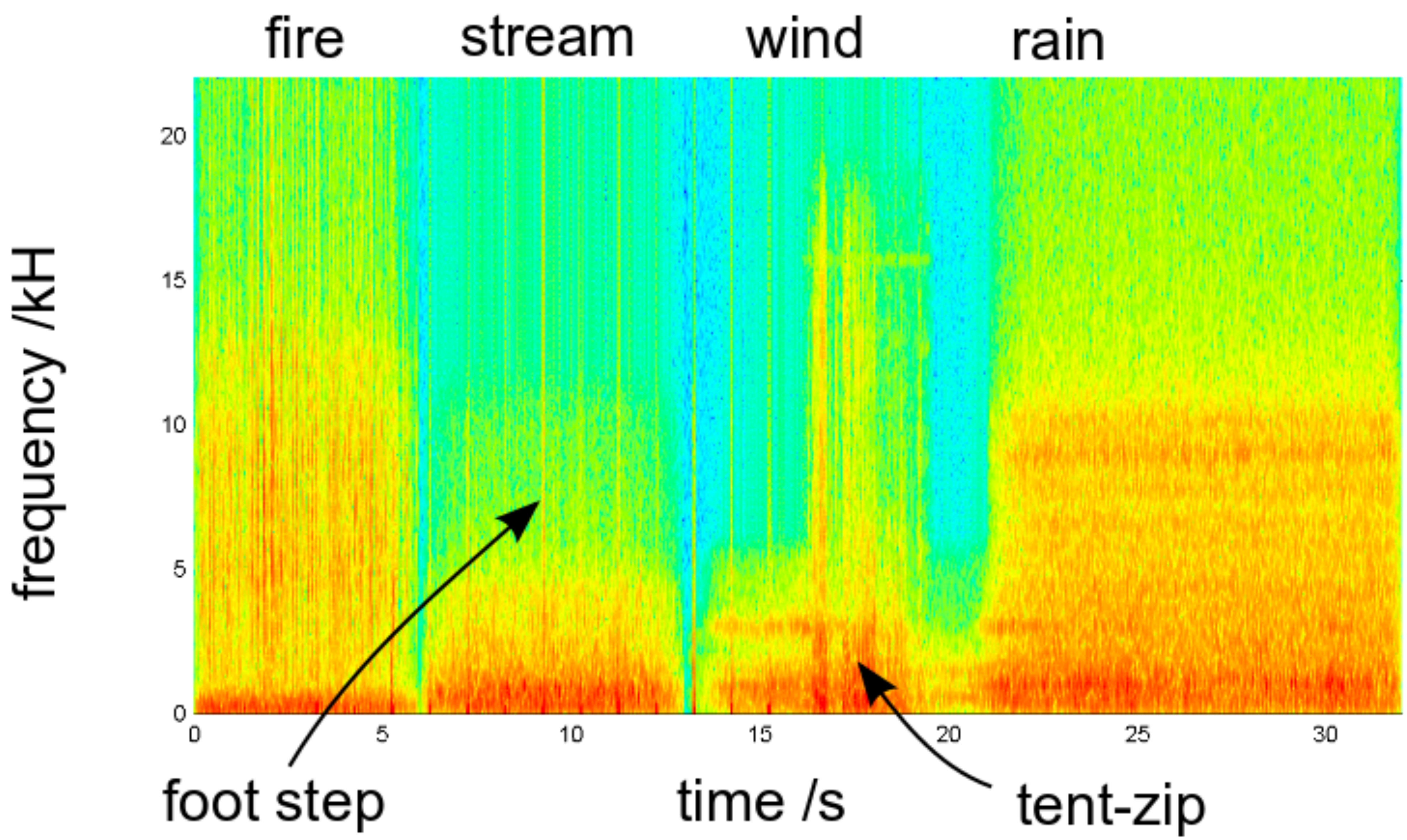


Model data in the frequency domain

e.g. automated speech recognition

Tasks: identify latent structure, denoising…

4. Generate artificial data with same statistics



Definition. Probabilistic models

What is a time series?

Formally, a collection of random variables indexed by time, t*

**Usually discrete time (digital data collection), but continuous time can be convenient in some cases*

$$\{X_1, X_2, \ldots, X_t \ldots\}$$

"stochastic process"

data = "realization"

Unlike the traditional case, NOT I.I.D. !!!

These **dependencies** are the main point; it's what makes prediction possible.

Fully specified by joint*:

$$\mathrm{P}(X_1 \leq x_1, \ldots, X_t \leq x_t \ldots)$$

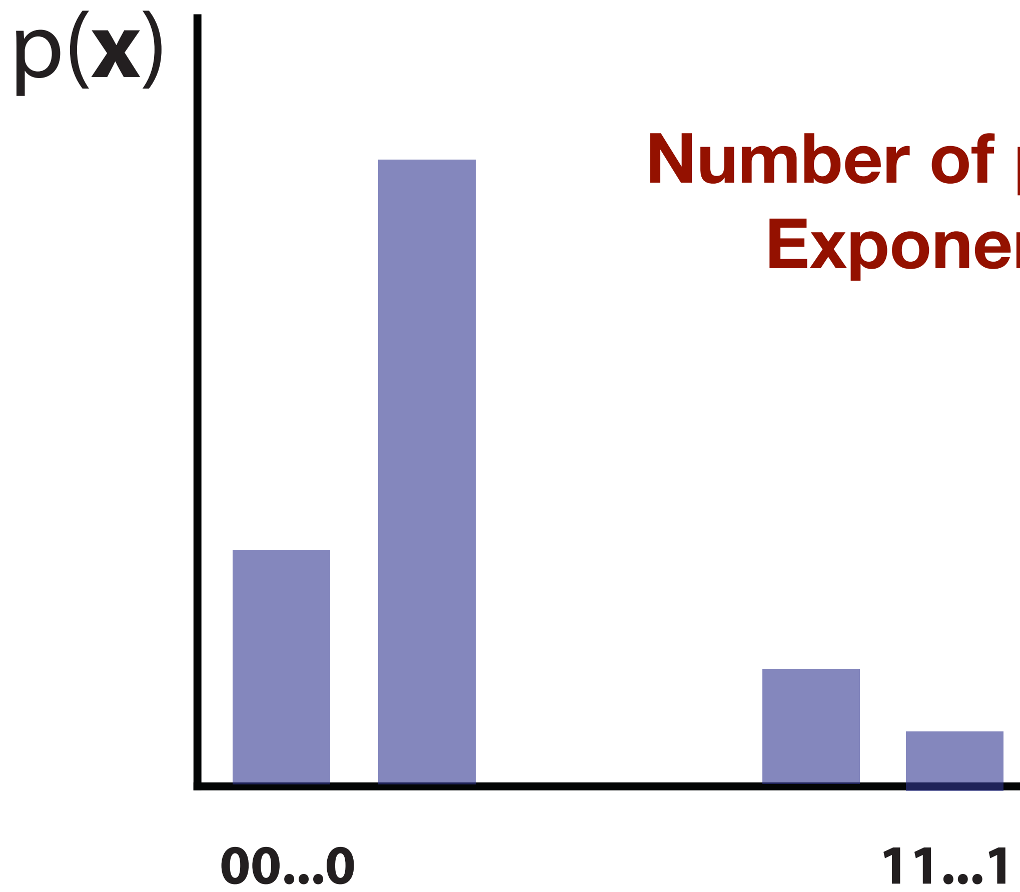
These **dependencies** are also the main problem: they make the math hard.

***Intractable in general, we limit ourselves to more structured classes of distributions*

Challenges in estimating complex dependencies

Simplest case:

$\{X_1, X_2, \ldots, X_t \ldots\}$ **= { 0, 1, 1, 0, 0 }** Sequence of length T



Number of parameters: Exponential in T

What helps us: regularities/structure

Chain rule:

$$\mathrm{P}(X_{1:T}) = \mathrm{P}(\mathrm{X}_1)\mathrm{P}(\mathrm{X}_2|\mathrm{X}_1)...\mathrm{P}(\mathrm{X}_\mathrm{T}|\mathrm{X}_{1:\mathrm{T}-1})$$

Markov assumption (order K=1):

$$\mathrm{P}(X_{1:T}) = \mathrm{P}(\mathrm{X}_1)\mathrm{P}(\mathrm{X}_2|\mathrm{X}_1)\mathrm{P}(\mathrm{X}_3|\mathrm{X}_2)...\mathrm{P}(\mathrm{X}_\mathrm{T}|\mathrm{X}_{\mathrm{T}-1})$$

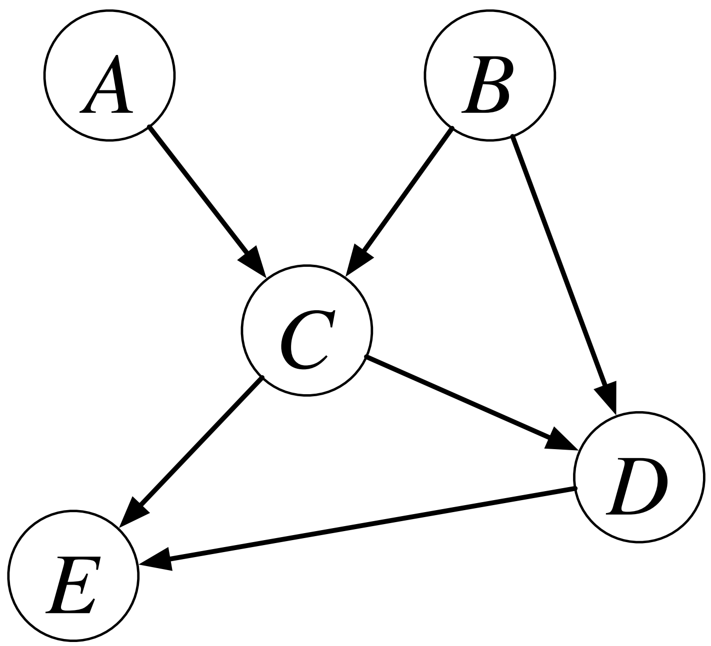
The Bayesian framework: a 3-step recipe

1. Representation:
how to formalize our knowledge about the problem in the language of probabilities

2. Inference:
Given a model, how to make predictions

3. Learning:
Given data, how to learn set the model parameters to match it

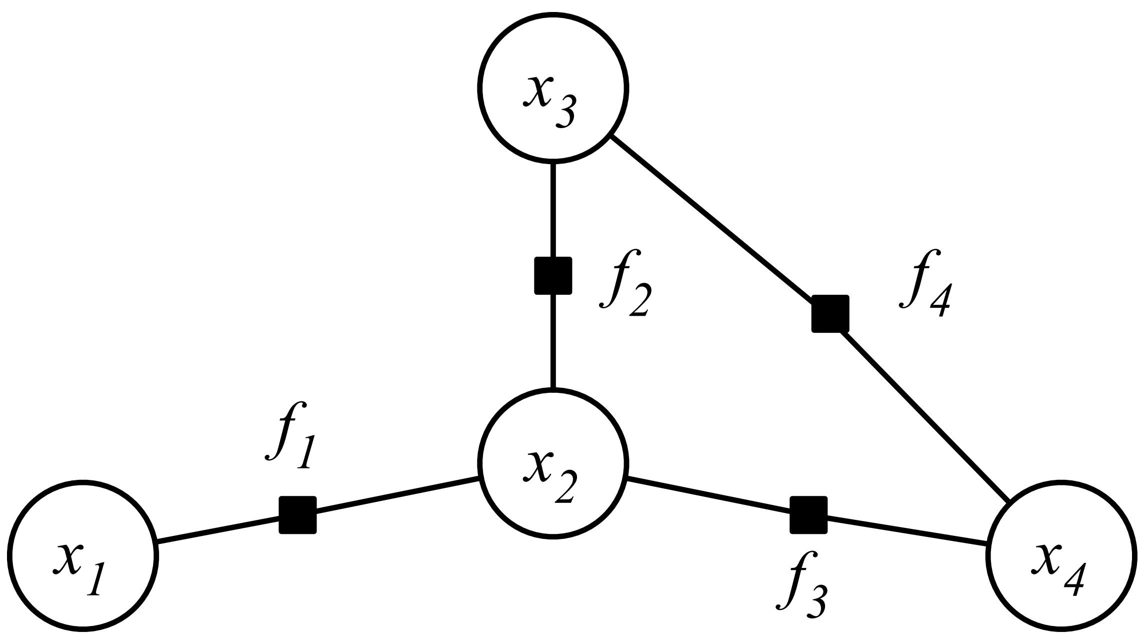
Representing knowledge through graphical models



Directed Acyclic Graph (DAG) models:

$$p(X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i | X_{\mathsf{pa}(i)})$$

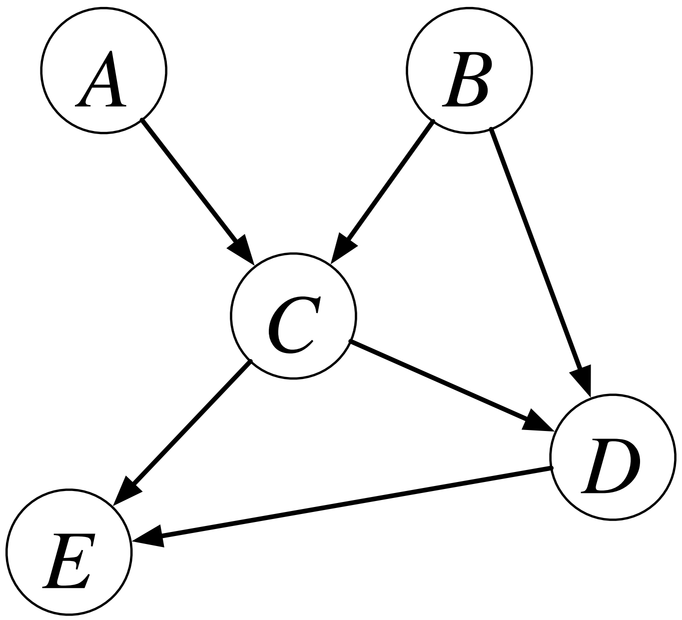
Representing knowledge through graphical models



Factor graphs models:

$$p(\mathbf{x}) = p(x_1, \dots, x_n) = \frac{1}{Z} \prod_j f_j(\mathbf{x}_{S_j})$$

D-separability $X \perp\!\!\!\perp Y | \mathcal{V}$

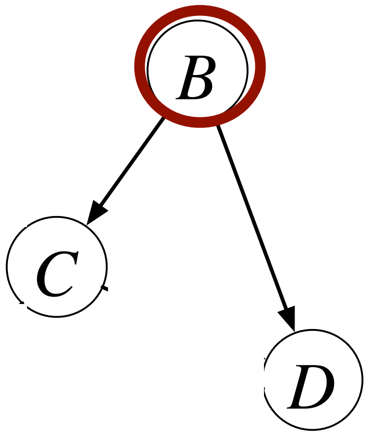


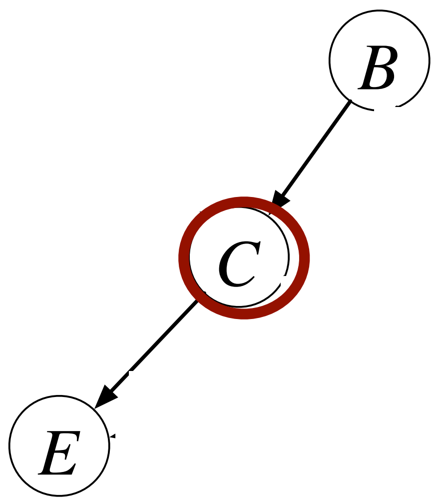
Definition: $\mathcal{V}$ d-separates X from Y if *every* undirected path[2] between X and Y is **blocked** by $\mathcal{V}$. A path is blocked by $\mathcal{V}$ if there is a node W on the path such that either:

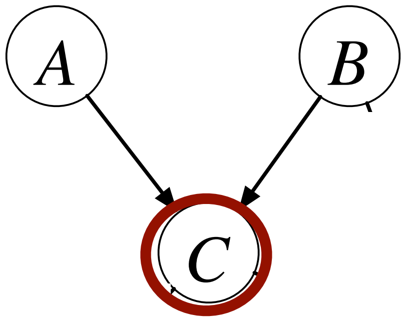
1. W has converging arrows along the path ($\rightarrow W \leftarrow$)[3] and neither W nor its descendants are observed (in $\mathcal{V}$), or
2. W does not have converging arrows along the path ($\rightarrow W \rightarrow$ or $\leftarrow W \rightarrow$) and W is observed ($W \in \mathcal{V}$).

Corollary: Markov Boundary for X: $\{\text{parents}(X) \cup \text{children}(X) \cup \text{parents-of-children}(X)\}$.

Conditioning: 3 cases to remember





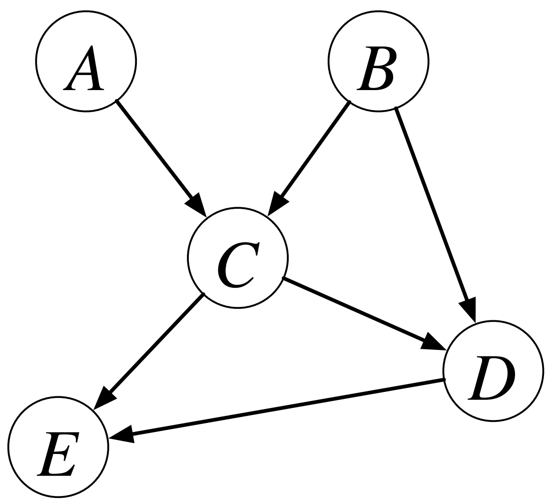


Inference

Sum rule: $P(x) = \sum_y P(x, y)$

Product rule: $P(x, y) = P(x)P(y|x)$

Efficient inference



$$p(A,B,C,D,E) = p(A)p(B)p(C|A,B)p(D|B,C)p(E|C,D)$$

Binary variables: {0,1}

$$p(A|C=c)$$

Parameter learning

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta)P(\theta)}{P(\mathcal{D})}$$

$P(\mathcal{D}|\theta)$	likelihood of θ
$P(\theta)$	prior probability of θ
$P(\theta|\mathcal{D})$	posterior of θ given $\mathcal{D}$

Why bother with probabilistic models when we have deep learning?

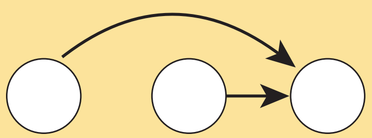
- Often in small data regime, where simple models do best
- Interpretability
- Clear path to incorporating domain-specific knowledge (priors)
- Explicit representation of uncertainty
- Can be combined with deep learning models in interesting ways:
  - Inspire RNN architectures (e.g. Deep Kalman)
  - Factors can be parametrized by ANNs
  - Bayesian deep learning (see Andrew Wilson's work)

Overview of the course

TIME
Lagged relationships

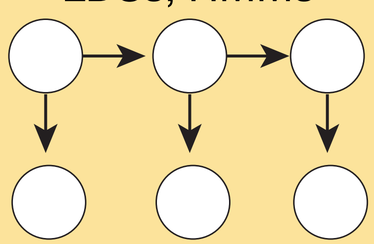
Model temporal dependencies directly

AR+ friends



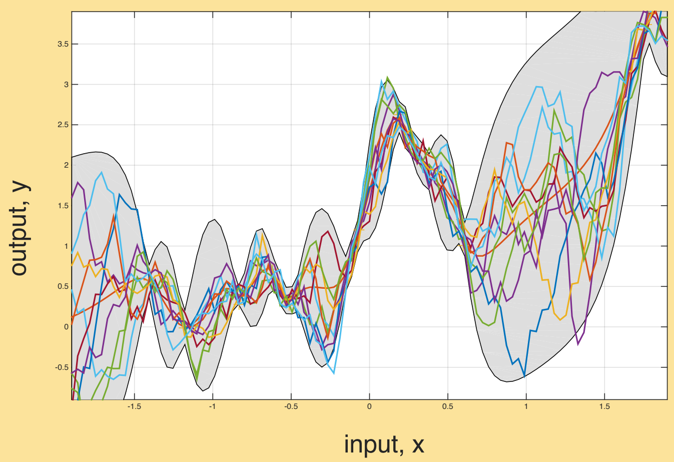
Latent structure

LDSs, HMMs



Distribution function

GP



FREQUENCY
Periodic structure

Stationary processes seasonality/periodicity

Probabilistic spectral analysis

Mixing it up

Non stationary spectral structure (?)

Guest lecture:
example state of the art
HE HE (CDS)