Time Series Analytics

Naren Ramakrishnan

VirginiaTech

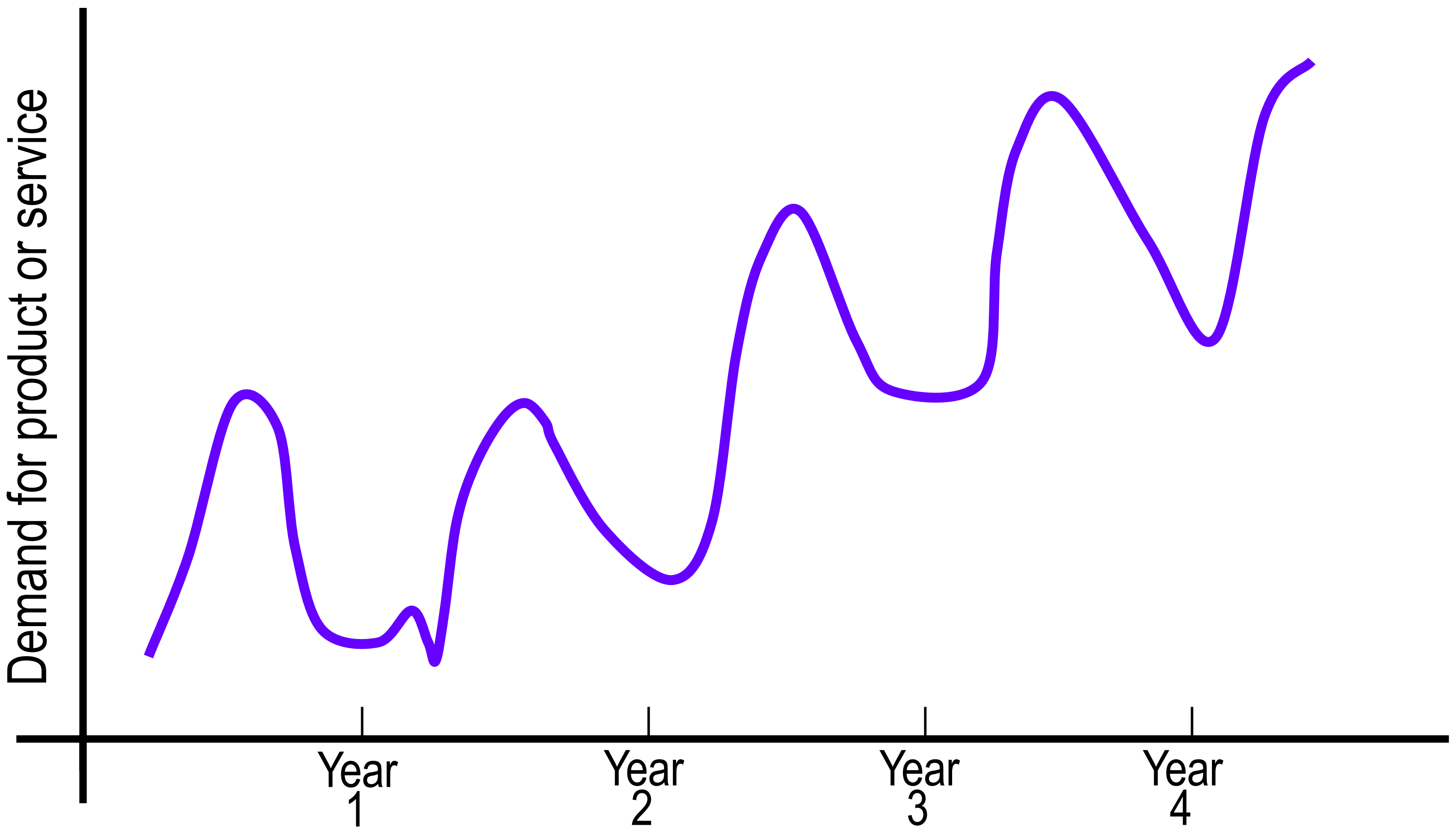
Recap

- Supervised learning
  - Classification, Regression
- Unsupervised learning
  - Clustering, Dimensionality Reduction

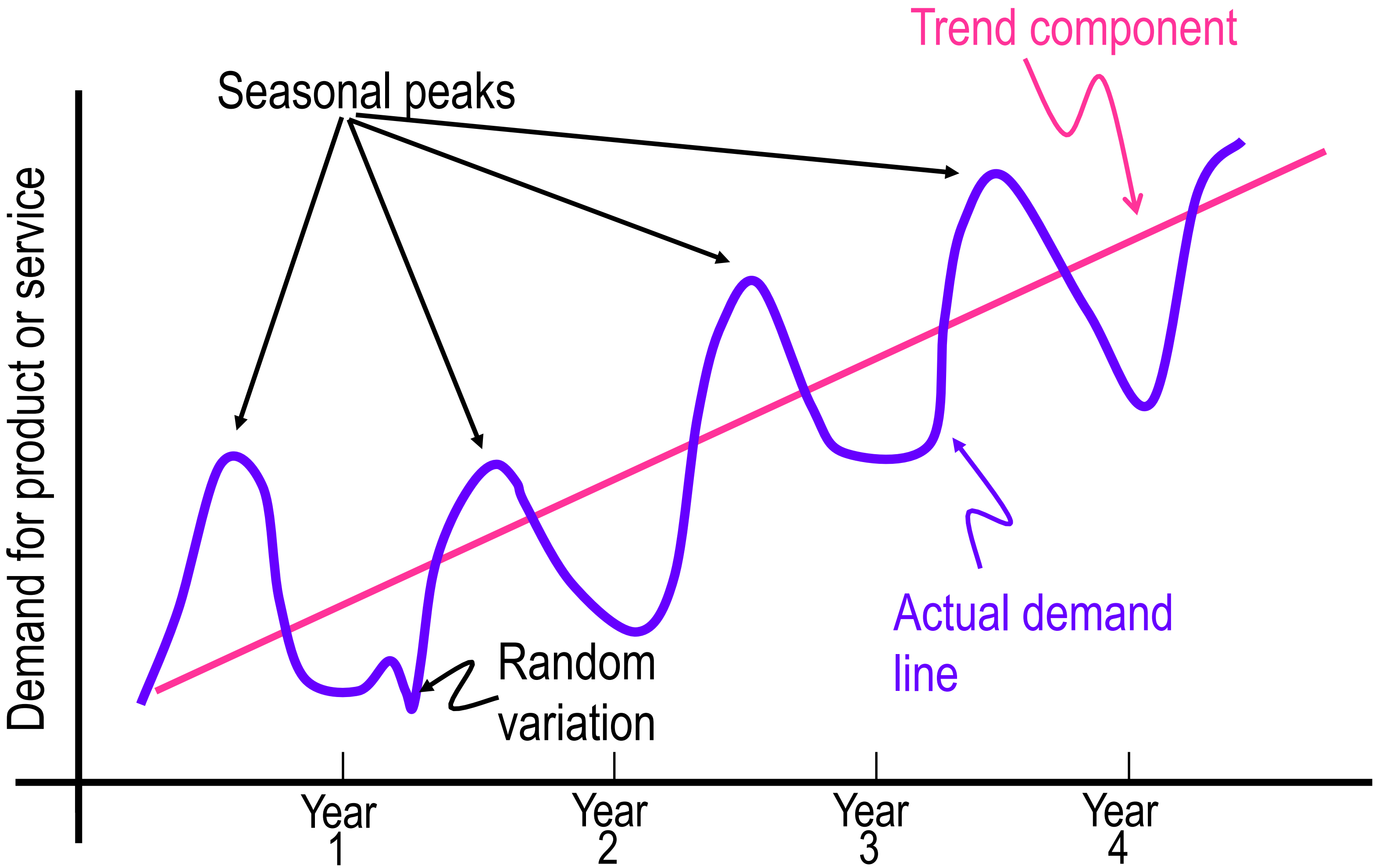
Today

- Time series analytics
  - Has connections to both supervised and unsupervised learning
  - Methods fine tuned for the temporal nature of data
- What is a time series?
  - A continuous-valued variable indexed by time (either discrete or continuous)

Product Demand over Time



Product Demand over Time



Borrowed from Heizer/Render - Principles of Operations Management, 5e, and Operations Management, 7e

What we can do with time series

- Forecast the next time point
  - Autoregression (AR)
  - Moving Average (MA)
  - Autoregressive Moving Average (ARMA)
- Extract key characteristics
  - Cycles (e.g., seasonality), Trends (e.g, growth)
- Cluster time series
  - What is a good similarity/distance measure?

Naive Approach

- Value in *next* period is the same as value in *most recent* period
  - May sales = 48 → June forecast = 48
- Usually leads to atrocious predictions

Simple Moving Average

- Assumes that an average is a good estimator of future behavior
  - Used if *little or no trend*
  - Used for smoothing

$$F_{t+1} = \frac{A_t + A_{t-1} + A_{t-2} + \ldots + A_{t-n+1}}{n}$$

F_{t+1} = Forecast for the upcoming period, t+1

n = Number of periods to be averaged

A_t = Actual occurrence in period t

Simple Moving Average

You are a manager in an electronics store. You aim to forecast iPod sales for months 4-6 using a 3-period moving average.

Month	Sales
1	4
2	6
3	5
4	?
5	?
6	?

Simple Moving Average

Month	Sales	Moving Average (n=3)
1	4	NA
2	6	NA
3	5	NA
4	?	(4+6+5)/3=5
5	?	
6	?	

Reality hits

Month	Sales	Moving Average (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3 ?	5
5	?	
6	?	

Forecast for Month 5

Month	Sales	Moving Average (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3	5
5	?	(6+5+3)/3=4.667
6	?	

Reality hits again

Month	Sales	Moving Average (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3	5
5	?7	4.667
6	?	

Reality again

Month	Sales	Moving Average (n=3)
1	4	NA
2	6	NA
3	5	NA
4	3	5
5	7	4.667
6	?	(5+3+7)/3=5

Weighted Moving Average

- Gives more emphasis to recent data

$$F_{t+1} = w_1 A_t + w_2 A_{t-1} + w_3 A_{t-2} + \ldots + w_n A_{t-n+1}$$

- Weights
  - decrease for older data
  - sum to 1.0

Weighted Moving Average: 3/6, 2/6, 1/6

Month	Sales	Weighted Moving Average
1	4	NA
2	6	NA
3	5	NA
4	?	31/6 = 5.167
5	?	
6	?	

Next iteration

Month	Sales	Weighted Moving Average
1	4	NA
2	6	NA
3	5	NA
4	3	31/6 = 5.167
5	7	25/6 = 4.167
6		32/6 = 5.333

Exponential Smoothing

- **Assumes that the most recent observations have the highest predictive value**
  - **gives more weight to recent time periods**

$$\boxed{F_{t+1} = F_t + \alpha \underbrace{(A_t - F_t)}_{e_t}}$$

F_{t+1} = Forecast value for time $t+1$

A_t = Actual value at time t

α = Smoothing constant

Need initial forecast F_t to start.

Exponential Smoothing – Example 1

i	Ai
Week	Demand
1	820
2	775
3	680
4	655
5	750
6	802
7	798
8	689
9	775
10	

Given the weekly demand data what are the exponential smoothing forecasts for periods 2-10 using α=0.10?

Assume $F_1=D_1$

Exponential Smoothing – Example 1

i	Ai	Fi
Week	Demand	α = *0.1*
1	820	820.00
2	775	
3	680	
4	655	
5	750	
6	802	
7	798	
8	689	
9	775	
10		

$F_2 = F_1 + \alpha(A_1 - F_1)$

$=820+.1(820-820)$

$=820$

Exponential Smoothing – Example 1

i	Ai	Fi
Week	Demand	α = *0.1*
1	820	820.00
2	775	820.00
3	680	
4	655	
5	750	
6	802	
7	798	
8	689	
9	775	
10		

$F_3 = F_2 + \alpha(A_2 - F_2)$

$=820+.1(775-820)$

$=815.5$

Exponential Smoothing – Example 1

i	Ai	Fi
Week	Demand	α = *0.1*
1	820	820.00
2	775	820.00
3	680	815.50
4	655	
5	750	
6	802	
7	798	
8	689	
9	775	
10		

This process continues through week 10

Exponential Smoothing – Example 1

i	Ai	Fi	
Week	Demand	α = *0.1*	α = *0.6*
1	820	820.00	820.00
2	775	820.00	820.00
3	680	815.50	793.00
4	655	801.95	725.20
5	750	787.26	683.08
6	802	783.53	723.23
7	798	785.38	770.49
8	689	786.64	787.00
9	775	776.88	728.20
10		776.69	756.28

What if the α constant equals 0.6?

Exponential Smoothing

- How to choose α
  - depends on the emphasis you desire to place on the most recent data
- Increasing α makes forecast more sensitive to recent data

Effect of Smoothing Constant α

$$F_{t+1} = F_t + \alpha\,(A_t - F_t)$$

or

$$F_{t+1} = \underbrace{\alpha}_{w_1} A_t + \underbrace{\alpha(1-\alpha)}_{w_2} A_{t-1} + \underbrace{\alpha(1-\alpha)^2}_{w_3} A_{t-2} + \ldots$$

$\alpha=$	Weights		
	Prior Period	2 periods ago	3 periods ago
	α	$\alpha(1-\alpha)$	$\alpha(1-\alpha)^2$
$\alpha = 0.10$	10%	9%	8.1%

Autoregression

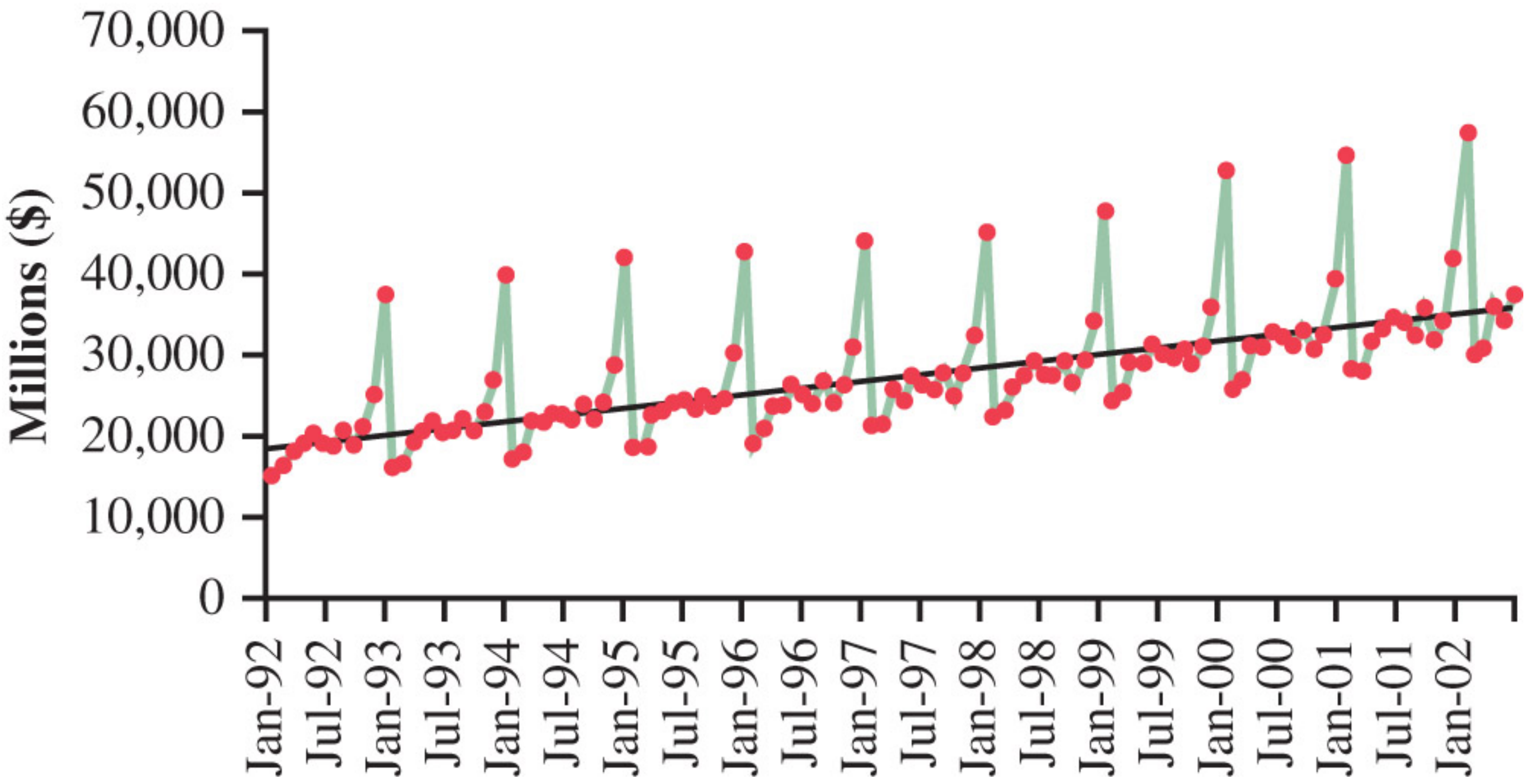
- Just like regression
  - Independent variables are older response variables (times *t-1*, *t-2*, …)
  - Dependent variable is the current response variable (at time *t*)
  - Setup a simple supervised learning scenario

Year	Y_i	Y_{i-1}	Y_{i-2}
92	4	---	---
93	3	4	---
94	2	3	4
95	3	2	3
96	2	3	2
97	2	2	3
98	4	2	2
99	6	4	2

ARMA

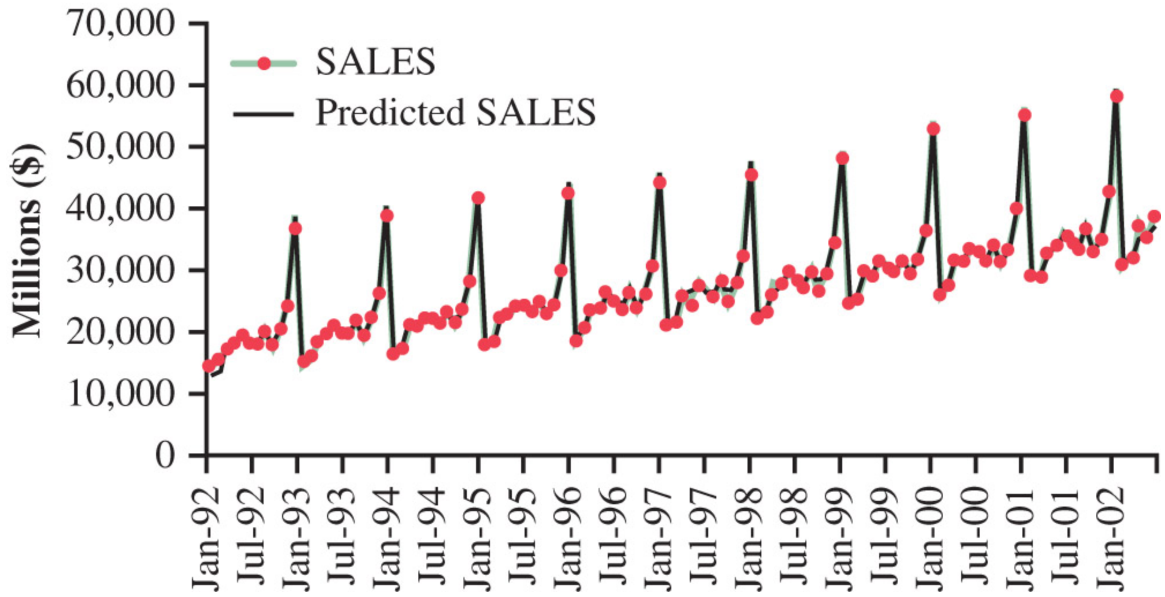
- Has elements of both
  - Autoregression (AR)
  - Moving average (MA)
- These methods are ideal for
  - Trend identification
    - (But restricted to linear methods so far)
      - Can be extended to non-linear relationships as well

Example of ARMA



Modeling seasonality

- Create new independent variables
  - E.g., denoting season, or calendar month



Applying a forecasting method

- Collect historical data
- Select a model
  - Moving average methods
    - Select *n* (number of periods)
    - For weighted moving average: select weights
  - Exponential smoothing
    - Select α
  - Autoregression
    - Define past lookup period, independent variables, seasonality

…but how do you evaluate a forecast?

Evaluating forecasting methods

a. MAD = Mean Absolute Deviation

$$\text{MAD} = \sum_{t=1}^{n} \frac{|A_t \ - Ft|}{n}$$

b. MSE = Mean Squared Error

$$\text{MSE} \ = \sum_{t=1}^{n} \frac{(A_t \ - Ft)^2}{n}$$

c. RMSE = Root Mean Squared Error

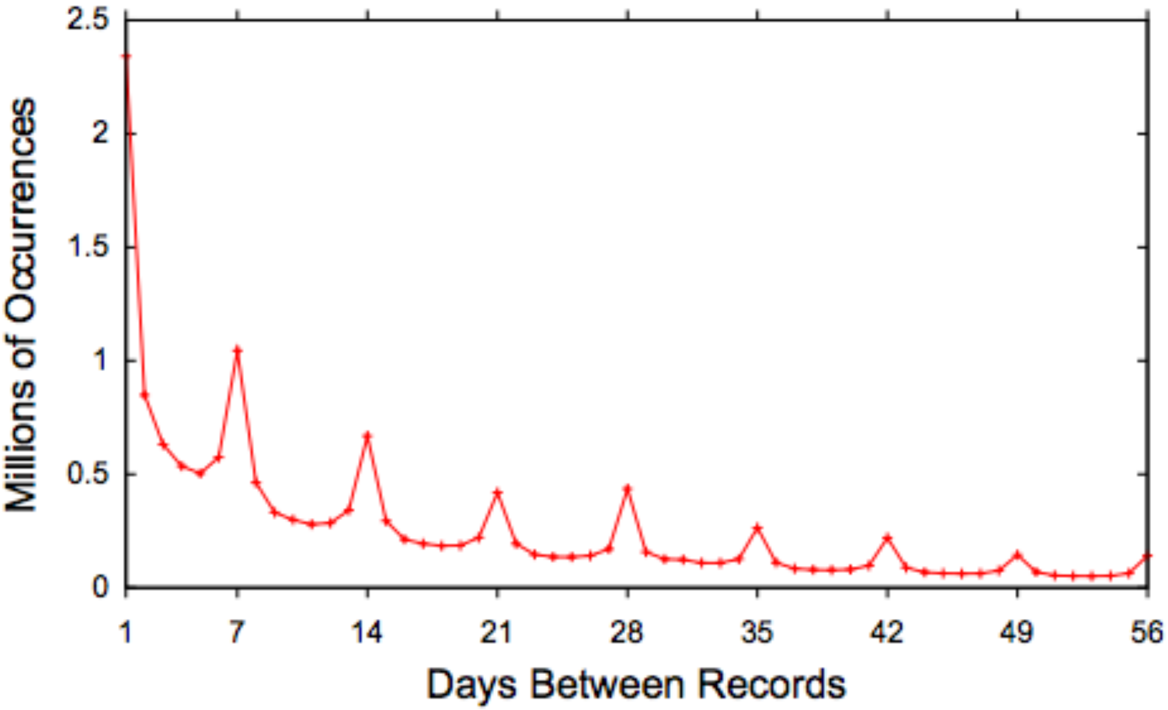
$$\text{RMSE} = \sqrt{MSE}$$

When do we use which method?

- AR or ES or MA or ARMA?
  - Depends on the stationarity properties of the time series
- Typically
  - Step changes, outliers are removed at the outset along with some smoothing
    - Many such transformations abound
  - The residuals are then subject to time series modeling
    - Under the assumption that they are stationary
  - Defining the time series itself is a creative activity!

Example: differencing

- Example from electronic medical records



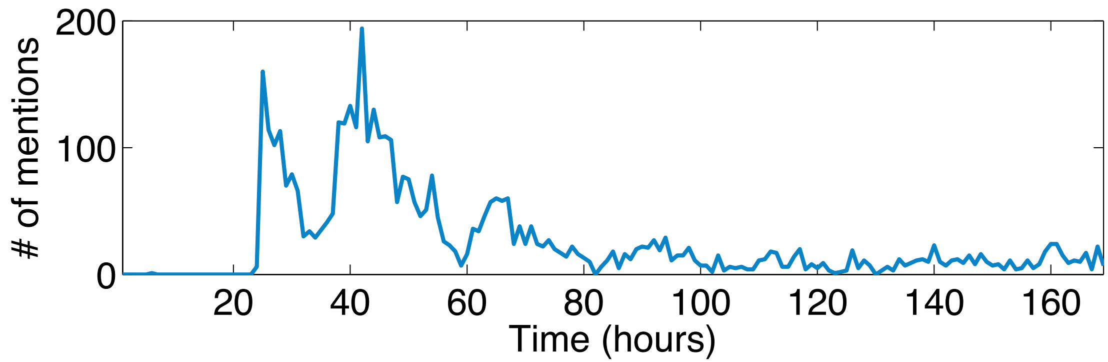
Clustering time series data

- Recap clustering algorithms
  - K-means
  - Hierarchical clustering
    - => All of them require a distance or similarity function
- Use MAD and MSE functions given earlier
  - And apply these algorithms as if you are applying them to points

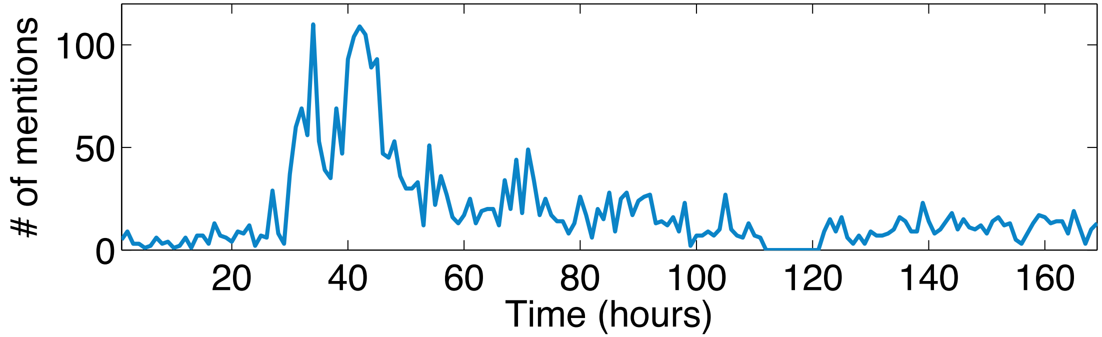
Example of time series clustering

- Rise and fall patterns of memes on social media

"you can put lipstick on a pig"

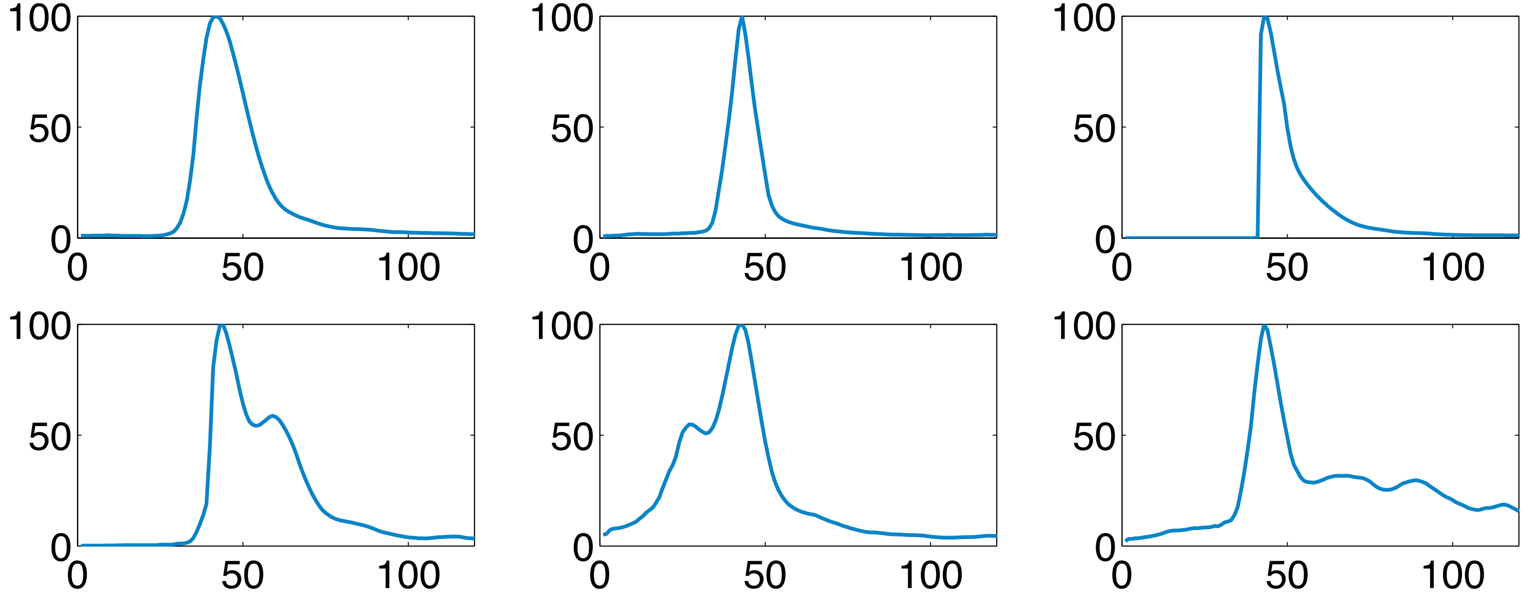


"yes we can"



Key clusters identified

- Modeling popularity across different social media channels
  - **four** classes on YouTube [Crane et al. ’08]
  - **six** classes on Meme [Yang et al. ’11]



Recap

- Given time series data
  – How do we extend it? (Forecasting)
  – How do we identify salient aspects of it? (Trend, seasonality identification)
  – How do we group them (Clustering)