# INFDTA01-2
# Data Mining – Week 5

Giulia Costantini

# Agenda

| Week | Topic |
|---|---|
| ~~1~~ | ~~Introduction; Intro clustering~~ |
| ~~2~~ | ~~Clustering; Intro genetic algorithms~~ |
| ~~3~~ | ~~Genetic algorithms~~ |
| ~~4~~ | ~~Practicum; (optional) GA + regression case study~~ |
| 5 | Forecasting (SES, DES) |
| 6 | Forecasting (TES) |
| 7 | Linear programming; outliers; summary/practicum |
| 8 | Check assignments + oral check |

- Forecasting
  - definition, applications
  - exponential smoothing (ES)
    - moving average (simple and weighted)
    - simple exponential smoothing
    - double exponential smoothing (Holt's trend-corrected ES)
    - triple exponential smoothing (multiplicative Holt-Winters ES) *[next lesson]*
  - measuring forecast accuracy *[next lesson]*
  - summary
- Practical assignment part 3

# What is forecasting?

- process of making statements about events whose actual outcomes (typically) have not yet been observed
- estimation of some variable of interest at some specified future date

*Past data is used to predict a future outcome*

- central issues are *risk* and *uncertainty*
  - the future often looks nothing like the past
  - "only guarantee with forecasting: your forecast is <u>wrong</u>!"
  - indicate the degree of uncertainty of a forecast (*prediction intervals*)

# Why forecasting?

- Typical forecasting problem → taking some data point over time and projecting that data into the future
  - Sales
  - Demand
  - Supply
  - Carbon emissions
  - Population
  - …

# How to do forecasting?

- A lot of methods
  - *Naïve approach*
  - *Time series*
    - **Moving average, weighted moving average, ES**, ARMA, ARIMA, extrapolation, trend estimation, growth curve, etc…
  - *Causal/econometric* (regression analysis)
  - *Judgmental*
    - Delphi, statistical surveys, scenario building, etc…
  - *Artificial intelligence*
    - Neural networks, machine learning, pattern recognition, etc…
  - Other
    - Simulation, probability forecasting, etc…

# Time series

- Data with natural temporal ordering → time series
  - sequence of data points
  - measured at successive points in time
  - spaced at uniform time intervals

- Time interval can be anything
  - Year
  - Month
  - Week
  - Day
  - Hour
  - …

# Naïve approach

Forecast for *any* future period

- historical average

Features

- simple & cost-effective
- used as benchmark for comparison with more sophisticated models

Example

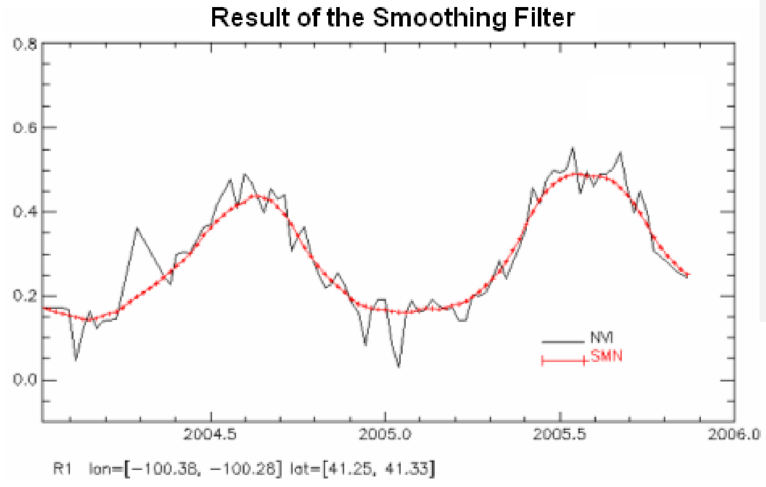
- Data points $= \{ 55, 63, 48, 50, 56 \}$
- Average $= 54.4$ → forecast for all future values

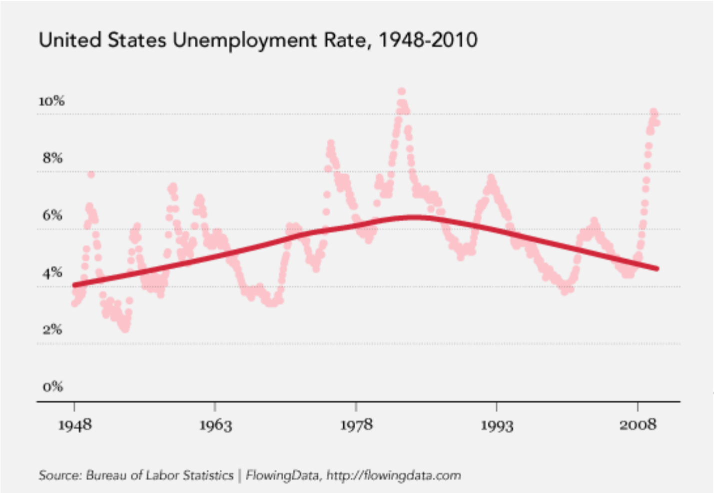
# Time series

- Smoothing
  - Help us see patterns (i.e., trends)
  - Smooth out the irregular roughness to see a clearer signal

- $\{x_t\}$ = data sequence
- $\{s_t\}$ = output sequence from smoothing

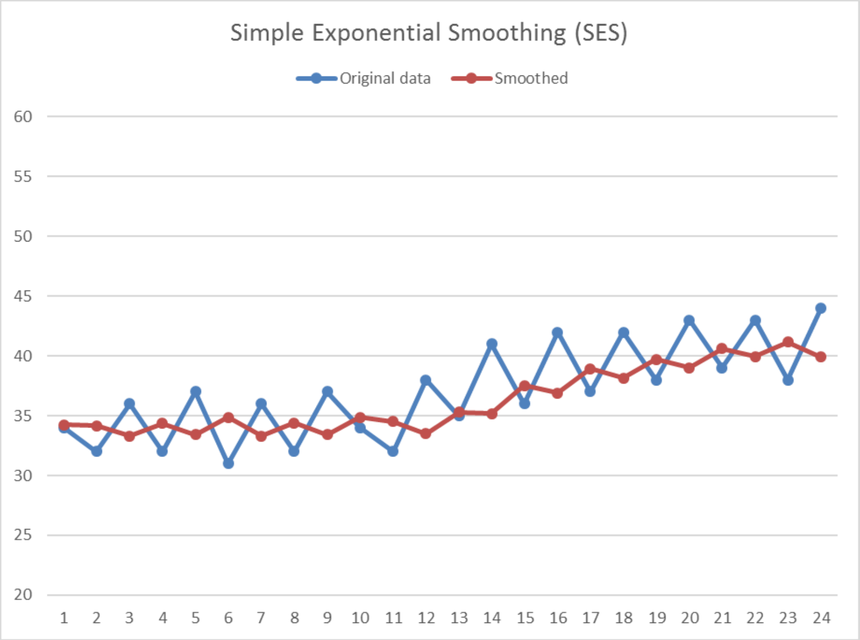
Examples



# Time series

- Smoothing
  - We **know** the original time series $x_t$ (blue dots)
  - We want to **find a** smoothed version of it $s_t$ (red dots)
  - There are various techniques possible (moving avg, SES, DES, etc…)




| Smoothing coefficient (alpha) | | |
|---|---|---|
| 0,4 | | |

| Time t | Demand x_t | Smoothed sequence s_t |
|---|---|---|
| 1 | 34 | 34,3 |
| 2 | 32 | 34,2 |
| 3 | 36 | 33,3 |
| 4 | 32 | 34,4 |
| 5 | 37 | 33,4 |
| 6 | 31 | 34,9 |
| 7 | 36 | 33,3 |
| 8 | 32 | 34,4 |
| 9 | 37 | 33,4 |
| 10 | 34 | 34,9 |
| 11 | 32 | 34,5 |
| 12 | 38 | 33,5 |
| 13 | 35 | 35,3 |
| 14 | 41 | 35,2 |
| 15 | 36 | 37,5 |
| 16 | 42 | 36,9 |
| 17 | 37 | 38,9 |
| 18 | 42 | 38,2 |
| 19 | 38 | 39,7 |
| 20 | 43 | 39,0 |
| 21 | 39 | 40,6 |
| 22 | 43 | 40,0 |
| 23 | 38 | 41,2 |
| 24 | 44 | 39,9 |

# Moving average

- **SMA** (simple moving average)
  - simplest way to smooth a time series
  - (unweighted) mean of the previous $k$ data points

$$s_t = \frac{1}{k}\sum_{i=0}^{k-1} x_{t-i} = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-k+1}}{k}$$

# Moving average

- Example
  - $\{x_t\} = \{10, 8, 9, 11, 10, 20, 9, 12, 11\}$
  - Suppose $k = 4$

- Smoothed sequence
  - $s_4 = \frac{10+8+9+11}{4} = 9.5$
  - $s_5 = \frac{8+9+11+10}{4} = 9.5$
  - …
  - $s_9 = \frac{20+9+12+11}{4} = 13$

# Moving average

- Choice of $k$: arbitrary
  - Small value → less smoothing; more responsive to recent changes
  - Large value → greater smoothing effect
  - **Demo excel**

- Problems
  - Cannot be used on the first $k-1$ terms
  - Consider only the last $k$ observations (previous ones are ignored completely)
  - Every observation in the last $k$ counts the same (has the same weigth)

# Moving average

- **WMA** (weighted moving average)
  - Similar to SMA but each value of the average has a *different weight* $w_i$

$$s_t = \sum_{i=0}^{k-1} w_i x_{t-i} = w_1 x_t + w_2 x_{t-1} + w_3 x_{t-2} + \cdots + w_{k-1} x_{t-k+1}$$

- Weighting factors
  - $\{w_1, w_2, \ldots w_{k-1}\}$ such that $\sum_{i-0}^{k-1} w_i = 1$
  - give more weight to **recent** terms
    - … but the other two problems remain!

# Exponential smoothing (SES)

- **SES** (*simple* exponential smoothing)
  - Commonly applied to financial market and economic data
  - Current smoothed statistics = weighted average of
    - the previous observation $x_{t-1}$ and
    - the previous smoothed statistic $s_{t-1}$

$$\boldsymbol{s_t = \alpha x_{t-1} + (1-\alpha) s_{t-1}}$$

# Exponential smoothing (SES)

- Not many observations needed to start
  - one at least, depending on how we initialize $s$
- Every observations counts, proportionally to time
  - Less recent count (exponentially) less

- Initialization
  - $s_1 = x_1$
  - $s_1 =$ average of the first 12 values of $x_t$ (better!)
  - …

# Exponential smoothing (SES)

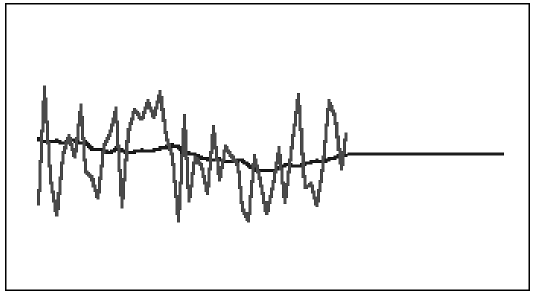
$$\boldsymbol{s_t = \alpha x_{t-1} + (1-\alpha)s_{t-1}}$$

- $\alpha$ = **data smoothing factor**
  - $0 < \alpha < 1$
  - Large values → less smooth, greater weight to recent changes
  - Small values → more smooth, less responsive to change
  - How to choose it?
    - Statistician judgment
    - Optimizing it, by minimizing the standard error of the smoothed statistics with respect to the time series (*method of least squares*): minimize $\sqrt{\frac{\sum(s_i-x_i)^2}{n-1}}$ ($n$ = number of observations in the time series)

# Exponential smoothing (SES)

- So, what is the forecast???
  - $s_{n+1}$ = last value available of the smoothed statistic
  - $f_{n+m}$ = forecast for $m$ time periods into the future

$$\boldsymbol{f_{n+m} = s_{n+1}\ \forall m \in \{1, 2, 3, \dots\}}$$



Examples:

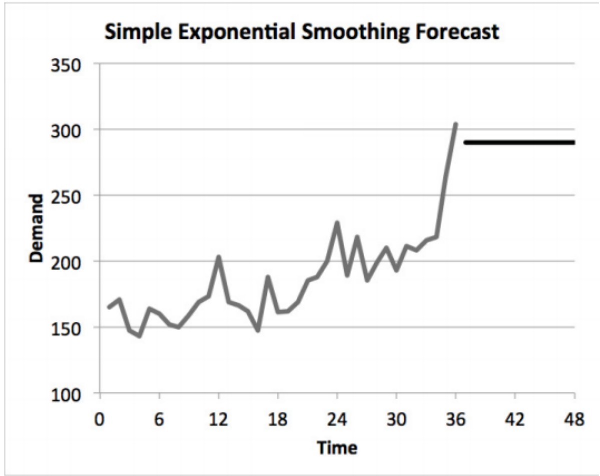
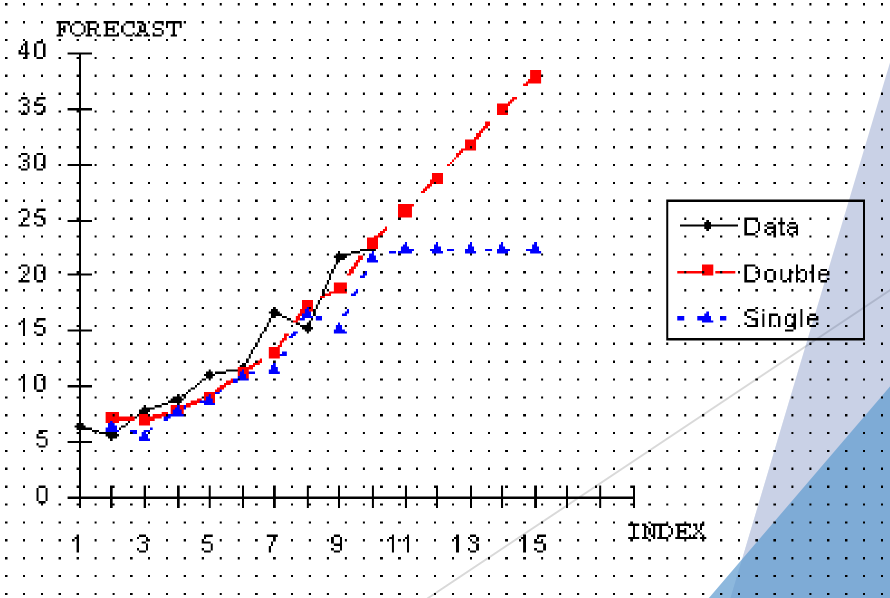



**Figure 8-9:** Graphing the final simple exponential smoothing forecast



# Exponential smoothing (SES)

$$s_t = \alpha x_{t-1} + (1-\alpha)s_{t-1} \qquad f_{n+m} = s_{n+1}\ \forall m \in \{1,2,3,\dots\}$$

- **Exercise**
  - Download from N@tschool the Excel file "Week 5 - SES and DES exercise.xlsx"
  - Go the first sheet and insert the formula of SES to compute the *smoothed sequence* associated to the time series
  - Fill in also the "*squared error*" column and use it to compute the error measure $\sqrt{\frac{\sum(s_i-x_i)^2}{n-1}}$
  - Compute the forecast for the following year (from 37 to 48)
  - Try different values for the smoothing factor $\alpha$ and see how that changes the shape of the smoothing sequence, the forecasts and the value of the error measure

# Exponential smoothing (DES)

- Trend in the data? SES is ***not*** good at all!
  - *Holt's trend-corrected exponential smoothing* is better
  - Also called *double* exponential smoothing (two smoothing factors)

- $\boldsymbol{s_t}$ is (as before) the smoothed statistic (*level*) at time $t$
- $\boldsymbol{b_t}$ is our best estimate of the *trend* at time $t$
- Forecast for the next time step $t + 1 \Rightarrow \boldsymbol{f_{t+1}} = s_t + b_t$

# Exponential smoothing (DES)

- Update equations

$$\boldsymbol{s_t = \alpha x_t + (1-\alpha)(s_{t-1} + b_{t-1})}$$

$$\boldsymbol{b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}}$$

$$\boldsymbol{f_{t+1} = s_t + b_t}$$

- Smoothing factors
  - $0 < \alpha < 1$ is the *data* smoothing factor
  - $0 < \boldsymbol{\beta} < 1$ is the *trend* smoothing factor
  - Can be optimized minimizing the error, similarly as before…
    - Standard error $= \sqrt{\frac{\sum(f_i - x_i)^2}{n-\mathbf{2}}}$

# Exponential smoothing (DES)

- Initialization
  - $s_2 = x_2$ , $b_2 = x_2 - x_1$
    - $f_3 = s_2 + b_2$: the first forecast available is at $t = 3$!
  - other possibility: fit a regression model on the first half of the data and use *slope* & *intercept* as starting values for $s$ and $b$

# Exponential smoothing (DES)

- And the forecast??? Given
  - the final value of the smoothed statistics, $s_n$
  - the final estimate of the trend, $b_n$
  - the number of time intervals $m$ into the future you want to forecast

$$f_{n+m} = s_n + m b_n$$

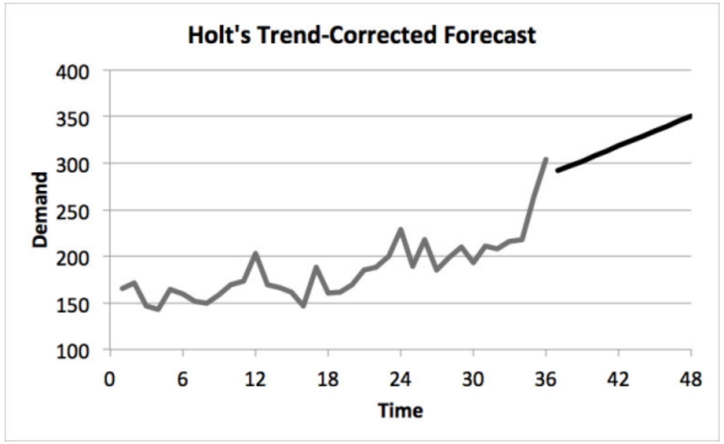
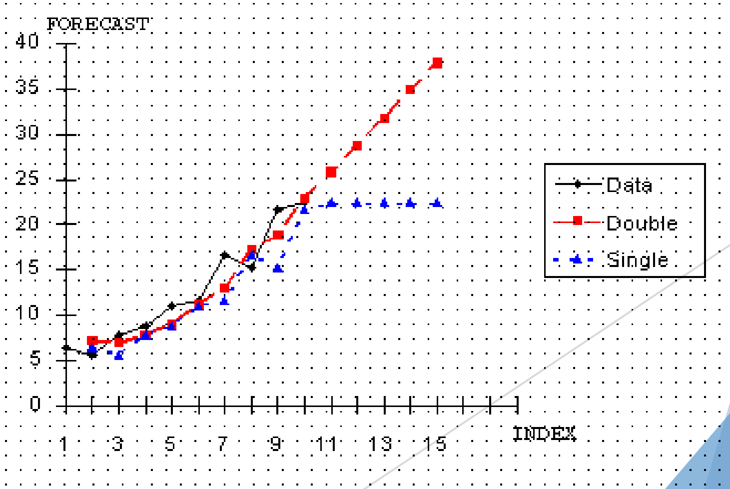



**Figure 8-19:** Graph of optimal Holt's forecast



# Exponential smoothing (DES)

- Example
  - time interval = months
  - $n = 36$ (3 years of data)
  - $s_n = 280,\ b_n = 26$

$$f_{n+m} = s_n + mb_n$$

- What's the forecast for month 39?

$$f_{39} = s_n + (39 - 36) \times b_n = 280 + 3 \times 26 = 358$$

- What's the forecast for the first month of the fourth year?
  - First month of fourth year = month 37!

$$f_{37} = s_n + (37 - 36) \times b_n = 280 + 26 = 306$$

# Exponential smoothing (DES)

$$s_t = \alpha x_t + (1-\alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

$$f_{n+m} = s_n + m b_n$$

$f_{t+1} = s_t + b_t$ equivalent to $f_t = s_{t-1} + b_{t-1}$

- **Exercise**
  - Go back to the Excel file “Week 5 - SES and DES exercise.xlsx”
  - Go the second sheet and insert the formulas of DES to compute the *smoothed and trend sequences* associated to the time series, plus the *forecast*
  - Fill in also the “*squared error*” column and use it to compute the error measure $\sqrt{\frac{\sum(f_i - x_i)^2}{n-2}}$
  - Compute the forecast for the following year (from 37 to 48)
  - Try different values for the smoothing factors $\alpha, \beta$ and see how those change the plot, the forecasts and the value of the error measure

# Summary

- *Forecasting*
  - Using past data to predict future values
- *Simple exponential smoothing*
  - when the level of the time series changes slowly in time
  - one smoothing factor $\alpha$ (data)
- *Double exponential smoothing*
  - when there is a linear trend in the data
  - two smoothing factors $\alpha, \beta$ (data; trend)

# Practical assignment – Part 3

- Implement **SES** and **DES** to do forecasting on the data series of *Chapter 8* (Sword)
  - Find the **optimal** values of the smoothing parameters ($\alpha$ for SES; $\alpha$ and $\beta$ for DES)
  - Using the optimal values of the parameters, compute the **forecasts** for the demand of the following year (time steps 37-48)
  - **Visualize** the results (without using Excel), similarly to figures 8-9 and 8-19 of the book
  - Detailed description of the assignment is on N@tschool