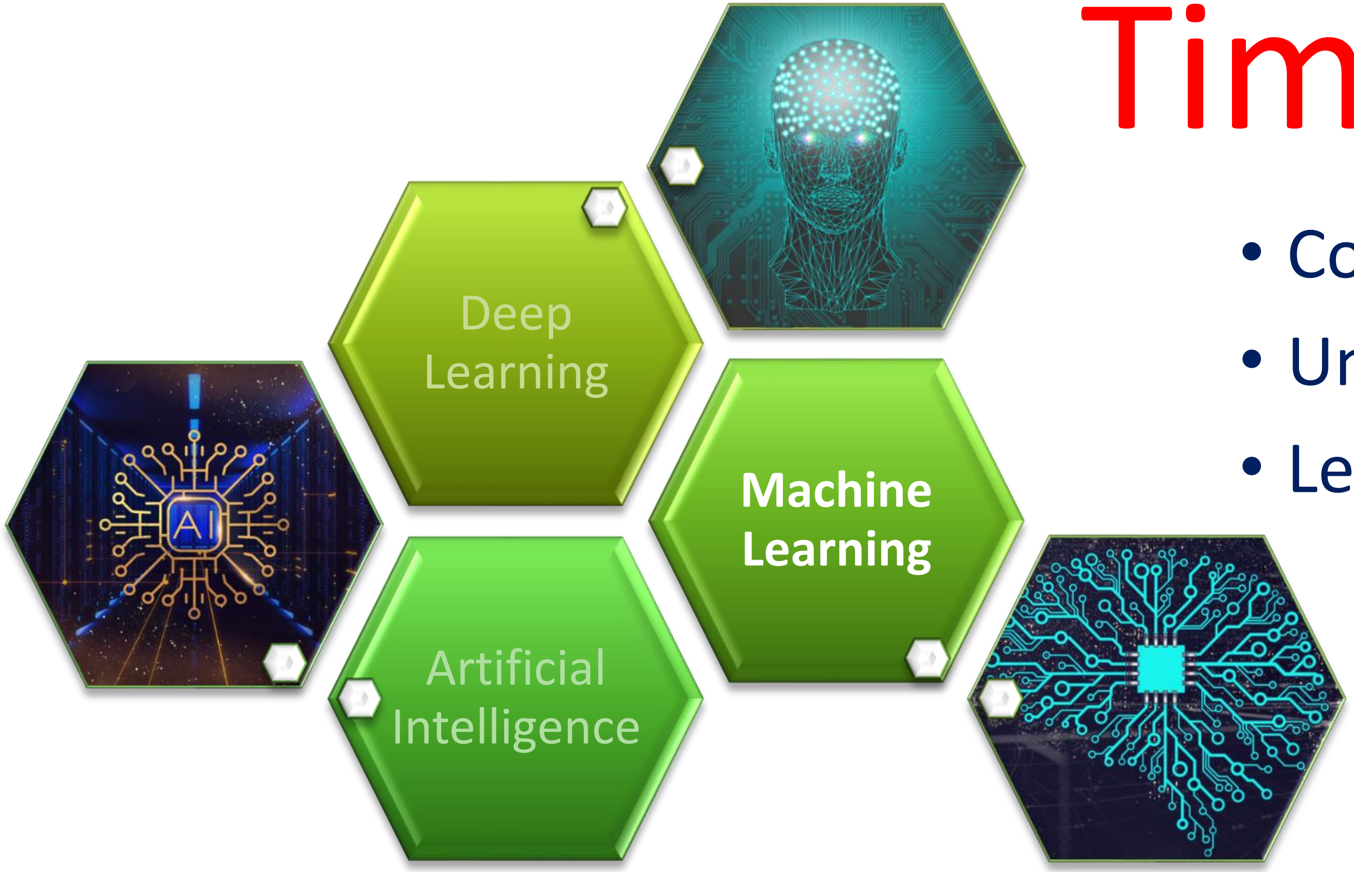


Time Series

- Course Code: CSE471
- Unit 1: **Time Series: An Introduction**
- Lecture 3: Data cleaning.

Smoothing Data

- Often real-world time series data is smoothed before analysis, especially for visualizations that aim to extract useful information out of it.

Why smoothing is required?

- Data preparation
- Feature generation
- Prediction
- Visualization

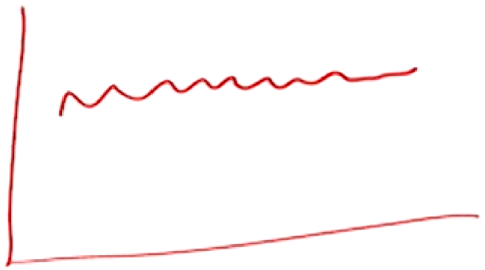
Exponential Smoothing Methods

In these methods, the weight assigned to a previous period's demand decreases exponentially as that data gets older.

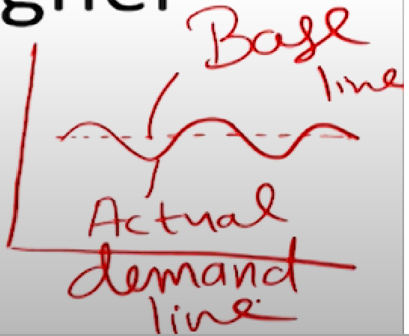
It means recent demand data receives a higher weight than does older demand data.

Exponential Smoothing Methods

In these methods, the weight assigned to a previous period's demand decreases exponentially as that data gets older.



It means recent demand data receives a higher weight than does older demand data.

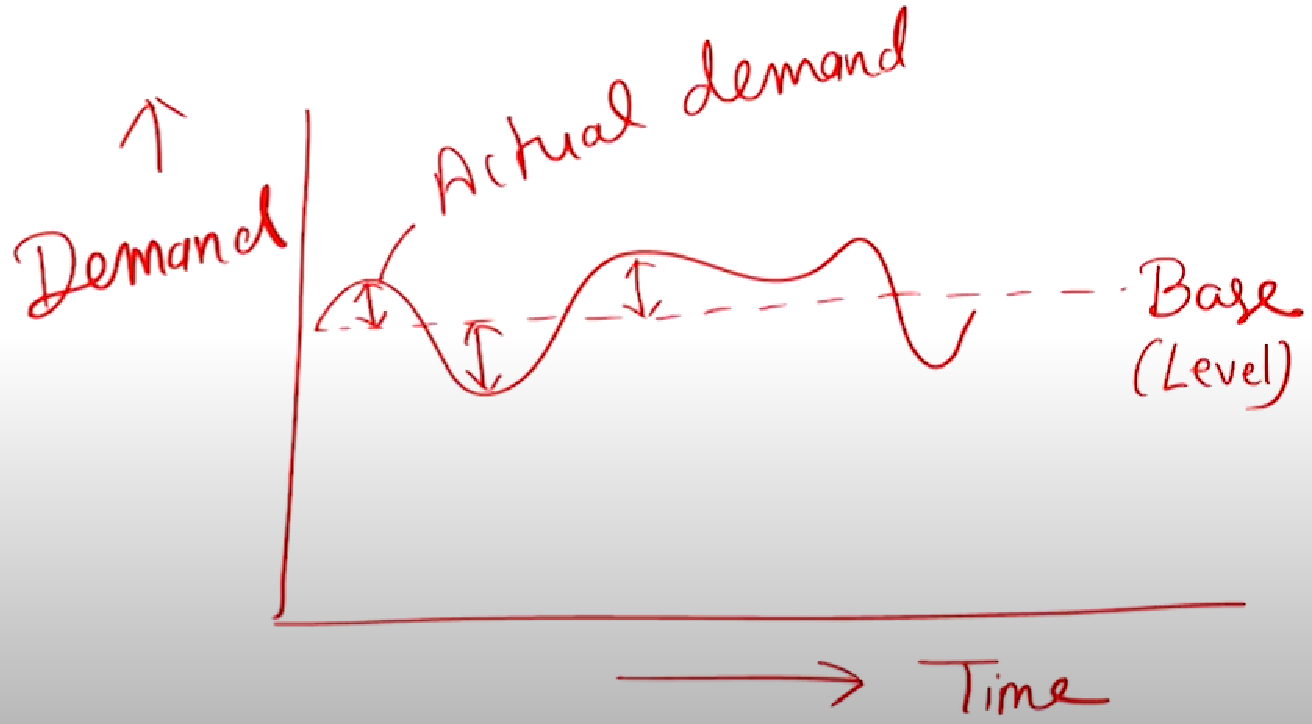


Simple moving av. → Wt moving av. → Exponential Smoothing.

When to use Exponential Smoothing

- When you are forecasting for a large number of items.
- The forecasting horizon is relatively short.
- There is little outside information available about cause and effect.
- Small effort in forecasting is desired. Efforts is measured by both a method's ease of application and by the computational requirements.
- Updating of the forecast as new data becomes available is easy.
- It is desired that the forecast is adjusted for randomness and tracks trends and seasonality.

New Base = Previous Bases + α(New Demand- Previous Base)



$S_{Dec,2019} \quad S_{Jan,2020} = F_{Feb,2020}$

$D_{Jan,2020}$

New Base = Previous Bases + α(New Demand- Previous Base)

$$S_t = S_{t-1} + \alpha\left(D_t - S_{t-1}\right)$$

Demand

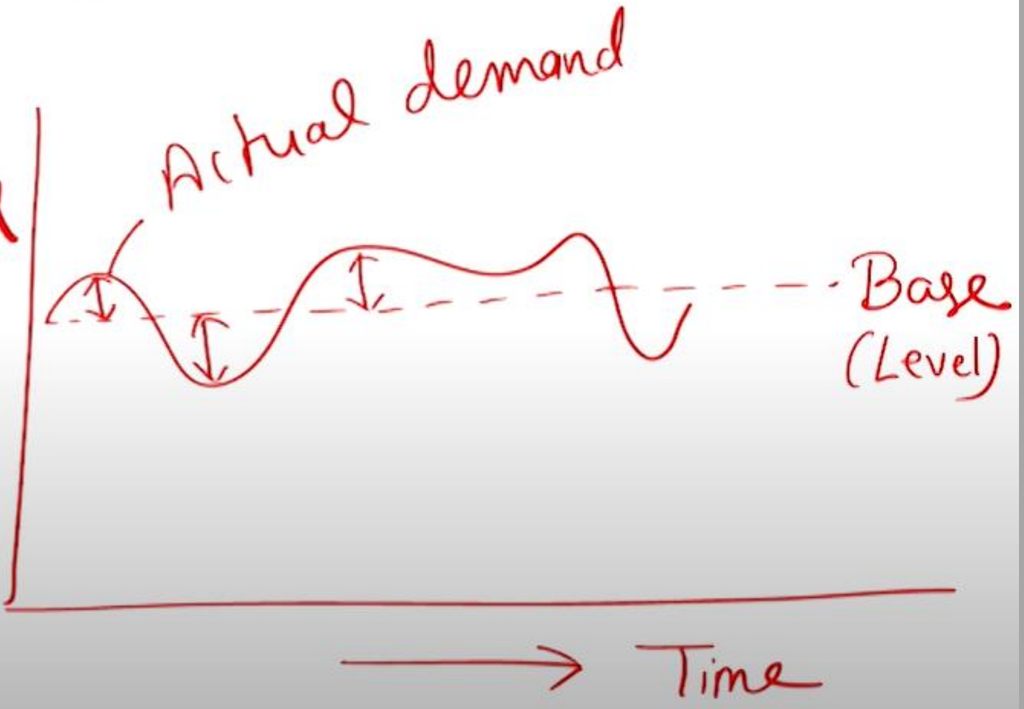
't' is Current period

α = Smoothing Constant

0 to 1

Jan 2020

$$F_{t,1} = S_t$$



$$S_{Dec,2019} \rightarrow S_{Jan,2020} = F_{Feb,2020}$$

$$D_{Jan,2020}$$

New Base = Previous Bases + α(New Demand- Previous Base)

$$S_t = S_{t-1} + \alpha(D_t - S_{t-1}) = \alpha D_t + (1-\alpha) S_{t-1}$$

Demand

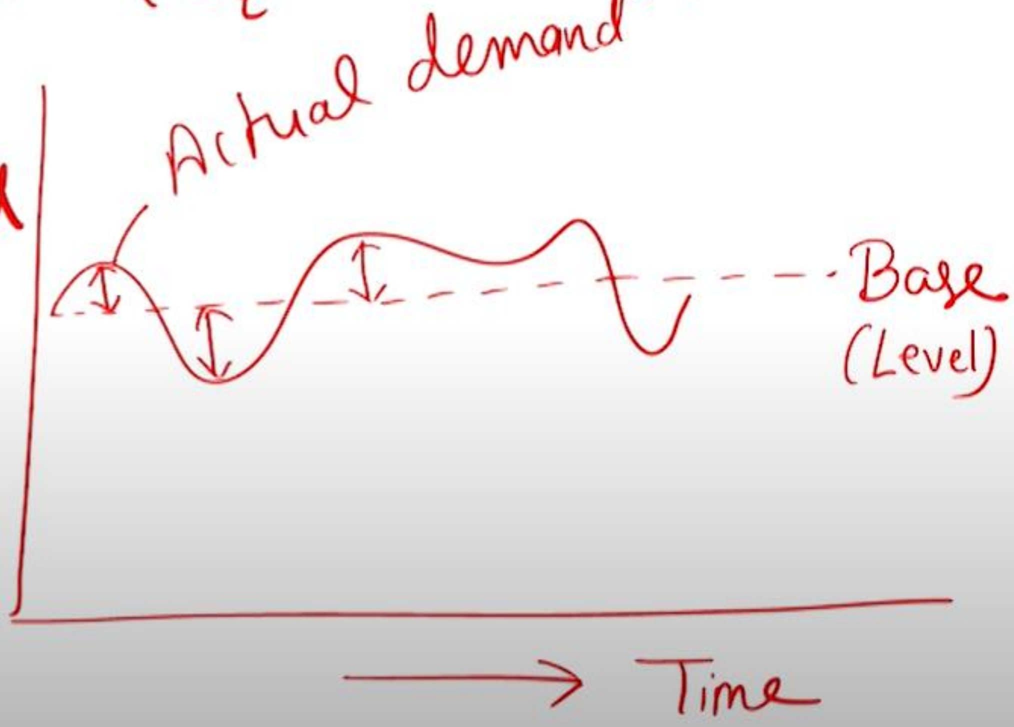
't' is current period

α = Smoothing Constant

0 to 1.

Jan 2020

$$F_{t,1} = S_t$$



- The smoothing constant, α, must be between 0.0 and 1.0. Popular Values .1 to .3
- A large α provides a high impulse response forecast. $\alpha = 0$, $\alpha = .1$
- A small α provides a low impulse response forecast.

$$S_{Dec,2019} \rightarrow S_{Jan,2020} = F_{Feb,2020}$$

$$D_{Jan,2020}$$

New Base = Previous Bases + α(New Demand- Previous Base)

$$S_t = S_{t-1} + \alpha(D_t - S_{t-1}) = \alpha D_t + (1-\alpha) S_{t-1}$$

Demand

$\alpha = 0$

$S_t = S_{t-1}$

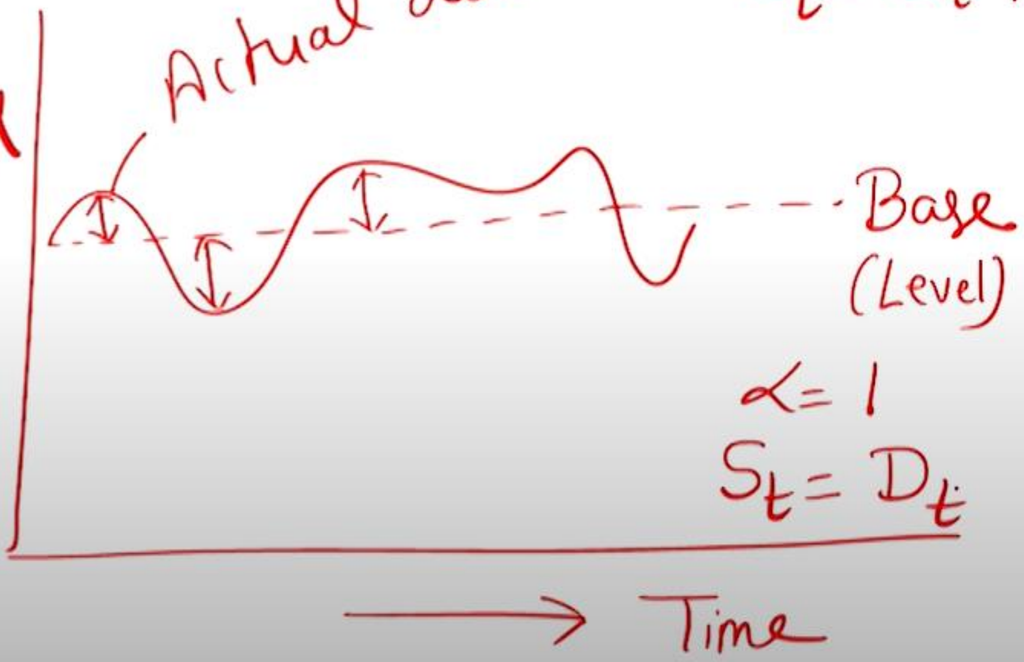
't' is current period

α = Smoothing Constant

0 to 1.

Jan 2020

$F_{t,1} = S_t$



- The smoothing constant, α, must be between 0.0 and 1.0.

Popular Values .1 to .3

- A large α provides a high impulse response forecast.

.8, .9 or 1.0

$\alpha = 0$, $\alpha = 1$

- A small α provides a low impulse response forecast.

- Moving Average
  - Representative Historical Data

Day	Calls	Day	Calls
1	159	7	203
2	217	8	195
3	186	9	188
4	161	10	168
5	173	11	198
6	157	12	159

- Moving Average

Use the moving average method with an AP = 3 days to develop a forecast of the call volume in Day 13.

$$F_{13} = (168 + 198 + 159)/3 = 175.0 \text{ calls}$$

$$\begin{array}{r} \cdot 6 \\ \cdot 3 \\ \cdot 1 \\ \hline \sum W = 1 \cdot 0 \end{array}$$

- Weighted Moving Average

Use the weighted moving average method with an AP = 3 days and weights of .1 (for oldest datum), .3, and .6 to develop a forecast of the call volume in Day 13.

$$F_{13} = \cdot 6\, D_{12} + \cdot 3\, D_{11} + \cdot 1\, D_{10}$$

$$F_{13} = \underbrace{\overset{D_{10}}{.1(168)} + \overset{D_{11}}{.3(198)} + \overset{D_{12}}{.6(159)}}_{\sum W = 1} = 171.6 \text{ calls} \checkmark \simeq 172 \text{ Calls.}$$

Note: The WMA forecast is lower than the MA forecast because Day 13's relatively low call volume carries almost twice as much weight in the WMA (.60) as it does in the MA (.33).

$\alpha = ?$

- **Exponential Smoothing**

  If a smoothing constant value of .25 is used and the exponential smoothing forecast for Day 11 was 180.76 calls, what is the exponential smoothing forecast for Day 13?

$$F_{12} = 180.76 + .25(198 - 180.76) = 185.07$$

$$F_{13} = 185.07 + .25(159 - 185.07) = 178.55$$

$F_{13} = S_{12}$

$S_t = S_{t-1} + \alpha(D_t - S_{t-1})$ or $\alpha D_t + (1-\alpha)S_{t-1}$

$\alpha = ?$

- **Exponential Smoothing**

If a smoothing constant value of .25 is used and the exponential smoothing forecast for Day 11 was 180.76 calls, what is the exponential smoothing forecast for Day 13?

$$F_{12} = 180.76 + .25(198 - 180.76) = 185.07$$

$$F_{13} = 185.07 + .25(159 - 185.07) = 178.55$$

$F_{13} = S_{12}$ $D_{12} = 159$ $S_t = S_{t-1} + \alpha(D_t - S_{t-1})$ or $\alpha D_t + (1-\alpha)S_{t-1}$

$S_{12} = S_{11} + \alpha(D_{12} - S_{11})$ S_{11}? $\alpha = ?$

- **Exponential Smoothing**

If a smoothing constant value of .25 is used and the exponential smoothing forecast for Day 11 was 180.76 calls, what is the exponential smoothing forecast for Day 13?

$\boxed{F_{11} = 180.76}$

$$F_{12} = 180.76 + .25(198 - 180.76) = 185.07$$

$$F_{13} = 185.07 + .25(159 - 185.07) = 178.55$$

$F_{13} = S_{12}$ $D_{12} = 159$ $S_t = S_{t-1} + \alpha(D_t - S_{t-1})$ or $\alpha D_t + (1-\alpha)S_{t-1}$

$S_{12} = S_{11} + \alpha(D_{12} - S_{11})$ S_{11}? $\alpha = ?$

- Exponential Smoothing

If a smoothing constant value of .25 is used and the exponential smoothing forecast for Day 11 was 180.76 calls, what is the exponential smoothing forecast for Day 13?

$F_{11} = S_{10} = 180.76$

$S_{11} = S_{10} + \alpha(D_{11} - S_{10})$

$S_{10} + \alpha(D_{11} - S_{10})$

$S_{11} = F_{12} = 180.76 + .25(198 – 180.76) = 185.07$ ✓

$S_{12} = F_{13} = \underset{S_{11}}{185.07} + .25(\underset{D_{12}}{159} – \underset{S_{11}}{185.07}) = 178.55$ ✓

175

172

Addressing seasonality in data

Seasonality is the repeated variation of a time series over a fixed period of time. For example, sales of ice cream may be higher in the summer and lower in the winter. **Seasonality can be removed from time series data for several reasons:**

- **To make the data easier to analyze**. Many statistical and machine learning algorithms are designed to work with stationary data. Seasonality can make the data non-stationary, which can make it more difficult to analyze.
- **To improve the accuracy of forecasts**. Forecasts of time series data are typically more accurate when the data is stationary. This is because seasonality can introduce noise into the data, which can make it more difficult to predict future values.
- **To make the data more comparable across different time periods.** Seasonality can make it difficult to compare data from different time periods. For example, if you are comparing sales data from different years, you may want to remove seasonality so that you can compare the underlying trends.

Preventing unintentional lookaheads

- Lookahead bias -occur in time series forecasting when the model is trained on data that includes future values.
- Preventing lookahead bias is important because it can lead to inaccurate forecasts. When a model is trained on data that includes future values, it is essentially learning to predict the future based on the past. This can lead to the model overfitting the data, and making predictions that are not accurate.

Why is it required?

- Lookahead bias is a serious problem because it can lead to inaccurate forecasts. This can have a significant impact on businesses and organizations that rely on time series forecasting. For example, a business that uses time series forecasting to set inventory levels could make incorrect decisions if the forecasts are inaccurate.

Ways to prevent lookaheads

- Use a sliding window. A sliding window only includes data that has already occurred, and it is updated as new data becomes available. This prevents the model from seeing future values.
- Use a causal model. A causal model is a model that only uses data that is causally related to the target variable. This means that the model cannot see future values, because they are not causally related to the target variable.