Forecasting

With Smoothing Models

Sanjay Sane

Naïve Forecast

- Naïve Forecast is the next value as the last observed value
- Whatever is the last value, just use it as the forecast for the next time point

Seasonal Naïve Forecast

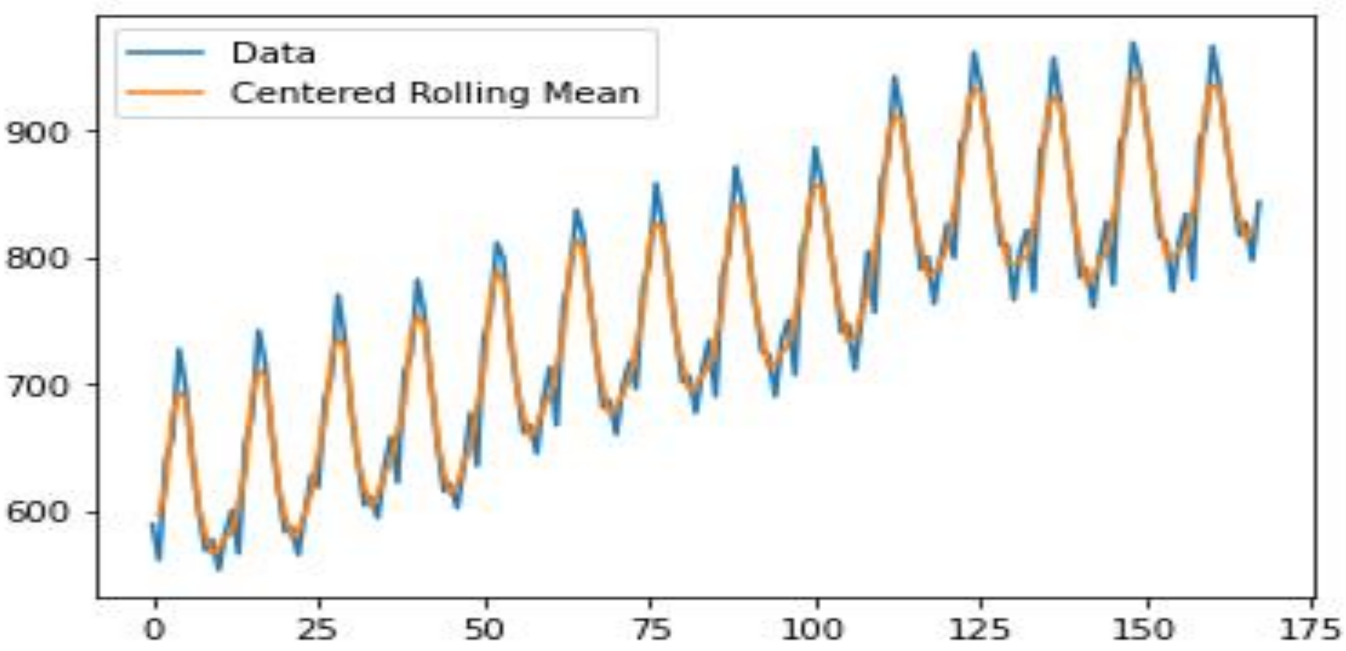
- The seasonal naive forecast is the last observed value from a similar period in the past.
- For example, if we want to know the sales for next Saturday, we can use the sales from the previous week Saturday.

Window Average Forecast

- Window average forecast calculates the average of the last window size(span) values of the series, often called trailing rolling mean
- Centered Rolling Mean, the same average but positioned at the center of the time window can be good for visualization

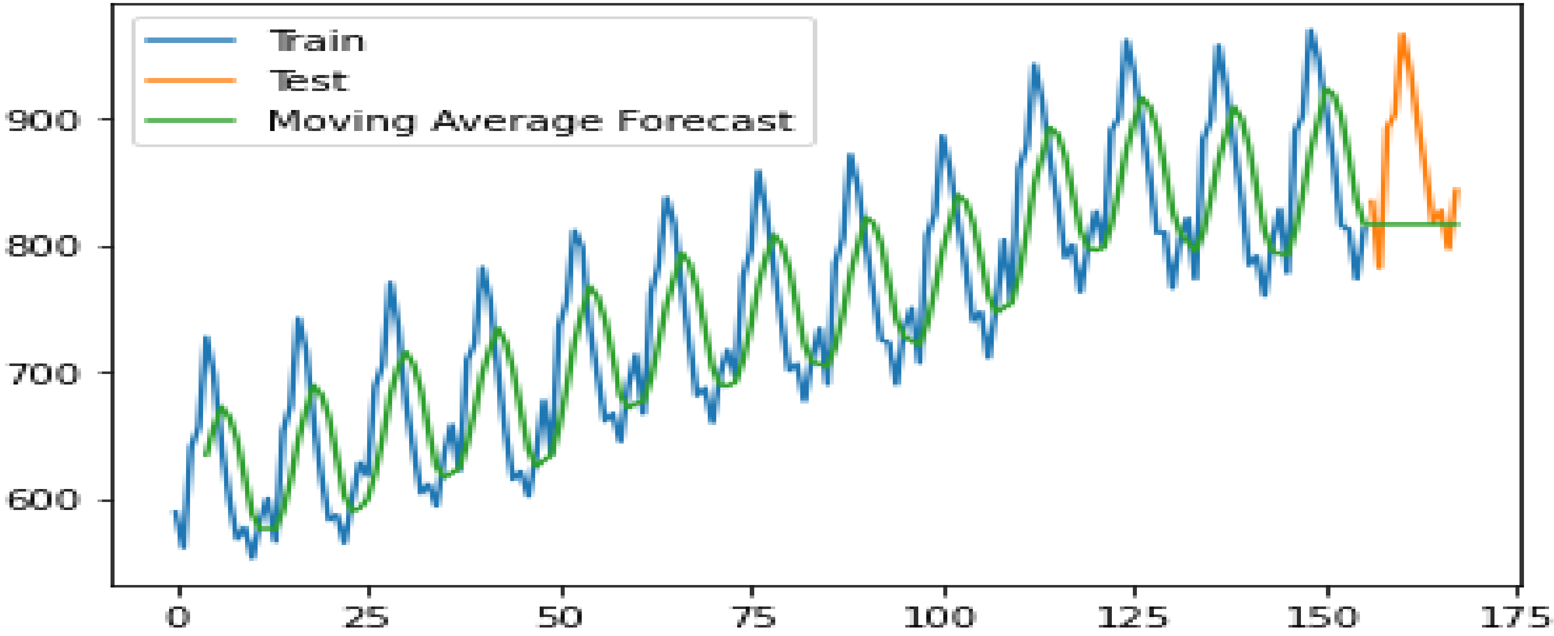
Centered Rolling Mean

```
span=3
fcast = y.rolling(span,center=True).mean()
plt.plot(y, label='Data')
plt.plot(fcast, label='Centered Rolling Mean')
plt.legend(loc='best')
plt.show()
```



Trailing Rolling Mean

```
y_train = df['Milk'][:-12]
y_test = df['Milk'][-12:]
span = 5
fcast = y_train.rolling(span).mean()
MA = fcast.iloc[-1]
MA_series = pd.Series(MA.repeat(len(y_test)))
MA_fcast = pd.concat([fcast,MA_series],ignore_index=True)
```



Sanjay Sane

Forecasting Models

Forecasting Model Types

- We will be covering:
  - Smoothing Models
  - ARIMA Models

Smoothing Models

- Simple Exponential Smoothing
- Double Exponential Smoothing
  - Holt Linear Trend
  - Holt Exponential Trend
  - Damped Methods
- Triple Exponential Smoothing
  - Holt-Winters Additive
  - Holt-Winters Multiplicative

Simple Exponential Smoothing

- In Simple Exponential Smoothing, weighted average of all past values is taken in such a way that the weights decrease exponentially into past
- Like Moving Average, this method is used for forecasting series that have no trend and no seasonality

$\alpha > \alpha(1-\alpha) > \alpha(1-\alpha)^2 > \ldots$

Calculation

say $\alpha = 0.2$

- The exponential smoother calculates a forecast at time t+1, F_{t+1}:
  $F_{t+1} = \propto y_t + \propto (1 - \propto) y_{t-1} + \propto (1 - \propto)^2 y_{t-2} + \cdots$
  - Where $\propto$ is a constant between 0 and 1 called smoothing constant
- The above equation can also be written as:

$$F_{t+1} = F_t + \propto e_t$$

  - Where F_t is forecast error at time t and e_t is forecast error at time t

Choice of $\propto$

- The smoothing constant $\propto$ determines the rate of learning
- A value close to 1 implies fast learning, i.e. the most recent values influence the forecast most
- A value close to 0 implies slow learning, i.e. the past observations influence the forecast most
- The default values of $\propto$ that have been mostly observed to work well are between 0.1 and 0.2.

Series with Additive Trend

- For series with trend, we can use Holt’s method, also known as double exponential smoothing
- Similar to Simple Exponential Smoothing, the level of the series is estimated from the data and is updated as more data would become available
- Level is estimated using maximum likelihood method

Holt's Linear Trend Method

- The k-step ahead forecast is given by combining the level estimate at time t (L_t) and trend estimate at time t (T_t):

$$F_{t+k} = L_t + kT_t$$

- The level and trend are updated by the equations:

$$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

  - Where α and β are smoothing constants whose values range from 0 to 1 and are set by the user or chosen iteratively by R function holt()

- Level equation shows L_t, Level at time t as weighted average of the observation at time t y_t and within sample one step ahead forecast at time t , $(L_{t-1}+T_{t-1})$
- Trend Equation shows T_t , trend estimate at time t as weighted average of (L_t-L_{t-1}) and T_{t-1}, the previous trend estimate

More Details: http://robjhyndman.com/hyndsight/hw-initialization/

Sanjay Sane

Exponential Trend Method

- The k-step ahead forecast is given by combining the level estimate at time t (L_t) and trend estimate at time t (T_t):

$$F_{t+k} = L_t \times {T_t}^k$$

- The level and trend are updated by the equations:

$$L_t = \propto y_t + (1 - \propto)(L_{t-1} \times T_{t-1})$$

$$T_t = \beta \left(\frac{L_t}{L_{t-1}} \right) + (1 - \beta) T_{t-1}$$

  - Where α and β are smoothing constants whose values range from 0 to 1 and are set by the user or chosen iteratively by R function holt()

- Level equation shows L_t, Level at time t as weighted average of the observation at time t y_t and within sample one step ahead forecast at time t , $(L_{t-1} \times T_{t-1})$

- Trend Equation shows T_t , trend estimate at time t as weighted average of (L_t / L_{t-1}) and T_{t-1}, the previous trend estimate

Damped Trend Methods

- It has been observed that Holt’s Linear Trend and Exponential Trend tend to over-forecast for longer forecast horizons
- Gardner and McKenzie (1985) suggested a parameter that dampens the trend line to a flat line some time in the future
- Methods with damped trend have been proven to be more successful when forecasts are to be predicted by automatic process
- There are two types of damped trend methods:
  - Additive Damped Trend
  - Multiple Damped Trend

Additive Damped Trend

- In association with the smoothing parameters α and β, damped methods also include a damping parameter φ; 0 <φ< 1 as:

$$F_{t+k} = L_t + (\varphi + \varphi^2 + \cdots + \varphi^k)T_t$$
$$L_t = \propto y_t + (1 - \propto)(L_{t-1} + \varphi T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)\varphi T_{t-1}$$

- If φ=1 then the method is Holt's Linear Method

Multiplicative Damped Trend

- Taylor(2003) introduced a damping parameter to the exponential trend

$$F_{t+k} = L_t \times {T_t}^{(\varphi+\varphi^2+\cdots+\varphi^k)}$$
$$L_t = \alpha y_t + (1-\alpha)L_{t-1} \times {T_{t-1}}^{\varphi}$$
$$T_t = \beta\left(\frac{L_t}{L_{t-1}}\right) + (1-\beta){T_{t-1}}^{\varphi}$$

Holt-Winters Seasonal Method

- This method comprises of the forecast equation and three smoothing equations each for level, trend and seasonal component
- We use m to denote the period of season
- The additive method of Holt-Winters can be preferred when the seasonal variations are roughly constant through the series
- The multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

Holt-Winters Additive Method

- The component form of the model:

$$F_{t+k} = L_t + kT_t + S_{t-m+k_m{}^+}$$
$$L_t = \propto (y_t - S_{t-m}) + (1 - \propto)(L_{t-1} + T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$
$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-m}$$

Where

S_t : Seasonal Estimate at time t

${k_m}^+$: [(k-1) mod m)]+1 which ensures that the estimates of the seasonal indices used for forecasting come from the final year

Holt-Winters Multiplicative Method

- The component form of the model: (Additive Trend)

$$F_{t+k} = (L_t + kT_t)S_{t-m+k_m{}^+}$$
$$L_t = \propto \left(\frac{y_t}{S_{t-m}}\right) + (1 - \propto)(L_{t-1} + T_{t-1})$$
$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$
$$S_t = \gamma\left(\frac{y_t}{L_t}\right) + (1 - \gamma)S_{t-m}$$

Where

S_t : Seasonal Estimate at time t

${k_m}^+$: [(k-1) mod m)]+1 which ensures that the estimates of the seasonal indices used for forecasting come from the final year

Questions?