Announcements:

① Wed, 24 Dec → postponed (No class)

↳ 5 Jan, Monday

② GRADING CRITERIA:

- PROJECT SUBMISSION – 30%
  - ↳ Details by Wednesday, 24 Dec
  - ↳ Submission Date: Before 15 Jan
- VIVA – 30% ("15 Jan")
- END-TERM – 40%
  - ↳ End of January

* Generating FORECASTS:

Evaluation Metric → MAPE

"Mean Absolute Percentage Error"

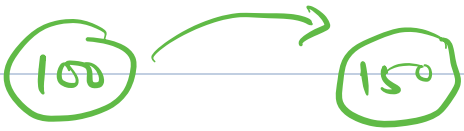
$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

y_i = Actual value

$\hat{y}_i$ = Forecasted value

N = No. of forecasts.

Example:



% increase = 50%

$$\frac{150 - 100}{100} = \frac{50}{100} = 0.5 \simeq 50\%$$

→ Using MAPE, we get a percentage of correct forecasts.

→ What if '$y_i = 0$'

↳ MAPE → undefined

↳ We use other methods like → MSE, RMSE

MAPE Desired Value is "< 5%"

Forecasting

Mobile Manufacturing Company (Apple)

(Data Scientist)

"How much will we sell in next 12 months"

* • Under-forecast
- Stock shortage
- Sales lost

* • Over-forecast
- inventory loss
- storage cost

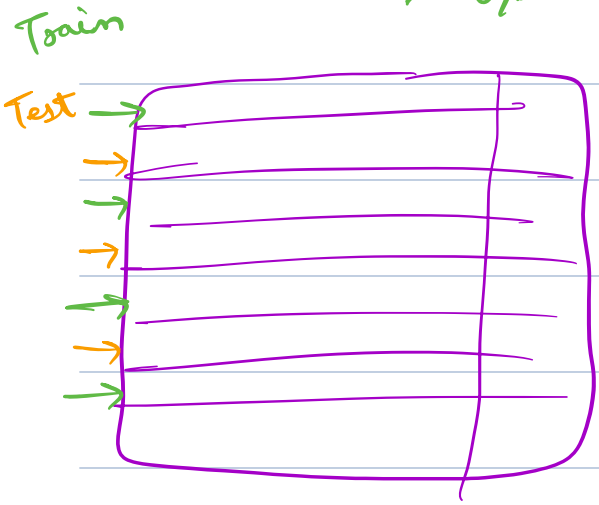
→ Give better forecasts, by analysing historic Time-series data.

*TRAIN - TEST SPLIT :

DATA → 80% → Train the Model

DATA → 20% → Test the Model

"Random"



Train-test split → Normal ML → Random

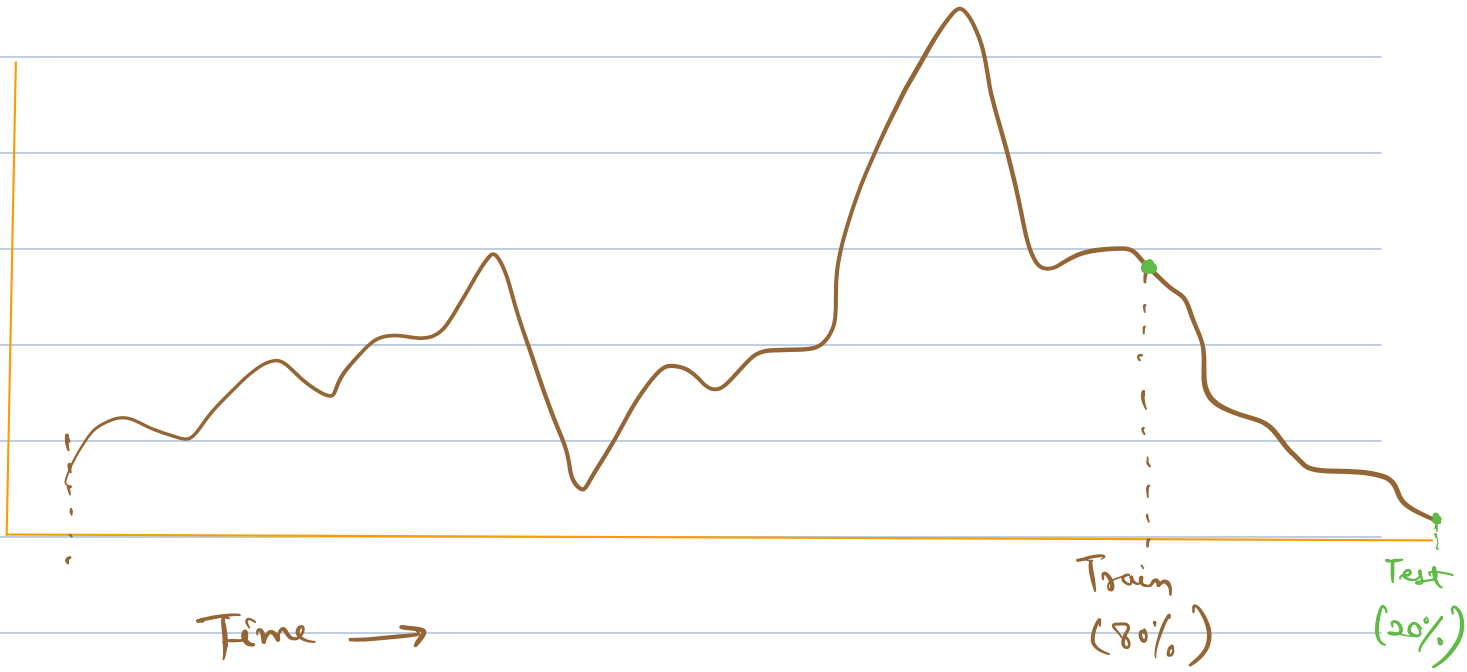
- Can we directly do train-test split on time-series data:
  - No
- Can we shuffle Time?
  - No

→ Train-Test split in Time Series analysis follow time sequence.

PAST data in sequence → FUTURE data

TRAIN | TEST

18 years of DATA → 17 years → Train set
→ Next 12 months → Test set



- We don't have 'train_y' and 'test_y' in train-test split?

→ We only have 1 column (Sales) in the dataset

→ We only have dates (Time), but we are using it as index

↳ We cannot create a 'y' column, if there are no columns.

FORECASTING METHODS :

① Simple Mean Forecasting:

→ All forecasted values are mean of Time Series.

→ Gives a flat line for forecasted values.

→ Very bad model / basic method.

Person - X : Gaining Weight

WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 5
60	65	70	65	65
			75	75
			70	70

② Naive Approach :

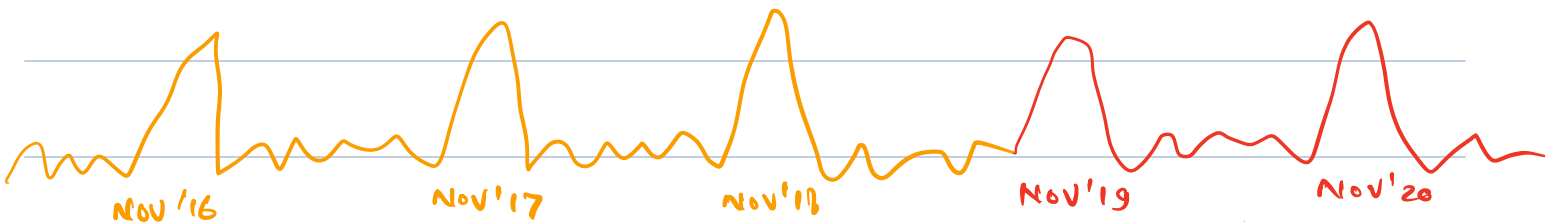
→ Take the last value of train data.

→ Forecast all next values as the last value.

→ Not an intelligent method.

③ Seasonal Naive Forecast :

Woolen clothes company → Sales will be higher in WINTERS



↳ previous same season value

[Last value of train data but of same season.]

→ Business repeat itself by seasons, not randomly.

→ Take last value of same season & forecast.

→ Better than previous 2 methods

→ Captures seasonality → Misses trend

→ Uses real historic patterns

↳ Business acceptable ✓

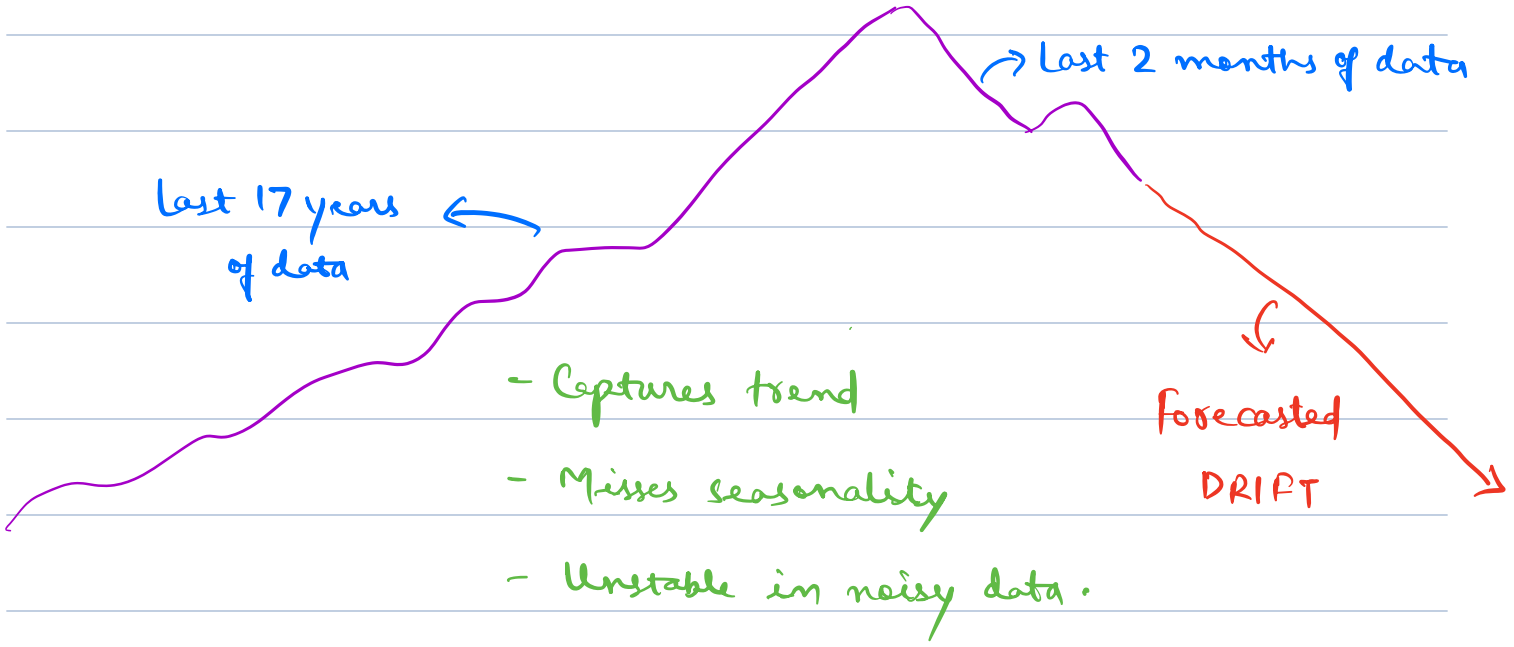
④ DRIFT method:

↳ We extend the line of time-series.

↳ Increasing / Decreasing

→ one last month can destroy the trend / slope.

→ Drift is extremely sensitive.



$y-t$ → Starting point

m → Slope

h → Steps into the future, I am predicting

- Why are we studying simple methods?

→ Sometimes, one of these simplest methods can be the best method.

→ These are also known as 'Benchmark' methods.

↳ Any method, you develop should be better than these models.

SMOOTHING METHODS :

1. Moving Average Forecasting → rolling (3, center = False)
2. Simple Exponential Smoothing
3. Holt's Method → Double Exponential Smoothing
4. Holt - Winters Method → Triple Exponential Smoothing