Supervised Machine Learning: Regression

Time Series Prediction

Time Series Data

- Time series is a sequential set of data points, measured typically over successive times
- Time series data are simply a collection of observations gathered over time
- Time series is a time oriented sequence of observations on a variable of interest
- It is clearly structured and numeric in nature
- Time series data is collected at some intervals
  - These intervals can be as large as years or as small as seconds
- Example:
  - Weekly sales – time interval is week
  - Daily temperature in Kamand – time interval is day

Time Series Data

- Time series is a sequential set of data points, measured typically over successive times
- Time series data are simply a collection of observations gathered over time
- Time series is a time oriented sequence of observations on a variable of interest
- It is clearly structured and numeric in nature
- Example:
  - Weekly sales – time interval is week
  - Daily temperature in Kamand – time interval is day
- Time series data is collected at some intervals
  - These intervals can be as large as years or as small as seconds

Date/ Time	Temperature (C)/ Humidity (%)	Pressure (Pa)	Rain (inches)	Light Intensity (lux)	Accelerations (g)	Force (N)	Moisture (%)
2017-09-06 18:44:32	23.00,56.00	617.64	0.01	3	0.52,0.31,-0.80,0.00,0.00,0.00,31.36,-159.01	0.02	81.00
2017-09-06 18:33:32	24.00,58.00	619.47	0.01	12	0.52,0.30,-0.79,0.00,0.00,0.00,31.45,-159.12	0.02	82.00
2017-09-06 18:22:39	24.00,58.00	623.37	0.00	71	0.52,0.31,-0.80,0.00,0.00,0.00,31.35,-158.88	0.02	83.00
2017-09-06 18:11:31	25.00,60.00	627.02	0.05	194	0.51,0.31,-0.80,0.00,0.00,0.00,30.80,-159.00	0.02	81.00

Time Series Data

- Time series data is given as:

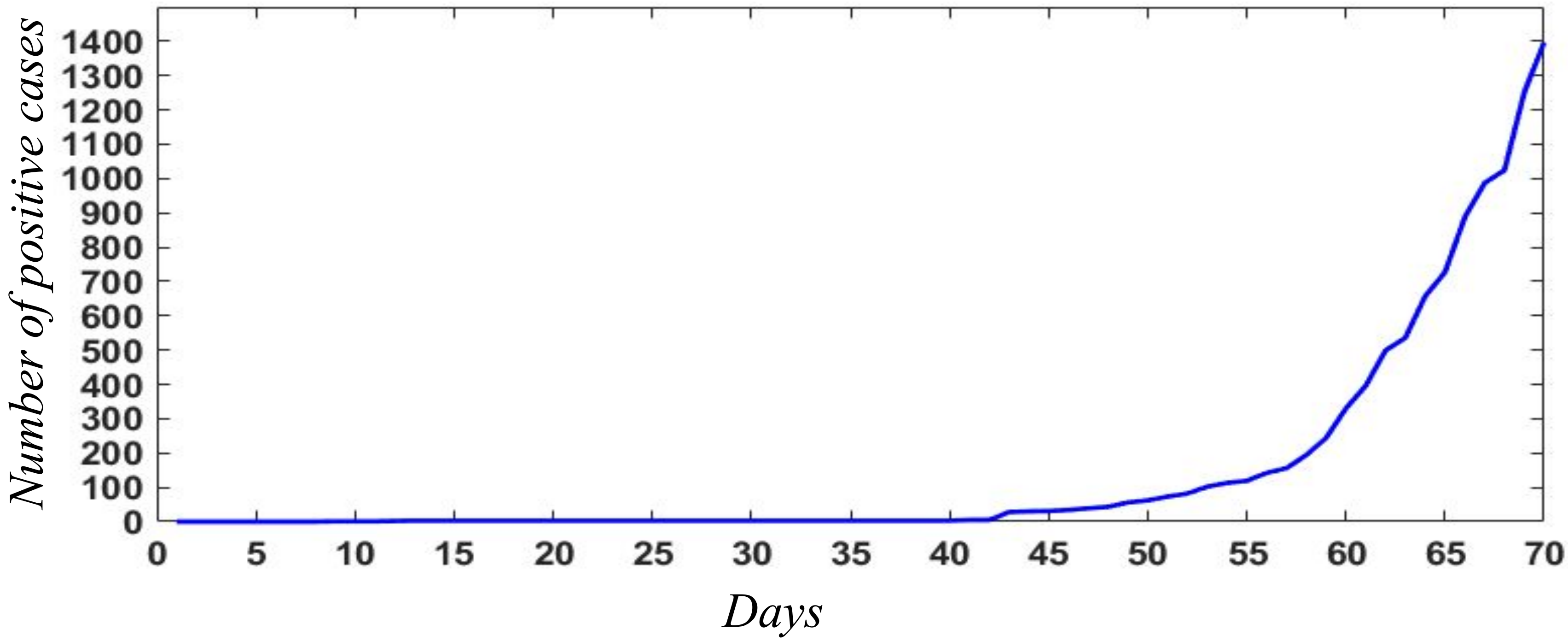
$$\mathbf{X} = (x_1, x_2, \ldots, x_t, \ldots, x_T)$$

  – x_t is the observation at time t
  – T be the number of observations

- Scope: We consider single variable x_t

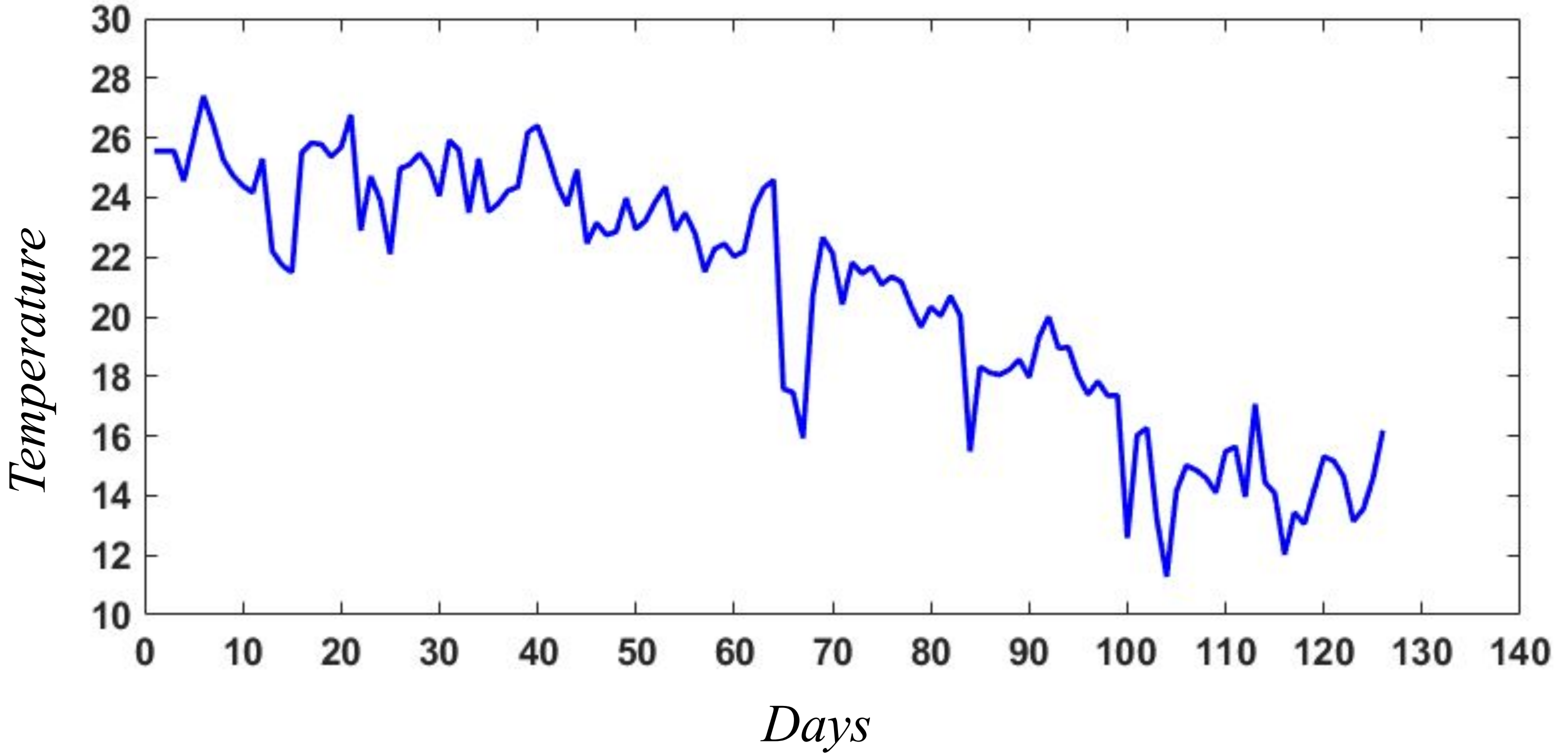
Time Series Data

- Trend: Shows how data moves over a period of time
  - COVID positive cases in India between 22 Jan 2020 to 31 March 2020



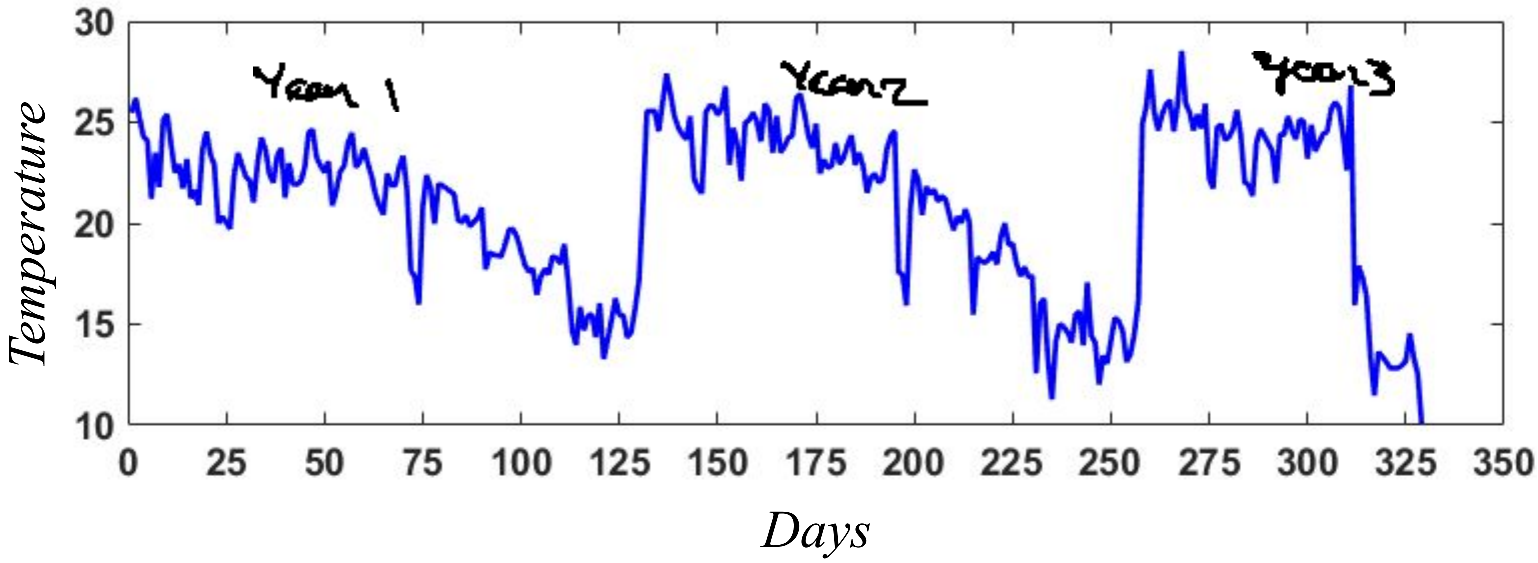
Time Series Data

- Trend: Shows how data moves over a period of time
  - Daily temperature at IIT Mandi from June-Nov 2018



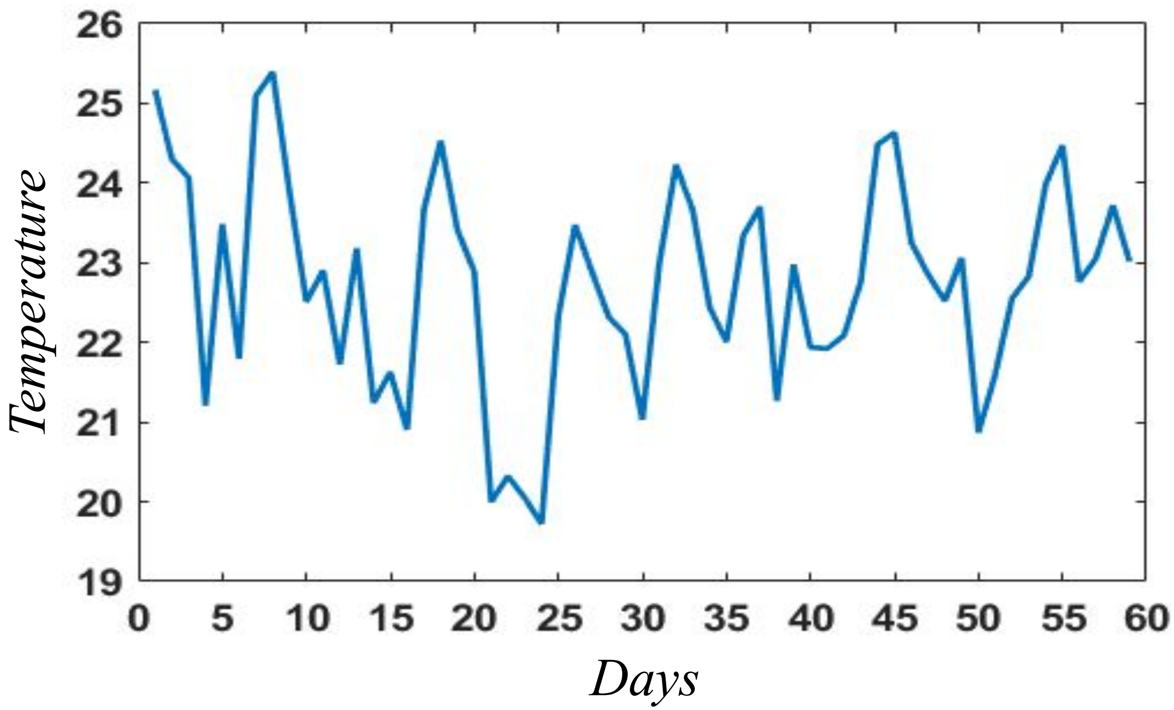
Time Series Data

- Seasonality: A type of pattern which repeats over a specific period of time
  - Daily temperature recorded in IIT Mandi for 3 years
    - Duration of recorded: July-Nov (2017-2019)



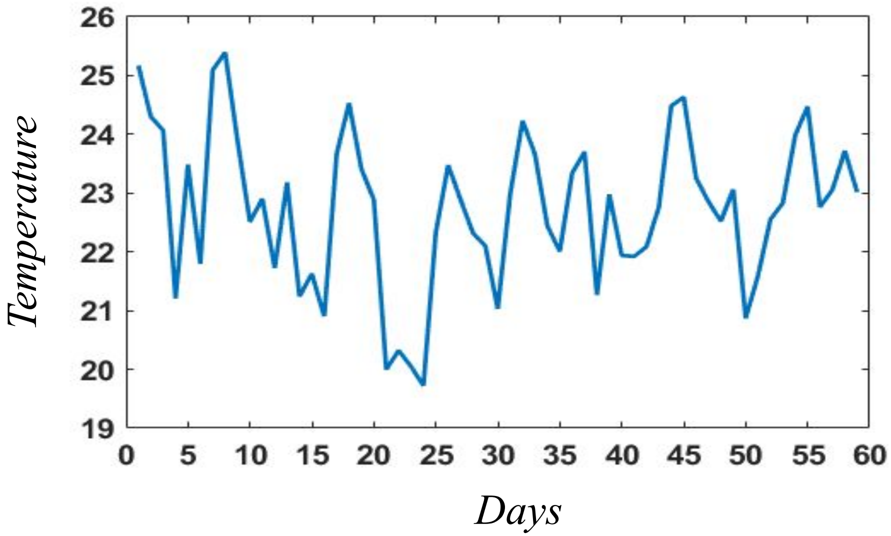
Time Series Data

- Random or error: Series does not have any trend, seasonality or cyclic component
  - Daily temperature recorded in IIT Mandi (1 July - 30 August 2019)



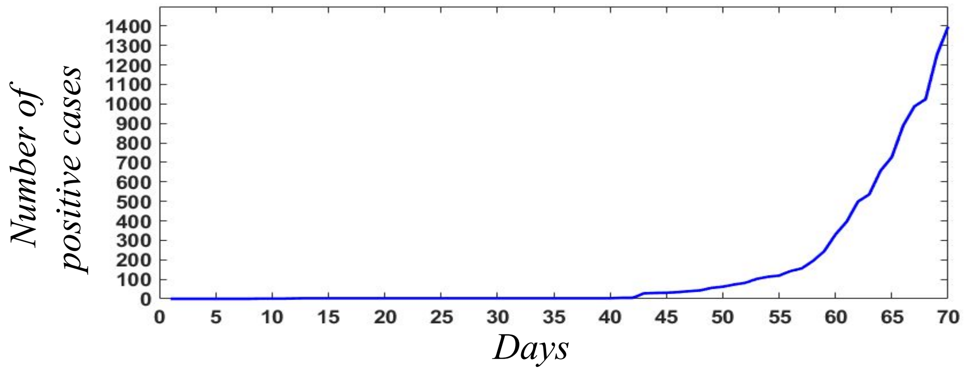
Stationary Time Series

- Stationary time series:
  - Statistical properties remain same at any given interval of time
  - Time independent kind of series
  - Mean and variance should be time independent
    - Mean and variance computed at any one part of the series should be similar to that of the mean and variance computed at another part
  - Stationary time series are easier to predict
- Example:
  - Daily temperature recorded in IIT Mandi (1 July - 30 August 2019)

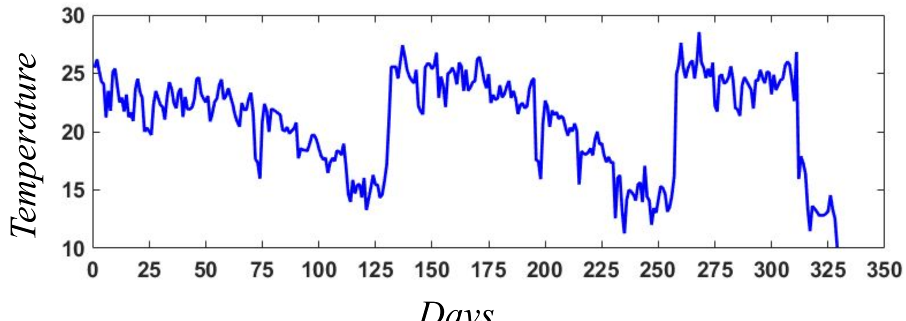


Non-stationary Time Series

- Non-stationary time series: Time series having trends or seasonality
  - Mean and variance are not time independent
- Example:
  - COVID positive cases in India between 22 Jan 2020 to 31 March 2020

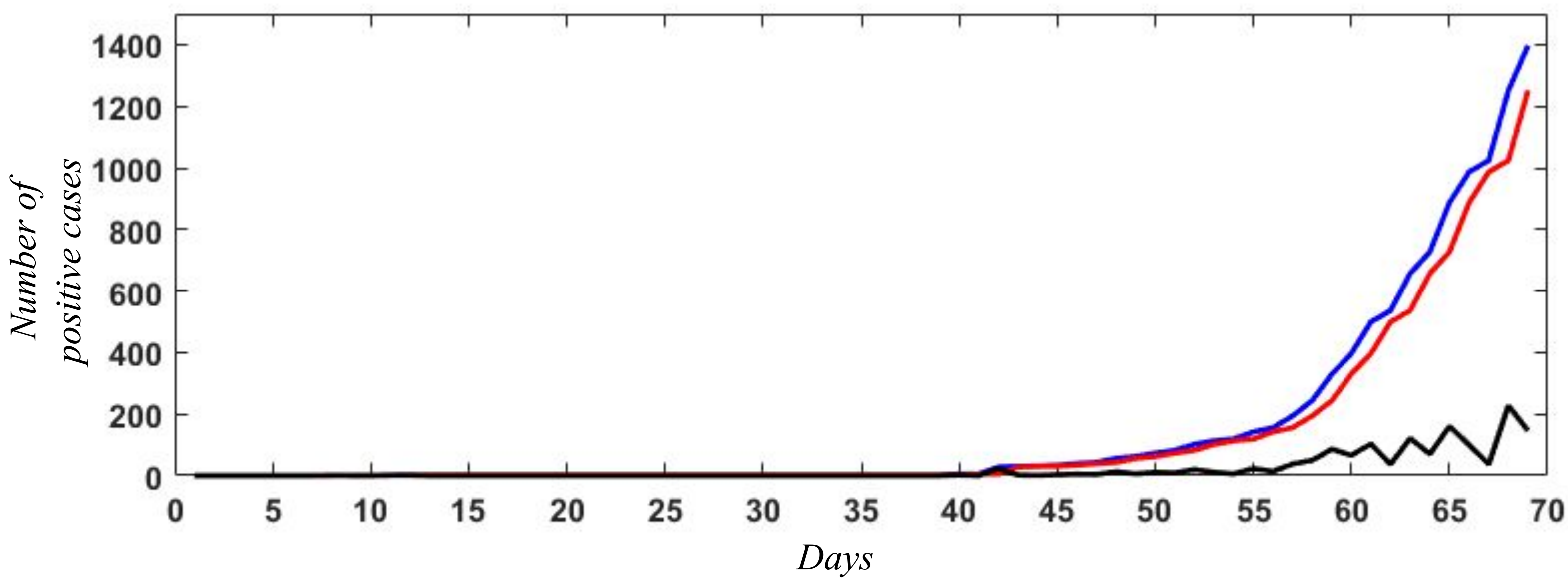


  - Daily temperature recorded in IIT Mandi for 3 years



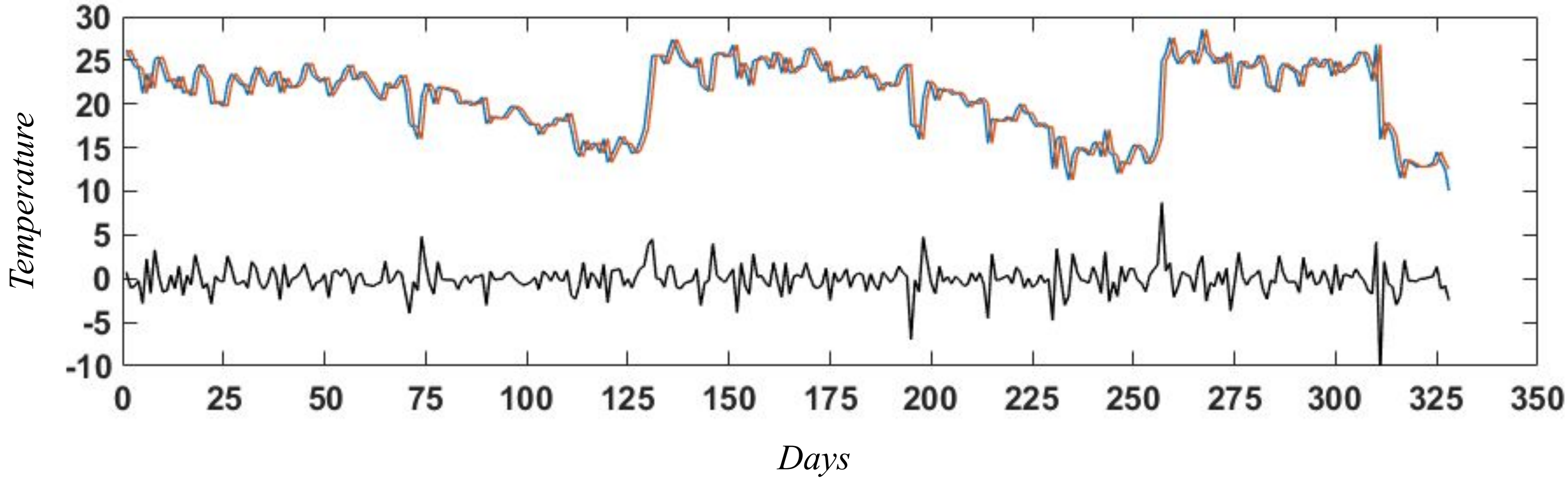
Differencing

- Non-stationary time series are made stationary by differencing
  - Difference between the original series and the lag series
  - Lag is the shift in the time series by a given number of observations
    - Lag 1: Shift by one time step
    - Lag 2: Shift by two time step
  - By differencing, non-stationary time series become more stationary



Differencing

- Non-stationary time series are made stationary by differencing
  - Difference between the original series and the lag series
  - Lag is the shift in the time series by a given number of observations
    - Lag 1: Shift by one unit
    - Lag 2: Shift by two unit
  - By differencing, non-stationary time series become more stationary
    - Mean and variance become almost same in the different parts
  - Example: Daily temperature recorded in IIT Mandi for 3 years



Time Series Data and Dependence

- Time series data is given as:

$$\mathbf{X} = (x_1, x_2, \ldots, x_t, \ldots, x_T)$$

  – x_t is the observation at time t

  – T be the number of observations

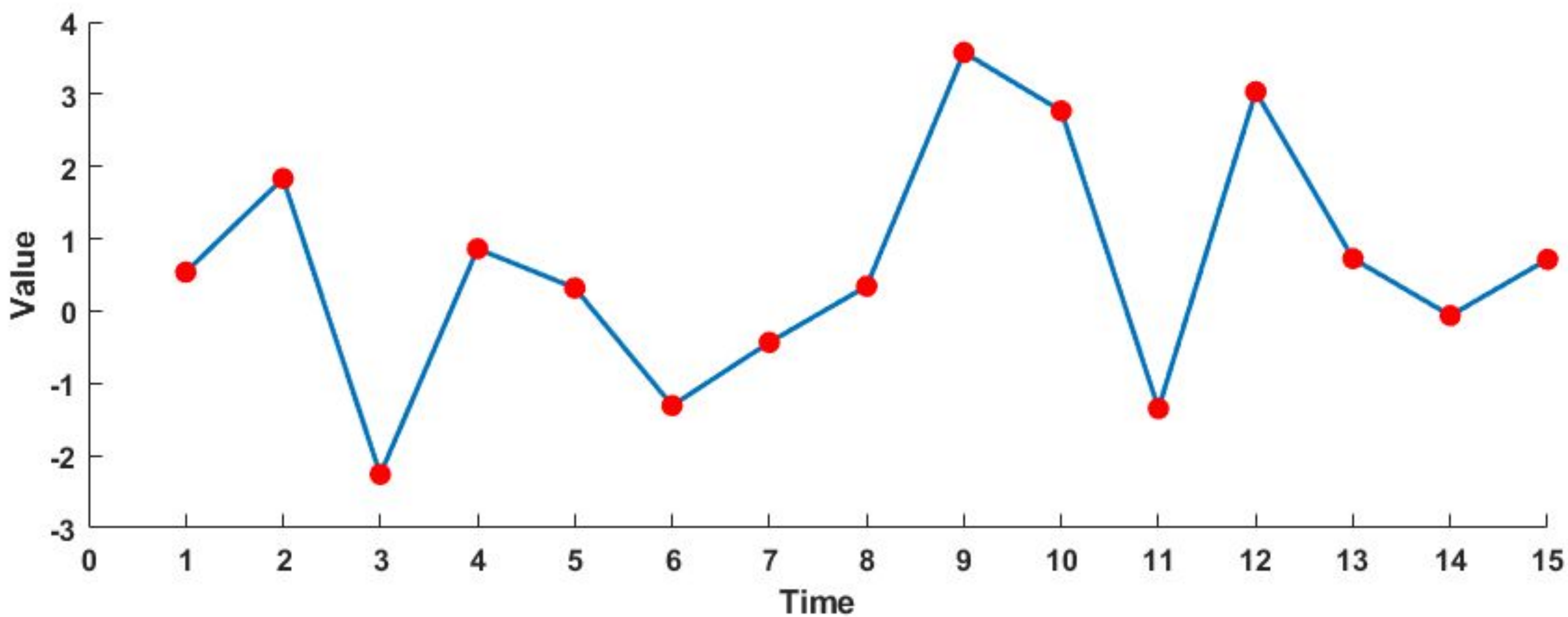
- In time series data, value of each element at time t (x_t) is dependent on the values elements at previous p time steps (x_{t-1}, x_{t-2}, …, x_{t-p}) – p time lag

  – Lag is the shift in the time series by a given number of observations

Time Series Data and Dependence

- Example: Data series in i.i.d
  - x_t is a random number drawn from $\mathcal{N}(0,1)$
- Each element at time t (x_t) is not dependent on the values elements at previous p time steps (x_{t-1}, x_{t-2}, …, x_{t-p}) – p time lag

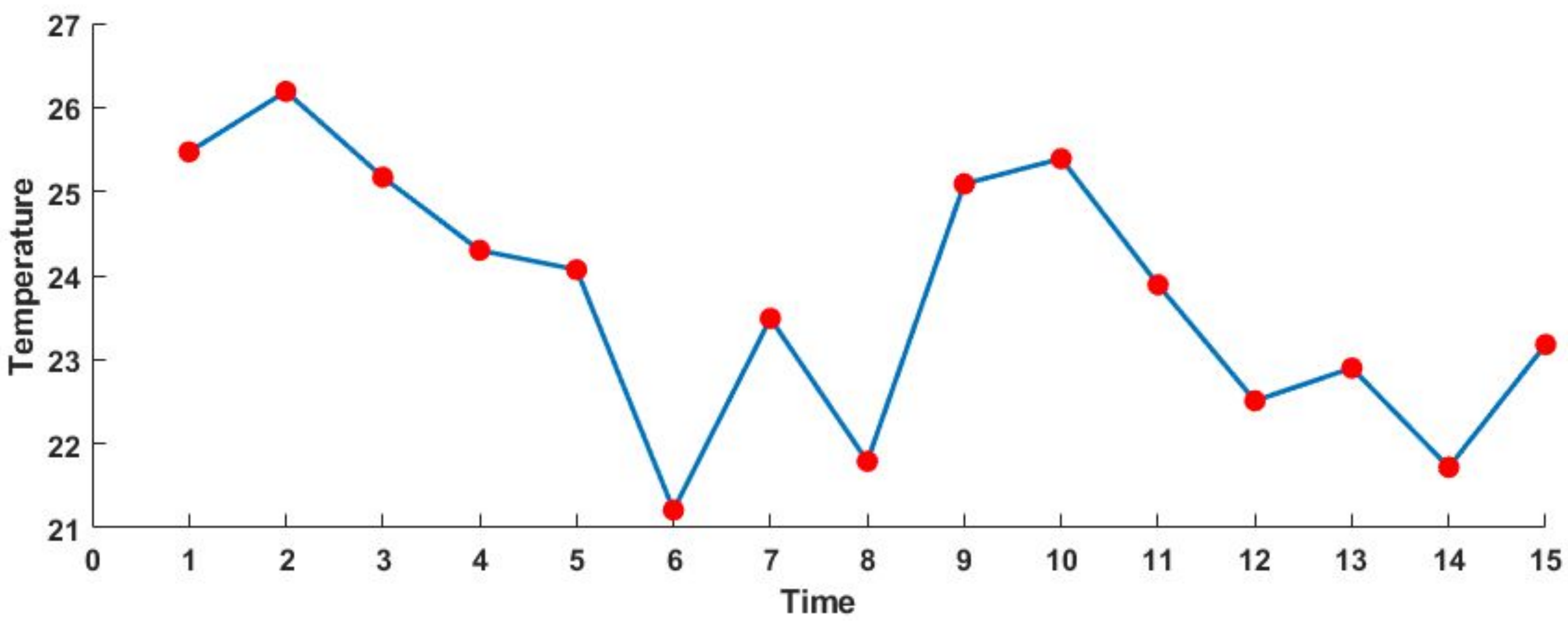
0.54	1.83	-2.26	0.86	0.32	-1.31	-0.43	0.34	3.58	2.77	-1.35	3.03	0.73	-0.06	0.71



Time Series Data and Dependence

- Example: Daily temperature at IIT Mandi
- Each element at time t (x_t) is dependent on the values elements at previous p time steps (x_{t-1} , x_{t-2}, …, x_{t-p}) – p time lag
  - *Temperature recorded for 15 days (1 Sept. 2019 – 15 Sept. 2019)*

25.47	26.19	25.17	24.3	24.07	21.21	23.49	21.79	25.09	25.39	23.89	22.51	22.9	21.72	23.18



Checking Dependency

- It's not always easy to just look at a time-series plot and say whether or not the series is independent
- x_t in a series is independent means that knowing previous values doesn't help you to predict the next value
  - Knowing x_{t-1} doesn't help to predict x_t
  - More generally, knowing x_{t-1}, x_{t-2}, ..., x_{t-p} doesn't help to predict x_t
    - p is the number of previous time step (time lag)
- Dependency of each element at time t (x_t) with the values of elements at previous p time steps (x_{t-1} , x_{t-2}, ..., x_{t-p}) is observed using autocorrelation

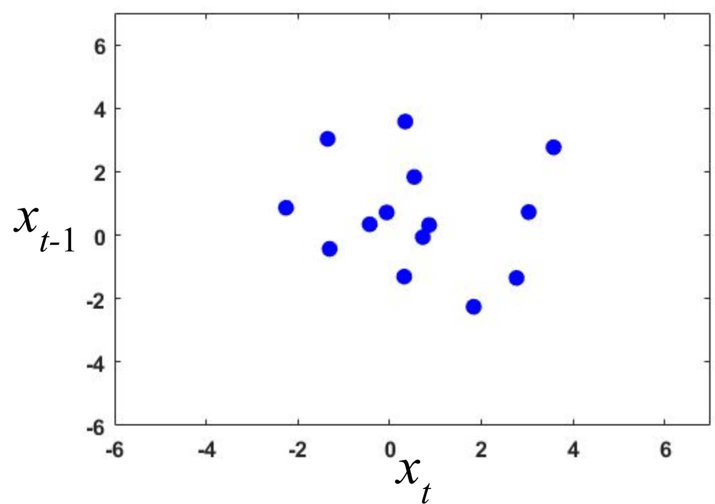
Checking Dependency - Autocorrelation

- The relationship between variables is called correlation
- Autocorrelation: The correlation calculated between the variable and itself at previous time steps
- Example: Data series in i.i.d
  - Autocorrelation between x_t and x_{t-p} – Pearson correlation coefficient between original series and lag-p series

Original Series x_t	0.54	1.83	-2.26	0.86	0.32	-1.31	-0.43	0.34	3.58	2.77	-1.35	3.03	0.73	-0.06	0.71
Lag-1 Series x_{t-1}		0.54	1.83	-2.26	0.86	0.32	-1.31	-0.43	0.34	3.58	2.77	-1.35	3.03	0.73	-0.06

  - Autocorrelation:

	x_t	x_{t-1}
x_t	1	**-0.1242**
x_{t-1}	**-0.1242**	1



Checking Dependency - Autocorrelation

- The relationship between variables is called correlation
- Autocorrelation: The correlation calculated between the variable and itself at previous time steps
- Example: Daily temperature at IIT Mandi
  - Autocorrelation between x_t (original series) and x_{t-1} (Lag-1 series)

Original Series

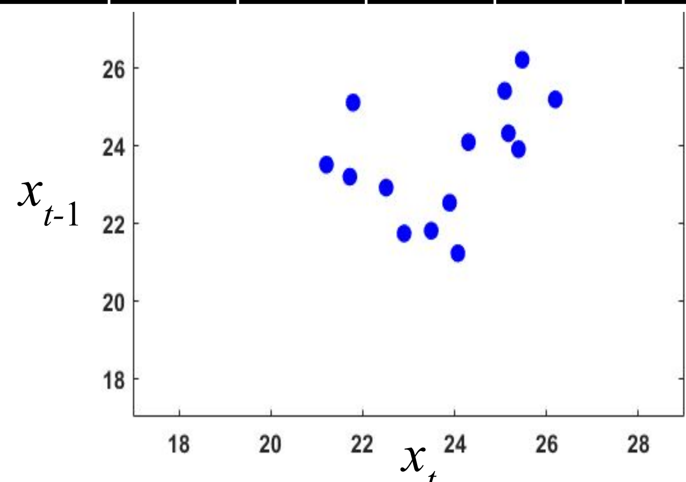
x_t	25.47	26.19	25.17	24.3	24.07	21.21	23.49	21.79	25.09	25.39	23.89	22.51	22.9	21.72	23.18

Lag-1 Series

x_{t-1}	25.47	26.19	25.17	24.3	24.07	21.21	23.49	21.79	25.09	25.39	23.89	22.51	22.9	21.72

  - Autocorrelation:

	x_t	x_{t-1}
x_t	1	**0.4054**
x_{t-1}	**0.4054**	1



Autoregression (AR)

Autoregression (AR)

- Regression on the values of same attribute
- Autoregression is a time series model that
  - uses observations from previous time steps as input to a linear regression equation to predict the value at the next time step
    - Output variable: value at next time step
    - Input variable: observations from previous time step
  - Output variable is a linear function of input variables

Autoregression (AR)

- Autoregression (AR): Regression on the values of same attribute
  - It is a time series model
  - Linear regression model that uses observations from previous p time steps as input to predict the value at the next time step
  - It makes an assumption that the observations at previous time steps are useful to predict the value at the next time step
  - The autocorrelation statistics help to choose which lag variables (p) will be useful in a model
  - Dependency of each element at time t (x_t) with the values of elements at previous p time steps (x_{t-1}, x_{t-2}, ..., x_{t-p}) is observed using autocorrelation
  - Autocorrelation: The correlation calculated between the variable and itself at previous time steps

Autoregression (AR) Model

- Autoregression (AR) is a linear regression model that uses observations from previous time steps as input to predict the value at the next time step
- An autoregression (AR) model makes an assumption that the observations at previous time steps are useful to predict the value at the next time step
- The autocorrelation statistics help to choose which lag variables (p) will be useful in a model
- Interestingly, if all lag variables (x_{t-1} , x_{t-2}, ..., x_{t-p}) show low or no correlation with the output variable (x_t), then it suggests that the time series problem may not be predictable
- This can be very useful when getting started on a new dataset

Autoregression (AR) Model

- Building an AR model depends on how many time lag (p) is considered

- AR(1) model: AR model using one time lag (p=1)
  - uses x_{t-1} i.e. value of previous time step to predict x_t

Illustration AR(1) Model – Prediction of Temperature

Date	Temp (x_{t-1})	Temp (x_t)	Date
		25.47	Sept 1
Sept 1	25.47	26.19	Sept 2
Sept 2	26.19	25.17	Sept 3
Sept 3	25.17	24.30	Sept 4
Sept 4	24.30	24.07	Sept 5
Sept 5	24.07	21.21	Sept 6
Sept 6	21.21	23.49	Sept 7
Sept 7	23.49	21.79	Sept 8
Sept 8	21.79	25.09	Sept 9
Sept 9	25.09	25.39	Sept 10
---	---	---	---
Oct 28	22.76	23.06	Oct 29
Oct 29	23.06	23.72	Oct 30
Oct 30	23.72	23.02	Oct 31

- T, the number of observations = 61

- Independent variable:
  - Temperature at the time t-1
- Dependent variable:
  - Temperature at the time t

AR(1) Model

- AR(1) model: AR model using one time lag (p=1)
  - uses x_{t-1} i.e. value of previous time step to predict x_t
- Given: Time series data: $\mathbf{X} = (x_1, x_2, \ldots, x_t, \ldots, x_T)$
  - x_t is the observation at time t
  - T be the number of observations
- AR(1) model is given as: $x_t = f(x_{t-1}, w_0, w_1) = \underline{w_0 + w_1\, x_{t-1}}$
  - The coefficients w_0 and w_1 are parameters of straight-line (regression coefficients) ***- Unknown***
- The regression coefficients are obtained as seen in simple linear regression (straight-line regression) using least square method

AR(1) Model - Training

- The regression coefficients are obtained as seen in simple linear regression (straight-line regression) using least square method
- Minimize the squared error between the actual data (x_t) at time t and the estimate of linear function (predicted variable ($\hat{x}_t$)) i.e. the function $f(x_{t\text{-}1}, w_0, w_1)$

$$\hat{x}_t = f(x_{t-1}, w_0, w_1) = w_0 + w_1\, x_{t-1}$$

$$\underset{w, w_0}{\text{minimize}}\, E(w_0, w_1) = \frac{1}{2}\sum_{t=2}^{T}\left(\hat{x}_t - x_t\right)^2$$

- The optimal $\hat{w}_0$ and $\hat{w}_1$ is given as

$$\hat{w}_1 = \frac{\sum_{t=1}^{T}\left(x_{t-1} - \mu_{t-1}\right)\left(x_t - \mu_t\right)}{\sum_{t=1}^{T}\left(x_{t-1} - \mu_{t-1}\right)^2}$$

$$\widehat{w}_0 = \mu_t - w_1\mu_{t\text{-}1}$$

- $\mu_{t\text{-}1}$: sample mean of variables at time t-1, $x_{t\text{-}1}$
- μ_t: sample mean of variables at time t, x_t

AR(1) Model: Testing

- For any test example at time t-1, $x_{t\text{-}1}$, the predicted value at time t, $\hat{x}_t$ is given by:

$$\hat{x}_t = f(x_{t-1}, w_0, w_1) = \hat{w}_0 + \hat{w}_1\, x_{t-1}$$

Evaluation Metrics for Time Series Prediction: Squared Error and Root Mean Squared Error

- The prediction accuracy is measured in terms of squared error: $E = (\hat{x}_t - x_t)^2$
  - x_t: actual value
  - $\hat{x}_t$: predicted value
- Let T_{test} be the total number of test samples
- The prediction accuracy of regression model is measured in terms of root mean squared error (RMSE):

$$E_{\text{RMS}} = \sqrt{\frac{1}{T_{test}} \sum_{t=1}^{T_{test}} (\hat{x}_t - x_t)^2}$$

- RMSE expressed in % as:

$$E_{\text{RMS}} = \sqrt{\frac{1}{T_{test}} \sum_{t=1}^{T_{test}} (\hat{x}_t - x_t)^2} \;*100$$

Evaluation Metrics for Time Series Prediction: Absolute Error and Mean Absolute Percentage Error (MAPE)

- Absolute error:

$$E_a = \frac{|x_t - \hat{x}_t|}{x_t}$$

  - x_t: actual value
  - $\hat{x}_t$: predicted value

- Let T_{test} be the total number of test samples
- The prediction accuracy of regression model is measured in terms of mean absolute percentage error:

$$E_{\mathrm{MAP}} = \left(\frac{1}{T_{test}} \sum_{t=1}^{T_{test}} \frac{|x_t - \hat{x}_t|}{x_t} \right) * 100$$

Illustration AR(1) Model – Prediction of Temperature: Training

Temp (x_{t-1})	Temp (x_t)	Date
	25.47	Sept 1
25.47	26.19	Sept 2
26.19	25.17	Sept 3
25.17	24.30	Sept 4
24.30	24.07	Sept 5
24.07	21.21	Sept 6
21.21	23.49	Sept 7
23.49	21.79	Sept 8
21.79	25.09	Sept 9
25.09	25.39	Sept 10
---	---	---
22.76	23.06	Oct 29
23.06	23.72	Oct 30
23.72	23.02	Oct 31

- T, the number of observations = 61

$$\hat{w}_1 = \frac{\sum_{t=1}^{60}(x_{t-1}-\mu_{t-1})(x_t-\mu_t)}{\sum_{t=1}^{60}(x_{t-1}-\mu_{t-1})^2}$$

$$\hat{w}_0 = \mu_t - w_1\mu_{t-1}$$

- μ_{t-1}: 22.81
- μ_t : 22.85
- $\hat{w}_1$: 0.523
- $\hat{w}_0$: 10.861

Illustration AR(1) Model – Prediction of Temperature: Test

- Predict Temperature for Nov 2

- $\hat{w}_1$: 0.523
- $\hat{w}_0$: 10.861

Nov 1

Temp (x_{t-1})	Temp (x_t)
22.30	-

- Predicted Temperature for Nov 2 : 22.52
- Actual Temperature on Nov 2 : 21.43
- Squared error : **1.19**
- Absolute error : **0.0509**

Autoregression Model

- AR(p) model: AR model using p time lags ($p < T$)
  - uses x_{t-1} , x_{t-2}, ..., x_{t-p} i.e. value of previous p time step to predict x_t

Illustration AR(p) Model – Prediction of Temperature

Temp (x_{t-3})	Temp (x_{t-2})	Temp (x_{t-1})	Temp (x_t)	Date
			25.47	Sept 1
		25.47	26.19	Sept 2
	25.47	26.19	25.17	Sept 3
25.47	26.19	25.17	24.30	Sept 4
26.19	25.17	24.30	24.07	Sept 5
25.17	24.30	24.07	21.21	Sept 6
24.30	24.07	21.21	23.49	Sept 7
24.07	21.21	23.49	21.79	Sept 8
21.21	23.49	21.79	25.09	Sept 9
---	---	---	---	---
22.83	23.98	24.47	22.76	Oct 28
23.98	24.47	22.76	23.06	Oct 29
24.47	22.76	23.06	23.72	Oct 30
22.76	23.06	23.72	23.02	Oct 31

- T, the number of observations = 61
- p = 3
- Independent variable:
  - Temperature at the time t-1 , t-2 and t-3
- Dependent variable:
  - Temperature at the time t

Autoregression Model

- AR(p) model: AR model using p time lags ($p < T$)
  - uses x_{t-1} , x_{t-2}, …, x_{t-p} i.e. value of previous p time step to predict x_t
- Given: Time series data: $\mathbf{X} = (x_1, x_2, \ldots, x_t, \ldots, x_T)$
  - x_t is the observation at time t
  - T be the number of observations
- AR(p) model is given as:

$$x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-p}, w_0, w_1, \ldots, w_p) = w_0 + w_1\, x_{t-1} + \ldots + w_p\, x_{t-p}$$

$$x_t = f(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{p} w_j x_{t-j} = \mathbf{w}^{\mathsf{T}}\mathbf{x}$$

$$\text{where } \mathbf{w} = [w_0, w_1, \ldots, w_p]^{\mathsf{T}} \text{ and } \mathbf{x} = [1, x_{t-1}, x_{t-2}, \ldots, x_{t-p}]^{\mathsf{T}}$$

  - The coefficients w_0, w_1, …, w_p are parameters of hyperplane (regression coefficients) ***- Unknown***

AR (p) Model - Training

- The regression coefficients are obtained as seen in multiple linear regression with p input variables using least square method
- Minimize the squared error between the actual data (x_t) at time t and the estimate of linear function (predicted variable ($\hat{x}_t$)) i.e. the function $f(\mathbf{x},\mathbf{w})$

$$\hat{x}_t = f(\mathbf{x},\mathbf{w}) = w_0 + \sum_{j=1}^{p} w_j x_{t-j} = w_0 + \mathbf{w}^{\mathsf{T}}\mathbf{x}$$

$$\underset{\mathbf{w}}{\text{minimize}}\, E(\mathbf{w}) = \frac{1}{2}\sum_{t=p+1}^{T}\left(\hat{x}_t - x_t\right)^2$$

- The autocorrelation statistics help to choose which lag variables (p) will be useful in a model

AR (p) Model - Training

- The optimal $\hat{\mathbf{w}}$ is given as $\boxed{\hat{\mathbf{w}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{x}^{(t)}}$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{t-p} & \dots & x_{t-3} & x_{t-2} & x_{t-1} \\ 1 & x_{(t+1)-p} & \dots & x_{t-2} & x_{t-1} & x_t \\ - & - & - & - & - & - \\ 1 & x_{(t+n)-p} & \dots & x_{t+n-3} & x_{t+n-2} & x_{t+n-1} \\ - & - & - & - & - & - \\ 1 & x_{T-p} & \dots & x_{T-3} & x_{T-2} & x_{T-1} \end{bmatrix} \qquad \mathbf{x}^{(t)} = \begin{bmatrix} x_t \\ x_{t+1} \\ - \\ x_{t+n} \\ - \\ x_T \end{bmatrix}$$

$\mathbf{X}$ is data matrix with time lag p

- The autocorrelation statistics help to choose which lag variables (p) will be useful in a model

AR (p) Model: Testing

- The value at time t, $\hat{x}_t$ is predicted by taking values from past p time steps (x_{t-1}, x_{t-2}, ..., x_{t-p}) as input:

$$\hat{x}_t = f(\mathbf{x}, \hat{\mathbf{w}}) = \hat{w}_0 + \sum_{j=1}^{p} \hat{w}_j x_{t-j} = \hat{\mathbf{w}}^{\mathsf{T}} \mathbf{x}$$

- The prediction accuracy is measured in terms of squared error:

$$E = \left(\hat{x}_t - x_t\right)^2$$

- Let T_{test} be the total number of test samples
- The prediction accuracy of regression model is measured in terms of root mean squared error:

$$E_{\mathrm{RMS}} = \sqrt{\frac{1}{T_{test}} \sum_{t=1}^{T_{test}} \left(\hat{x}_t - x_t\right)^2}$$

- Mean absolute percentage error (MAPE) is also used as a measure

Illustration AR(p) Model – Prediction of Temperature: Checking Dependency

Temp (x_{t-3})	Temp (x_{t-2})	Temp (x_{t-1})	Temp (x_t)	Date
			25.47	Sept 1
		25.47	26.19	Sept 2
	25.47	26.19	25.17	Sept 3
25.47	26.19	25.17	24.30	Sept 4
26.19	25.17	24.30	24.07	Sept 5
25.17	24.30	24.07	21.21	Sept 6
24.30	24.07	21.21	23.49	Sept 7
24.07	21.21	23.49	21.79	Sept 8
21.21	23.49	21.79	25.09	Sept 9
---	---	---	---	---
22.83	23.98	24.47	22.76	Oct 28
23.98	24.47	22.76	23.06	Oct 29
24.47	22.76	23.06	23.72	Oct 30
22.76	23.06	23.72	23.02	Oct 31

- $p = 3$
- T, the number of observations = 61
- Autocorrelation between x_t and x_{t-1} : 0.54
- Autocorrelation between x_t and x_{t-2} : 0.25
- Autocorrelation between x_t and x_{t-3} : -0.08
- An autocorrelation is deemed significant if

$$|\text{autocorrelation}| > \frac{2}{\sqrt{T}} = 0.25$$

- Time lag p=2 is sufficient as x_t is significant with x_{t-1} and x_{t-2}

Illustration AR(p) Model – Prediction of Temperature: Training

Temp (x_{t-2})	Temp (x_{t-1})	Temp (x_t)	Date
		25.47	Sept 1
	25.47	26.19	Sept 2
25.47	26.19	25.17	Sept 3
26.19	25.17	24.30	Sept 4
25.17	24.30	24.07	Sept 5
24.30	24.07	21.21	Sept 6
24.07	21.21	23.49	Sept 7
21.21	23.49	21.79	Sept 8
23.49	21.79	25.09	Sept 9
---	---	---	---
23.98	24.47	22.76	Oct 28
24.47	22.76	23.06	Oct 29
22.76	23.06	23.72	Oct 30
23.06	23.72	23.02	Oct 31

- p = 2
- T, the number of observations = 59
- Multiple linear regression with number of input variables = 2

$$\hat{\mathbf{w}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{x}^{(t)} \; ; \quad \hat{\mathbf{w}} \in \mathbf{R}^3$$

Illustration AR(p) Model – Prediction of Temperature: Test

- Predict Temperature for Nov 2

$$\hat{\mathbf{w}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{x}^{(t)} \; ; \quad \hat{\mathbf{w}} \in \mathbf{R}^3$$

Oct 31 Temp (x_{t-2})	Nov 1 Temp (x_{t-1})	Temp (x_t)
23.02	22.30	--

- **AR(2) model:**
- Predicted Temperature for Nov 2 : 22.49
- Actual Temperature on Nov 2 : 21.43
- Squared error : **1.13**
- Absolute error : **0.0495**
- **AR(1) model:**
- Predicted Temperature for Nov 2 : 22.52
- Actual Temperature on Nov 2 : 21.43
- Squared error : **1.19**
- Absolute error : **0.0509**

Summary: Autoregression

- Autoregression (AR): Regression on the values of same attribute
  - It is a time series model
  - Linear regression model that uses observations from previous p time steps as input to predict the value at the next time step
  - It makes an assumption that the observations at previous time steps are useful to predict the value at the next time step
  - The autocorrelation statistics help to choose which lag variables (p) will be useful in a model
- AR model can be performed on time series data with single variable or with multiple variables
- In this course we are limited only on the time series data with single variable

Text Books

1. J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Third Edition, Morgan Kaufmann Publishers, 2011.

2. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.