25.10.2024

Statistical Methods in AI (CS7.403)

Lecture-22: ML for Time Series

Ravi Kiran (ravi.kiran@iiit.ac.in)

https://ravika.github.io

@vikataravi

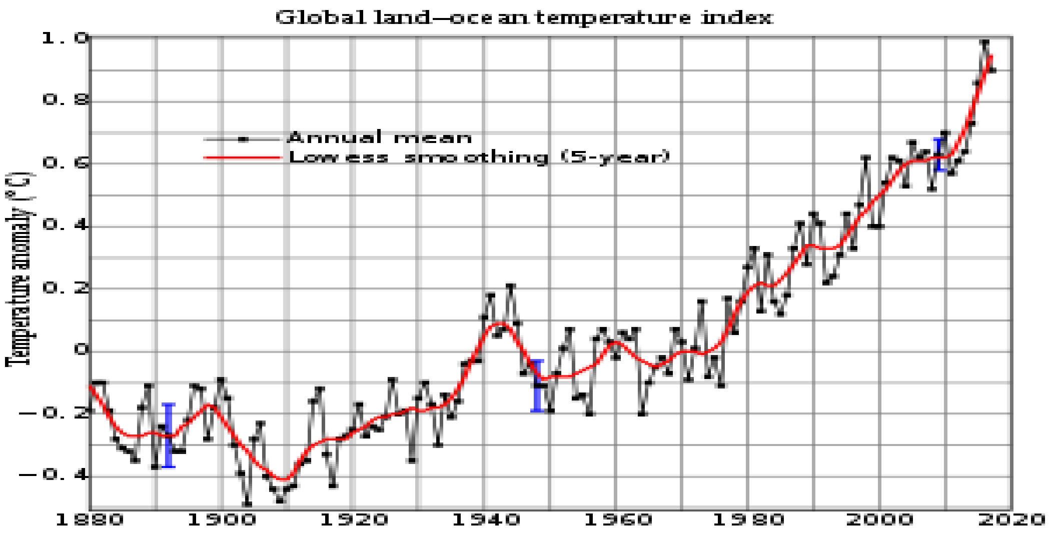
Center for Visual Information Technology (CVIT)

IIIT Hyderabad

Examples

BSE SENSEX

Global Land Ocean temperature



Examples

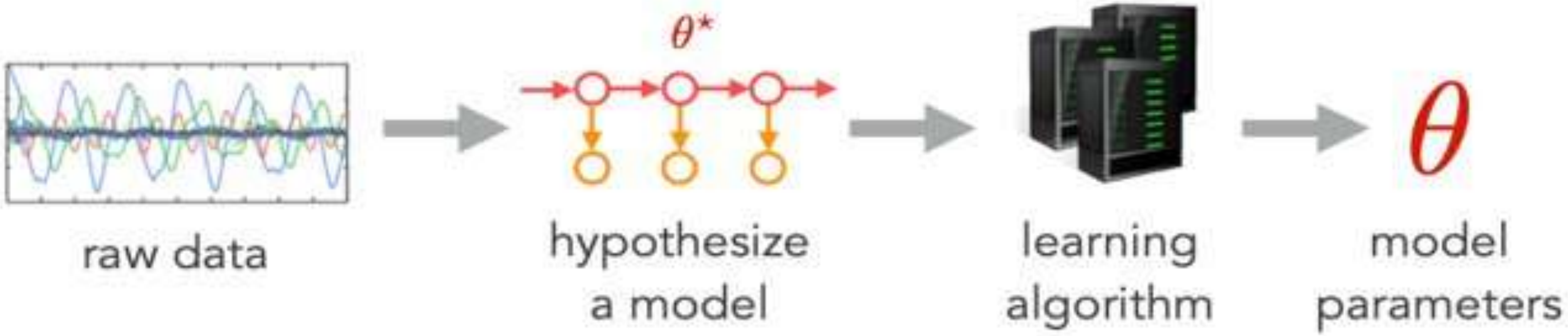
Day	No. of Packets of Milk sold
Monday	90
Tuesday	88
Wednesday	85
Thursday	75
Friday	72
Saturday	90
Sunday	102

Year	Population(in Million)
1921	251
1931	279
1941	319
1951	361
1961	439
1971	548
1981	685

Time Series

- Time series is a sequence of observations often ordered in time.
- Popular Problem: Given a sequence, predict future samples.
- Applications:
  - Meteorology,
  - Finance,
  - Marketing etc.

ML View Point



Using the model + learned parameters θ:

- Track
- Predict
- Simulate
- Plan
- ...

Learning Problem: find parameters θ s.t. $\theta \approx \theta^*$

From: B. Boots

Notation and Problem

- Notation: x[0], x[1], x[2], …, x[N].
- X[t], Where t is the time or index in the sequence.
- Assumption: Measurement at time t depends on three previous ones.
  - i.e., t-1, t-2 and t-3
- Why 3? We can have a different number.

Data

Raw Data	
Time	Sample
1	X_1
2	X_2
3	X_3
4	X_4
5	X_5
6	X_6
7	X_7

Rearranged Data			
Feature-1	Feature-2	Feature-3	Y_i
X_1	X_2	X_3	X_4
X_2	X_3	X_4	X_5
X_3	X_4	X_5	X_6
X_4	X_5	X_6	X_7

Feature Vector	
Feature	Y_i
V_1	X_4
V_2	X_5
V_3	X_6
V_4	X_7

A Simple Model

- $X[t] = w1\ X[t-1] + w2\ X[t-2] + w3\ X[t-3] + n$
  – Where n is noise.

- Problem:
  – Given the sequence X[0], X[1], ... X[N]
  – Find coefficients w1, w2, w3

- Find the coefficients w1,w2,w3 such that prediction error is minimal.

Performance Metrics For Time Series Data

- Four common techniques are:
  - mean absolute deviation,
  - mean absolute percent error,
  - the mean square error,
  - root mean square error.

$X_i : ACTUAL$

$\hat{X}_i : PREDICTED$

$$\text{MAD} = \sum_{i=1}^{n} \frac{\left|X_i - \hat{X}_i\right|}{n}$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left|X_i - \hat{X}_i\right|}{\hat{X}_i}$$

$$\text{MSE} = \sum_{i=1}^{n} \frac{\left(X_i - \hat{X}_i\right)^2}{n}$$

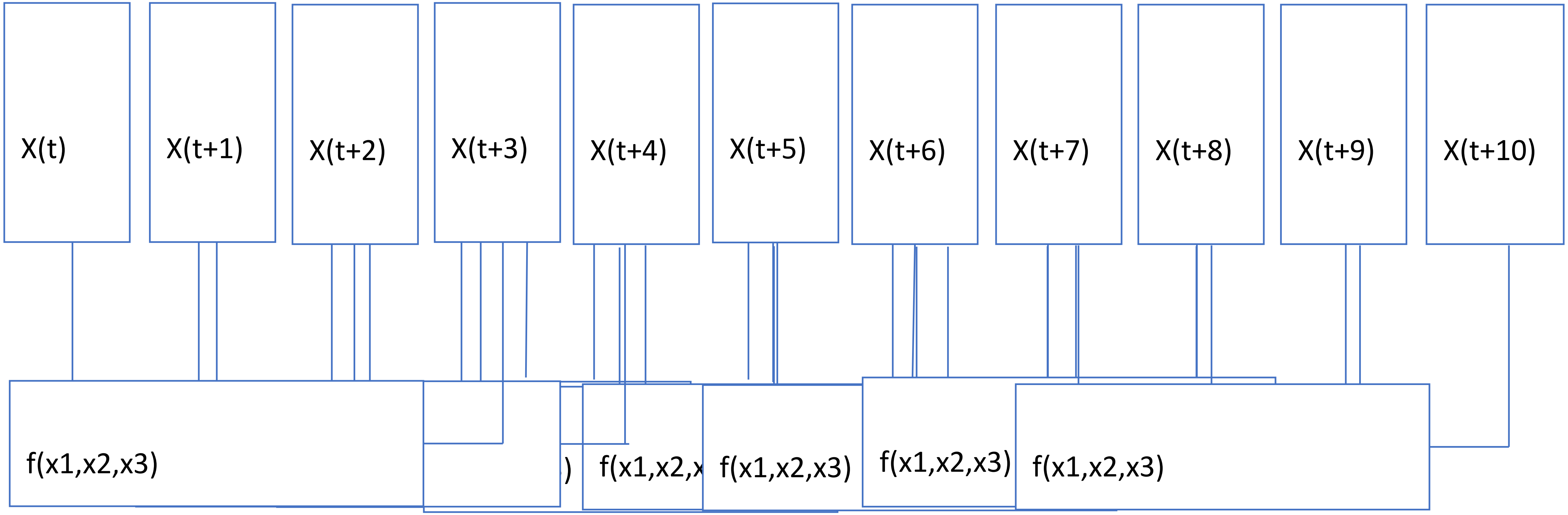
$$\text{RMSE} = \sqrt{\text{MSE}}$$

More Powerful Model

- Xt = f(W, Xt-1, Xt-2, Xt-3) + n
- Problem:
  - Given the sequence X0, X1, ... XN
  - Find coefficients W
- Data may be modeled as in the above linear case.
- f() may be seen as a MLP

$$\min W \sum_{t=3}^{N} (X_t - f(W, X_{t\text{-}1}, X_{t\text{-}2}, X_{t\text{-}3}))^2$$

Time series prediction



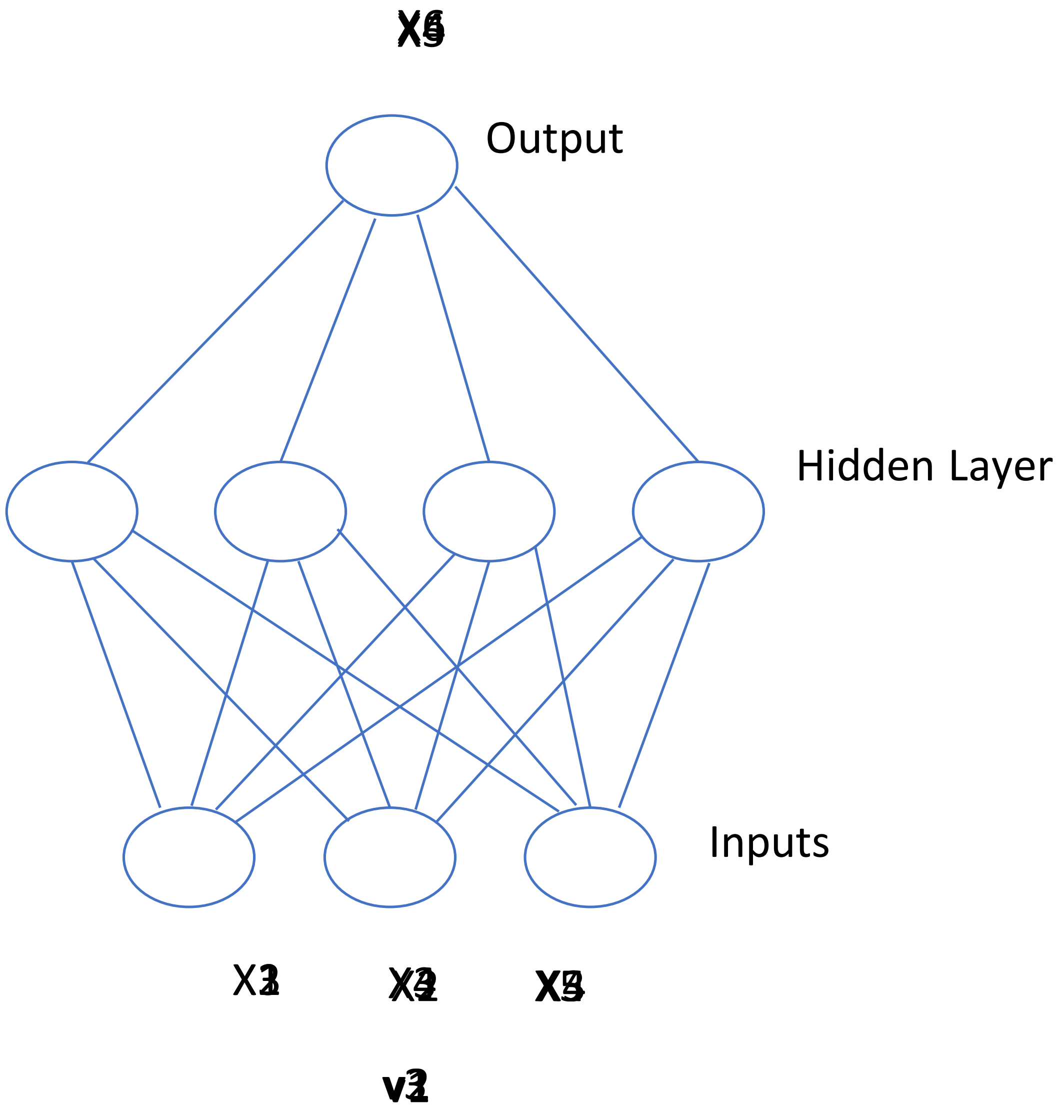
Data (Revisit)

Raw Data	
Time	Sample
1	X_1
2	X_2
3	X_3
4	X_4
5	X_5
6	X_6
7	X_7

Rearranged Data			
Feature-1	Feature-2	Feature-3	Y_i
X_1	X_2	X_3	X_4
X_2	X_3	X_4	X_5
X_3	X_4	X_5	X_6
X_4	X_5	X_6	X_7

Feature Vector	
Feature	Y_i
V_1	X_4
V_2	X_5
V_3	X_6
V_4	X_7

Neural Networks for Time Series Forecasting



Summary

- Predicting future samples is a new problem
- However, the solution is similar to what we know.
  - Cast as regression.
- Model can be linear
  - Linear regression
- Or nonlinear
  - MLP
- On how many past samples, the future sample will depend?
  - Order/model to be guessed?

AR and MA Modelling of Timeseries

Classical Models (AR and MA)

- Auto Regressive (**AR**) Model assumes $X_t = \alpha X_{t-1} + \epsilon_t$ (ϵ_t is random uncorrelated)

- **AR**: A model of order p is $X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \epsilon_t$

- Moving Average (**MA**) model assumes $X_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1}$

- **MA**: a model of order q is $X_t = \mu + \epsilon_t + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}$

Classical Models (ARMA & ARIMA)

- **ARMA (p,q):**

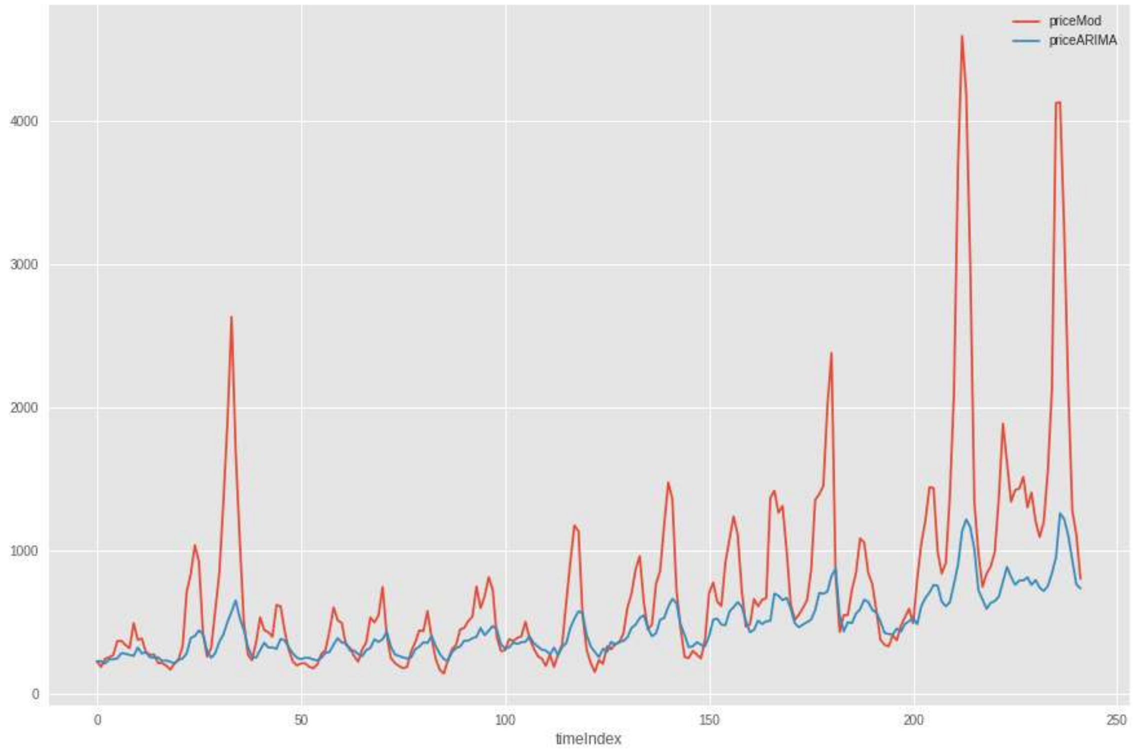
$$X_t = \mu + \epsilon_t + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}$$

With $\beta_0 = 1$

- **ARIMA (p, d, q):**
  - A process is ARIMA (p, q, d) if $\nabla^d X$ is ARMA (p,q).
  - Where $\nabla X_t = X_t - X_{t-1}$ and $\nabla^2 X_t = \nabla(\nabla X_t)$

Prediction using ARIMA and MLP

ARIMA



MLP