Understanding Time Series in Machine Learning

1 What is a Time Series?

A **time series** is a sequence of data points collected in chronological order, where the timing of observations is critical. Unlike regular datasets, time series data exhibits dependencies between observations, making it unique.

Why is it important?

- Enables forecasting of future values (e.g., sales, stock prices, weather).
- Identifies trends, seasonality, and anomalies.
- Models real-world time-dependent processes.

Examples of Applications:

- Stock price prediction
- Weather forecasting
- Energy consumption prediction
- Demand forecasting in retail

2 Components of a Time Series

Every time series can be decomposed into:

- **Trend**: Long-term increase or decrease (e.g., rising sales over years).
- **Seasonality**: Regular, repeating patterns (e.g., higher ice cream sales in summer).
- **Noise**: Random, unpredictable variations (e.g., sales dip due to a holiday).
- **Cyclic Behavior**: Longer, irregular patterns (e.g., economic cycles).

3 Basic Statistics in Time Series

Understanding the statistical properties is key:

- **Mean and Variance**: If constant over time, the series is **stationary**; if not, it's non-stationary.

- **Autocorrelation (ACF)**: Measures dependence on past values.
- **Partial Autocorrelation (PACF)**: Measures direct relationships with lagged values.

These are visualized using ACF and PACF plots to guide model parameter selection.

4 Why Stationarity Matters

Many models, like ARIMA, require stationary data (constant mean and variance). To achieve stationarity:

- **Differencing**: Subtract current value from previous value.
- **Log Transformation**: Stabilize variance.

5 ARIMA: Auto-Regressive Integrated Moving Average

ARIMA models predict future values using:

- **AutoRegression (AR)**: Past values.
- **Integrated (I)**: Differencing to achieve stationarity.
- **Moving Average (MA)**: Past forecast errors.

ARIMA is defined by parameters (p, d, q):

- p: Number of past observations.
- d: Number of differencings.
- q: Number of lagged forecast errors.

5.1 Demo: ARIMA(1,1,1) on Temperature Data

Consider the following daily temperature data:

Date	Temperature (°C)
2025-07-01	30
2025-07-02	32
2025-07-03	29
2025-07-04	31
2025-07-05	30

Table 1: Temperature Data

Step 1: Check Stationarity Calculate first differences:

$$\text{Diff}(t) = T(t) - T(t-1)$$

$$\text{Diff}(2025-07-02) = 32 - 30 = 2, \quad \text{Diff}(2025-07-03) = 29 - 32 = -3, \quad \text{etc.}$$

Differences: [2, -3, 2, -1]. The mean of differences is close to 0, suggesting stationarity after $d = 1$.

Step 2: Model with ARIMA(1,1,1) For simplicity, assume $p = 1$, $q = 1$. The model uses the previous differenced value and previous error to predict the next difference, then adds it to the last observation.

6 SARIMA: Seasonal ARIMA

SARIMA extends ARIMA to handle seasonality with additional parameters (P, D, Q, S):

- P: Seasonal AR order.
- D: Seasonal differencing.
- Q: Seasonal MA order.
- S: Seasonal period (e.g., 12 for monthly data).

6.1 Demo: SARIMA on Monthly Sales Data

Consider monthly sales with a yearly seasonal pattern:

Month	Sales (Units)
2024-01	100
2024-02	110
2024-03	120
2024-04	130
2024-05	140
2024-06	150
2024-07	200
2024-08	190
2024-09	180
2024-10	170
2024-11	160
2024-12	150

Table 2: Monthly Sales Data

Step 1: Seasonal Differencing For $S = 12$, compute seasonal differences:

$$\text{Diff}_{12}(t) = S(t) - S(t - 12)$$

Since we have one year, we apply first differencing ($D = 1$):

$$\text{Diff}(2024 - 02) = 110 - 100 = 10, \quad \text{etc.}$$

This removes the seasonal trend.

Step 2: Apply SARIMA(1,1,1)(1,1,1,12) The model combines ARIMA(1,1,1) with seasonal components to forecast future sales.

7 Step-by-Step Time Series Modeling

1. Visualize data to identify trends and seasonality.
2. Check stationarity using the Augmented Dickey-Fuller (ADF) test.
3. Make data stationary via differencing or transformations.
4. Use ACF/PACF plots to select model parameters.
5. Build and fit ARIMA/SARIMA model.
6. Validate on test data.
7. Forecast future values.

8 Applications and Analogy

Applications:

Domain	Application
Finance	Stock price forecasting, risk analysis
Healthcare	Patient vital signs monitoring
Retail	Sales and inventory forecasting
Transportation	Traffic flow prediction
Energy	Electricity demand forecasting

Table 3: Time Series Applications

Analogy: Forecasting tomorrow's weather:

- AR: Uses yesterday's weather.
- I: Accounts for recent changes.
- MA: Adjusts for past forecast errors.
- SARIMA: Adds seasonal patterns (e.g., summer vs. winter).