MIS4311
Machine Learning Applications

Fall 2025

Lecture #8

Time Series

Time series data is a sequence of data points collected or recorded at specific time intervals.

Examples include daily stock prices, monthly sales figures, yearly climate data, and many more.

The primary characteristic of time series data is its **temporal order**, meaning the sequence in which the data points are recorded matters.

Time Series

Unique Characteristics of Time Series Data

Time series data has several unique characteristics that distinguish it from other types of data:

1.Trend: This is the long-term movement or direction in the data. For example, the general increase in a company's sales over several years.
2.Seasonality: These are patterns that repeat at regular intervals, such as higher ice cream sales during the summer.
3.Cyclic Patterns: Unlike seasonality, cyclic patterns are not of a fixed period. These could be influenced by economic cycles or other factors.
4.Irregular Components: These are random or unpredictable variations in the data.

Time Series

Types of Time Series Data

Univariate vs. Multivariate Time Series:

- Univariate: A single variable or feature recorded over time (e.g., daily temperature).
- Multivariate: Multiple variables recorded over time (e.g., daily temperature, humidity, and wind speed).

Regular vs. Irregular Time Series:

- Regular: Data points are recorded at consistent time intervals (e.g., hourly, daily).
- Irregular: Data points are recorded at inconsistent time intervals.

Time Series

Preprocessing Time Series Data

Before diving into the analysis and forecasting, it's essential to preprocess your time series data to ensure accuracy and reliability. Handle missing values, outliers, and transform your data for better analysis.

1) Data Collection and Cleaning

Time series data often comes from various sources such as databases, APIs, or CSV files. The first step is to load your data into a suitable format, usually a Pandas DataFrame.

Time Series

2) Handling Missing Values and Outliers

Missing values and outliers can significantly affect your analysis. Here are common methods to handle them:

1. **Filling Missing Values:** Use methods like forward fill, backward fill, or interpolation.
2. **Removing Outliers**: Detect and remove outliers using statistical methods like Z-score or IQR (Interquartile Range).

3) Data Transformation

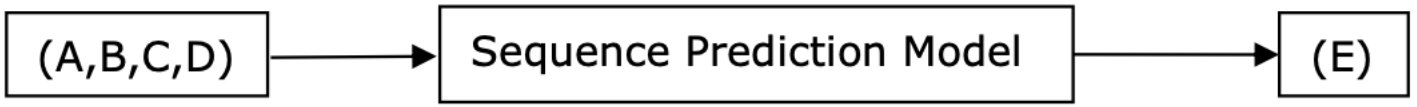
Transforming your data can help in identifying patterns and making it ready for analysis. Common transformations include:

1. **Smoothing:** Techniques like moving average can help in reducing noise.
2. **Differencing:** Used to remove trends and seasonality.
3. **Scaling and Normalization:** Ensure your data fits within a specific range for better model performance.

Time Series

Time series data means the data that is in a series of particular time intervals. If we want to build sequence **prediction in machine learning**, then we have to deal with sequential data and time.

Consider the following example to understand sequence prediction. Here A,B,C,D are the given values and you have to predict the value E using a **Sequence Prediction Model.**



Time Series

Sequence Prediction Models

Time series forecasting is the task of predicting future values based on historical data.
Examples across industries include forecasting of weather, sales numbers and stock prices.
It has also been applied to predicting price trends for cryptocurrencies such as Bitcoin and Ethereum.

One of the most commonly used is **Autoregressive Moving Average (ARMA),** which is a statistical model that predicts future values using past values. This method for making time series predictions is flawed, however, because it **doesn't capture seasonal trends**. It also assumes that the time series data is **stationary**, meaning that its statistical properties wouldn't change over time. This type of behavior is an idealized assumption that doesn't hold in practice, however, which means ARMA may provide skewed results.

Time Series

Sequence Prediction Models

An extension of ARMA is the **Autoregressive Integrated Moving Average (ARIMA)** model, which **doesn't assume stationarity** but does still assume that the data exhibits **little to no seasonality**.

Fortunately, the **seasonal ARIMA (SARIMA)** variant is a statistical model that can work with **non-stationary data and capture some seasonality.**

Additional popular time series forecasting packages are Prophet and DeepAR.

Prophet is an additive model developed by Facebook where non-linear trends are fit to seasonality effects such as daily, weekly, yearly and holiday trends.

DeepAR is a package developed by Amazon that enables time series forecasting with recurrent neural networks.

Time Series

Sequence Prediction Models

Python provides many easy-to-use libraries and tools for performing time series forecasting in Python.

Specifically, the stats library in Python has tools for building ARMA models, ARIMA models and SARIMA models with just a few lines of code.

Since all of these models are available in a single library, you can easily run many Python forecasting experiments using different models in the same script or notebook when conducting time series forecasting in Python.

Next Week

- Visualizing audio data and speech recognition

Thank you for your participation ☺