Berkeley Engineering | BerkeleyHaas

PROFESSIONAL CERTIFICATE IN MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Let’s give everyone a couple of minutes to join…

Module 10

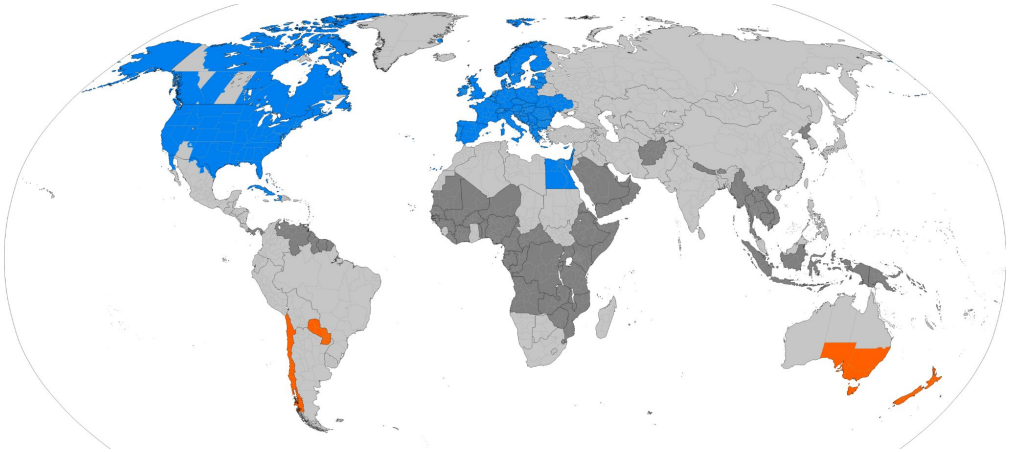
Time Series Analysis and Forecasting

Office Hours with Viviana Márquez
November 7, 2024

AGENDA

- Housekeeping
- Required activities for Module 10
- Content review Module 10: Time Series Analysis and Forecasting
- Questions

Daylight saving time!



Always check Canvas for the most up-to-date information regarding office hours!
Tool to convert to your timezone: https://www.worldtimebuddy.com/

Europe DST ends

North America DST ends

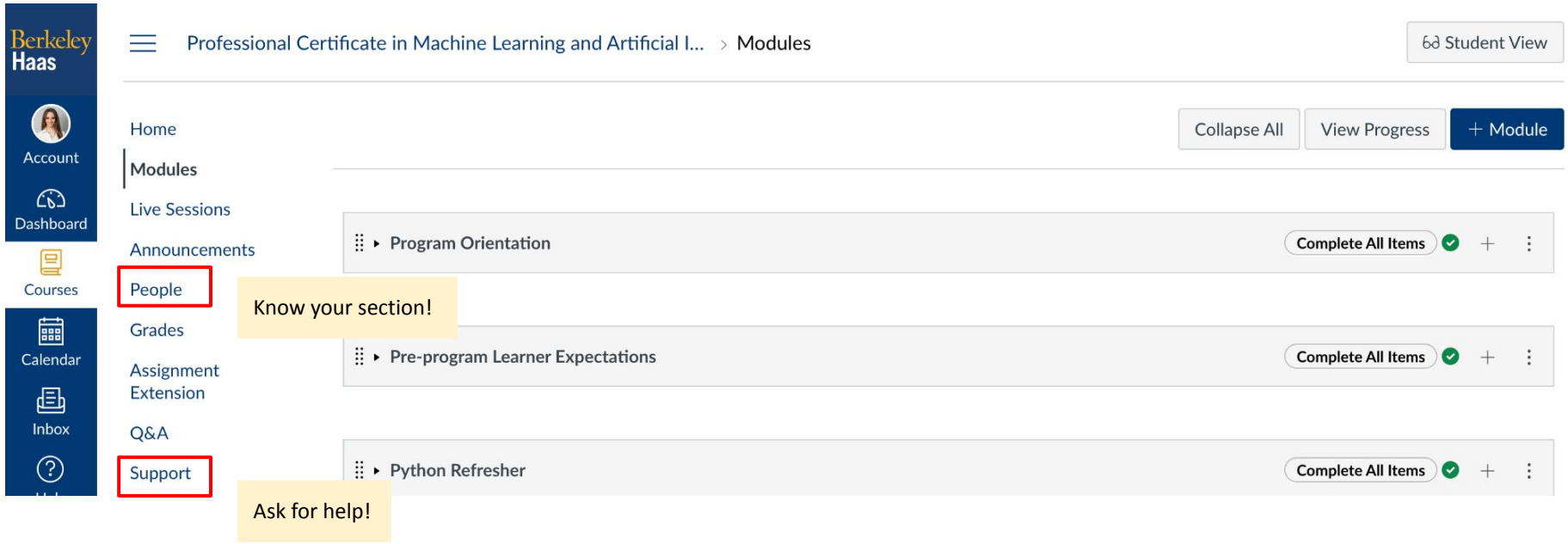
What's coming next?

- **Mod 10:** Time Series Analysis and Forecasting
- **Mod 11:** Practical Applications II
  - Similar to PA I, but this time you'll have to develop a regression project
  - 1:1s announced for Mod 12-Mod 15 (11/20-1/7)
- **Mod 12:** Classification and k-Nearest Neighbors
- **Thanksgiving break** 🦃🍂🍁 11/27 - 12/3
- **Mod 13:** Logistic Regression
- **Mod 14:** Decision Trees
- **Mod 15:** Gradient Descent and Optimization
- **Winter holiday break** ⛄❄️🌌 12/23-1/5

CAPSTONE PROJECT

- **How to have a great 1:1?**
  - Book as soon as the Calendly links are released!
  - Come prepared to the office hour
    - Have datasets, examples, ideas, questions
  - Only one 30 min consultation per learner in Mod12 - Mod15
    November 20, 2024 - January 7, 2024

CANVAS



- **Section A:** Vikesh Koul
- **Section B:** Jessica Cervi
- **Section C:** Viviana Márquez
- **Section D:** Francesca Vera
- **Section E:** Mani Kannappan

AGENDA

- ✅ Housekeeping
- Required activities for Module 10
- Content review Module 10: Time Series Analysis and Forecasting
- Questions

Required Activities for Module 10

- Knowledge Check 10.1: The Forecasting Problem
- Knowledge Check 10.2: Modeling and Autocorrelations
- Codio Assignment 10.1: Computing Autocorrelation
- Knowledge Check 10.3: Decomposition
- Codio Assignment 10.2: Decomposition
- Knowledge Check 10.4: ARMA
- Codio Assignment 10.3: ARMA Continued
- Codio Assignment 10.4: Time Series and Sales

AGENDA

- ✅ Housekeeping
- ✅ Required activities for Module 10
- Content review Module 10: Time Series Analysis and Forecasting
- Questions

Content review Module 10: Time Series Analysis and Forecasting

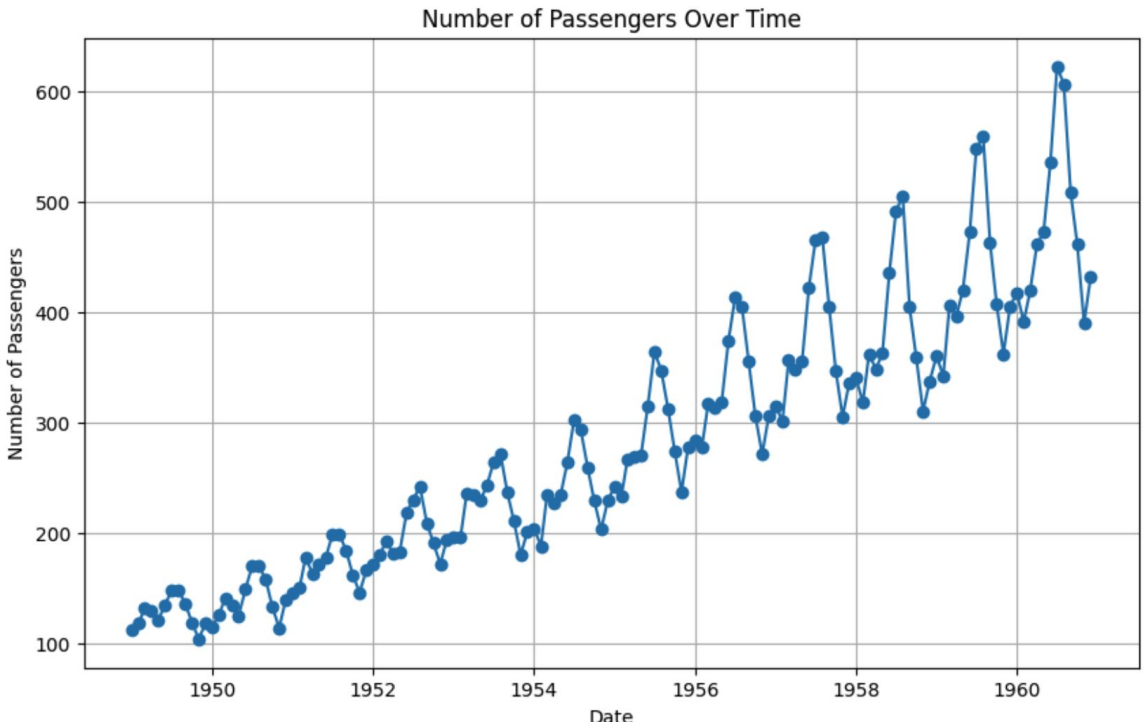
- What is a Time Series?
- Key Concepts
  - Serial Correlation
  - Trend
  - Seasonality
  - Stationarity
- Autocorrelation (ACF) and Partial Autocorrelation (PACF)
- Time Series models
- ARIMA example

What is a Time Series?

- A **time series** is data with temporal measurements. A collection of data points measured at specific regular intervals over time.
- Each observation has a timestamp, making the order of observations essential.
- Unlike regular datasets, where the order might not matter, the sequence of time in a time series dataset provides critical context.

Examples

	year	month	passengers
0	1949	Jan	112
1	1949	Feb	118
2	1949	Mar	132
3	1949	Apr	129
4	1949	May	121



Examples

- **Finance:** Stock prices, interest rates, cryptocurrency prices, and exchange rates recorded daily, hourly, or even by the minute.
- **Weather:** Temperature, rainfall, humidity, wind speed, and other climate data collected hourly or daily.
- **Retail:** Daily or weekly sales data, customer footfall in stores, inventory levels, and seasonal demand fluctuations.
- **Healthcare:** Patient heart rate, blood pressure, and other vital signs recorded over time, as well as disease incidence data.
- **Economics:** Monthly unemployment rates, GDP growth rates, inflation rates, and consumer price indices.
- **Energy:** Electricity usage by the hour, gas consumption, and renewable energy production data (e.g., solar or wind output).
- **Social Media:** Daily post engagement metrics, such as likes, shares, and comments, or sentiment analysis over time.
- **Traffic:** Hourly or daily traffic volume on highways, internet traffic data, and public transit ridership statistics.

Time Series Analysis

- A **time series analysis** is a technique used to model the relationship between the response variable ***y*** and time.
- The sequence of time in a time series dataset provides critical context, allowing us to identify seasonal patterns or observe how values evolve.
- The primary goal of time series analysis is **forecasting**—predicting future values based on historical data patterns. Other applications include:
  - **Anomaly detection:** Identifying unexpected events in time series data.
  - **Speech and audio processing:** speech recognition or music genre classification.
  - **Healthcare diagnostics:** Diagnosing arrhythmias from ECG data or seizures from EEG patterns

KEY CONCEPTS

- **Serial correlation (autocorrelation):**
  Observations closer in time tend to be more similar.

- **Trend:**
  The general (smoothed) behavior in the data.

- **Seasonality:**
  Regular and predictable fluctuations by periods of time.

- **Stationarity:**
  The statistical properties of the data (mean, variance, autocorrelation) remain constant over time, making it easier to model and analyze.

KEY CONCEPTS

- **Serial correlation (autocorrelation):**
  Observations closer in time tend to be more similar.

- **Trend:**
  The general (smoothed) behavior in the data.

- **Seasonality:**
  Regular and predictable fluctuations by periods of time.

- **Stationarity:**
  The statistical properties of the data (mean, variance, autocorrelation) remain constant over time, making it easier to model and analyze.

Code

- **Mod 10 - TSA Classical decomposition model.ipynb**
  https://colab.research.google.com/drive/16y96d5v1HTu1yNAsuXHvKXV5oDFC_bjR?usp=sharing

KEY CONCEPTS

- **Serial correlation (autocorrelation):**
  Observations closer in time tend to be more similar.

- **Trend:**
  The general (smoothed) behavior in the data.

- **Seasonality:**
  Regular and predictable fluctuations by periods of time.

- **Stationarity:**
  The statistical properties of the data (mean, variance, autocorrelation) remain constant over time, making it easier to model and analyze.

Autocorrelation

- **Autocorrelation**
  It's a statistical measure that indicates how a time series is correlated with its own past values. In simpler terms, it helps you understand whether and how the values of a time series at different times are related to each other.

$$\rho_k = \frac{Cov(X_t, X_{t-k})}{Var(X_t)}$$

- Interpretation
  - **Positive:** When the value at time ***t*** is similar to the value at time ***t-k***, meaning if one value is high, the previous value is likely also high (and vice versa for low values).
  - **Negative:** When the value at time ***t*** is dissimilar to the value at time ***t-k***, indicating an inverse relationship.
  - **Zero autocorrelation:** Indicates no relationship between the current value and its past values.

Lagged Values

- **Lagged values** refer to the past values of a time series at specified intervals (or "lags"). For instance, if you have a time series data point Xt, its lagged values can be defined as:
  - **Xt-1:** the value one time unit before t (one lag)
  - **Xt-2:** the value two time units before t (two lags)
  - **Xt-k:** the value k time units before t (k-th lags)

- In the context of autocorrelation, lagged values are used to measure how the current value of the time series is related to its past values. For example:
  - If you want to analyze how today's stock price (at time t) is influenced by its price yesterday (at time t-1), you would use the first lag.
  - To analyze how the price two days ago (at time t−2) affects today's price, you'd use the second lag.

- Understanding autocorrelation and lagged values is crucial in time series analysis as they help identify patterns (like seasonality), inform the selection of appropriate models for forecasting (e.g., ARIMA models), etc.

Autocorrelation and Partial Autocorrelation Functions

- **Autocorrelation Function (ACF):**
  - Measures the correlation of a time series with its own lagged values.
  - Helps to identify the degree of correlation between observations at different time lags.

- **Partial Autocorrelation Function (PACF):**
  - Measures the correlation of a time series with its own lagged values after removing the effects of intermediate lags.
  - Useful for identifying the order of the AR (AutoRegressive) part of ARIMA models.

Time Series Models

- Statistical methods:
  - **ARMA** (Autoregressive Moving Average)
  - **ARIMA** (Autoregressive Integrated Moving Average)
  - **SARIMA** (Seasonal ARIMA)
  - **VAR** (Vector Autoregression)
  - **VARMA** (Vector Autoregression Moving Average)
  - **ETS** (Exponential Smoothing State Space Model)
  - **Holt-Winters Seasonal Model**
- Decomposition models
  - **STL** (Seasonal Decomposition of Time Series)
- Neural Networks
  - **LSTM** (Long Short-Term Memory)
  - **GRU** (Gated Recurrent Units)
  - **Facebook Prophet**

ARIMA

- ARIMA is a popular statistical model used for forecasting time series data. It combines three key components:
  - **Autoregression (AR)**: Uses the relationship between an observation and a number of lagged observations (previous values) to make predictions.
  - **Integrated (I):** Involves differencing the data to make it stationary, which helps stabilize the mean of the time series by removing trends or seasonality.
  - **Moving Average (MA):** Models the relationship between an observation and a residual error from a moving average model applied to lagged observations.

- Parameters are denoted as (p, d, q):
  - **p:** Number of lag observations (AR part)
  - **d:** Number of times the raw observations are differenced (I part)
  - **q:** Size of the moving average window (MA part)

Code

- **LSTM | Prophet | SARIMA** example with the same dataset https://www.kaggle.com/code/danishmubashar/future-predict-by-lstm-prophet-sarima

AGENDA

- ✅ Housekeeping
- ✅ Required activities for Module 10
- ✅ Content review Module 10: Time Series Analysis and Forecasting
- Questions

QUESTIONS?

AGENDA

- ✅ Housekeeping
- ✅ Required activities for Module 10
- ✅ Content review Module 10: Time Series Analysis and Forecasting
- ✅ Questions

APPENDIX