

Week 1

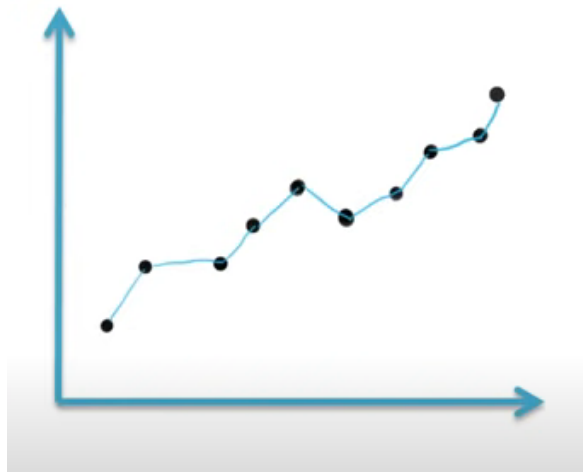
1. Tổng quan về dữ liệu chuỗi thời gian

- Dữ liệu chuỗi thời gian là những điểm dữ liệu theo thời gian và có khoảng cách đều nhau giữa các những quan sát liên tiếp (vd: giá chứng khoán trong ngày...)
- Phân tích dữ liệu chuỗi thời gian là các điểm dữ liệu được lấy theo thời gian có thể có các cấu trúc bên trong (vd: trend, seasonal variation...)

2. Thành phần của chuỗi thời gian

a. Xu hướng (trend):

- Là 1 mẫu dữ liệu cho thấy sự chuyển động của chuỗi các giá trị chuyển động cao hơn hoặc thấp hơn trong 1 thời gian dài.
- Thường xảy ra trong 1 khoảng thời gian rồi biến mất, không lặp lại nhưng có nhiều khả năng nó sẽ có xu hướng trở lại.



Hình 2.1: Xu hướng tích cực



Hình 2.2: Xu hướng tiêu cực

b. Seasonality (Tính thời vụ)

- Tính thời vụ là 1 đặc điểm của chuỗi thời gian, dữ liệu trải qua những thay đổi thường xuyên và có thể dự đoán, tái diễn theo hàng năm.
- Bất kỳ biến động hoặc mô hình có thể dự đoán được sự tái diễn hoặc lặp lại trong khoảng thời gian 1 năm

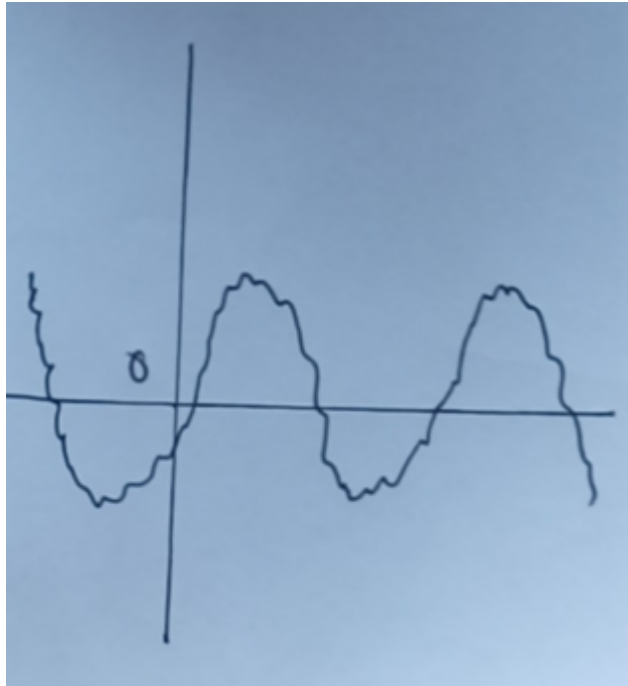
?? sự khác nhau giữa trend và seasonality

Trend	Seasonality
Đề cập đến sự thay đổi dài hạn của dữ liệu theo thời gian	Đề cập đến 1 mô hình lặp lại trong khoảng thời gian ngắn hơn (≤ 1 năm)
<ul style="list-style-type: none"> - Là sự thay đổi dần dần của dữ liệu. Có thể tăng hoặc giảm nhưng tương đối ổn định. - Một xu hướng có thể xuất phát từ nhiều yếu tố khác nhau (thay đổi về dân số, công nghệ, tình hình kinh tế....) 	<ul style="list-style-type: none"> - Được gây ra bởi nhiều yếu tố khác nhau (thời tiết, ngày lễ, các sự kiện diễn ra định kỳ...) - Thường thể hiện mô hình nhất quán về mức độ cao, thấp của dữ liệu

→ Phải phân biệt giữa xu hướng và tính thời vụ trong dữ liệu chuỗi thời gian vì chúng có tác động khác nhau đến việc phân tích và lập mô hình dữ liệu. Hiểu xu hướng giúp xác định hướng tổng thể của dữ liệu và dự đoán các giá trị trong tương lai, trong khi hiểu tính thời vụ giúp xác định các mô hình và tính đến tác động của các yếu tố mùa vụ trong dự báo.

c. White Noise

- Một chuỗi thời gian là nhiễu trắng nếu các biến độc lập và có phân bố giống hệt nhau với giá trị trung bình bằng 0
- Tất cả các biến có cùng phương sai và mỗi giá trị không có mối tương quan bằng 0 với tất cả các giá trị khác trong chuỗi
- Là 1 khái niệm quan trọng trong phân tích và dự báo chuỗi thời gian vì:
 - + Khả năng dự đoán: Nếu chuỗi thời gian là nhiễu trắng thì nó là ngẫu nhiên, không thể dự đoán
 - + Chẩn đoán mô hình: Một loạt lỗi từ mô hình dự báo chuỗi thời gian lý tưởng nhất phải là nhiễu trắng.
- Chuỗi thời gian là nhiễu trắng nếu:
 - + Giá trị trung bình bằng 0
 - + Độ lệch chuẩn hoặc độ biến động của chuỗi thời gian theo thời gian là không đổi
 - + Mối tương quan giữa độ trễ bằng 0



Hình 2.3. White noise

3. Phân rã chuỗi thời gian

- Phân rã chuỗi thời gian liên quan đến việc coi chuỗi là sự kết hợp của các thành phần cấp độ, xu hướng, tính thời vụ và độ nhiễu.
- Sự phân rã cung cấp một mô hình trừu tượng hữu ích để suy nghĩ về chuỗi thời gian nói chung và để hiểu rõ hơn các vấn đề trong quá trình phân tích và dự báo chuỗi thời gian
- Phân rã chuỗi thời gian thành các thành phần có hệ thống và không có hệ thống:
 - + **Có hệ thống**: Thành phần của chuỗi thời gian có tính nhất quán hoặc lặp lại và có thể được mô hình hóa: mức độ(level), xu hướng(trend) và tính thời vụ(seasonality)
 - + **Không có hệ thống** : Các thành phần của chuỗi thời gian không thể được mô hình hóa trực tiếp (noise)
- Kết hợp các thành phần của chuỗi thời gian:
 - + Mô hình cộng: $y(t) = \text{level} + \text{trend} + \text{seasonality} + \text{noise}$

Một mô hình cộng tính là tuyến tính trong đó những thay đổi theo thời gian được thực hiện một cách nhất quán với cùng một lượng.

+ Mô hình nhân: $y(t) = \text{level} * \text{trend} * \text{seasonality} * \text{noise}$

Mô hình nhân là mô hình phi tuyến, chẳng hạn như bậc hai hoặc hàm mũ. Sự thay đổi tăng hoặc giảm theo thời gian.

4. Mô hình chuỗi thời gian

- Mô hình chuỗi thời gian là 1 lĩnh vực quan trọng, được dùng để phân tích và dự đoán các dữ liệu thay đổi theo thời gian.
- Các khái niệm cơ bản:
 - + Chuỗi thời gian(time series): 1 tập hợp các quan sát dữ liệu được thực hiện theo thời gian
 - + Phân tích chuỗi thời gian(time series analysis): Là quá trình phân tích dữ liệu chuỗi thời gian để hiểu các cấu trúc và các yếu tố ảnh hưởng đến nó. Phương pháp: Biểu đồ, phân tích tổng quan, kiểm tra tính ngẫu nhiên
 - + Trạng thái thời gian (time series component): Một chuỗi thời gian thường được chia thành 3 phần chính:
 - Thành phần chuỗi (Trend component): sự thay đổi dài hạn hoặc xu hướng trong chuỗi thời gian
 - Thành phần mùa vụ (seasonal component): Các biến đổi chu kỳ có thời gian cố định (do yếu tố thời tiết hoặc các yếu tố trong kinh doanh)
 - Thành phần dao động ngẫu nhiên (residual component): Phần còn lại sau khi loại bỏ 2 phần trên, thể hiện sự biến động ngẫu nhiên của chuỗi thời gian
 - + Chuẩn bị dữ liệu (Data processing): Là quá trình xử lý dữ liệu chuỗi thời gian trước khi áp dụng mô hình (xử lý giá trị thiếu, chỉnh trạng thái thời gian và kiểm tra tính ổn định)
 - + Hàm tự tương quan (ACF): Là 1 độ đo thống kê sử dụng để đo lường mối quan hệ tương quan giữa các giá trị trong chuỗi thời gian và các giá trị trước đó

- + Hàm tự tương quan riêng (PACF): Tương tự ACF nhưng loại bỏ ảnh hưởng của các khoảng thời gian trung gian
- + Mô hình ARIMA: Kết hợp AR, MA, I để phân tích chuỗi thời gian
- + Mô hình học sâu: LSTM *
- + Dự đoán và đánh giá mô hình: Đưa ra dự đoán và đánh giá hiệu suất của mô hình (RMSE)
- + Chạy dự đoán : Dự đoán giá trị tương lai của chuỗi thời gian

5. Thao tác cơ bản của dữ liệu chuỗi thời gian với pandas và python

- Import thư viện:

```
import pandas as pd
import matplotlib.pyplot as plt
from datetime import datetime
```

- Đọc dữ liệu

```
data = pd.read_csv('https://raw.githubusercontent.com/haiyen0607/githubtest/main/AAPL.csv')
```

- Đặt lại ngày giờ

```
data_copy['Date'] = pd.to_datetime(data_copy['Date'])
```

- Lấy dữ liệu khoảng thời gian cụ thể:

```
data_new1 = (data_new.index >= '2022-12-31') & (data_new.index <= '2023-09-29')
data_2023 = data_new[data_new1]
data_2023
```

- Vẽ đồ thị:

```
import seaborn as sns
sns.pairplot(data_value)
```

END