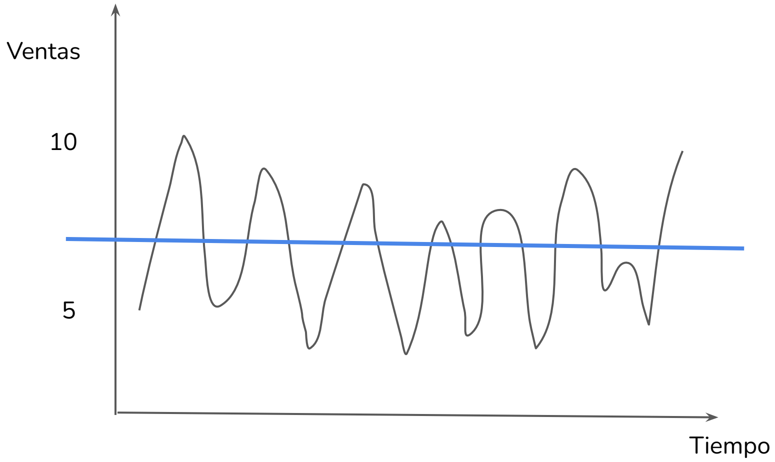
Un poco de series temporales



Hablando de Series Temporales

Mini Agenda

- Series de tiempo:
  - Definición
  - Componentes
  - Descomposición

- Algunos modelos básicos:
  - Media constante
  - Transformación logarítmica
  - Single Exponential Smoothing
  - Random walk
  - Tendencia lineal
  - Tendencia cuadrática

Un poco de Series Temporales

¿ Qué son las series temporales ?

Una serie de tiempo es un conjunto de observaciones tomadas en *intervalos regulares*, ordenados por el momento en que se produjeron.

Time series analysis (*análisis de series de tiempo*) comprende métodos para *proyectar la evolución*, obtener *estadísticas y otras características* de una variable a lo largo del tiempo.

Time series forecasting (*Pronóstico de series de tiempo*) es el uso de los modelos para predecir futuros valores.

Un poco de Series Temporales

Existen muchos ejemplos de datos que se pueden expresar como series de tiempo:

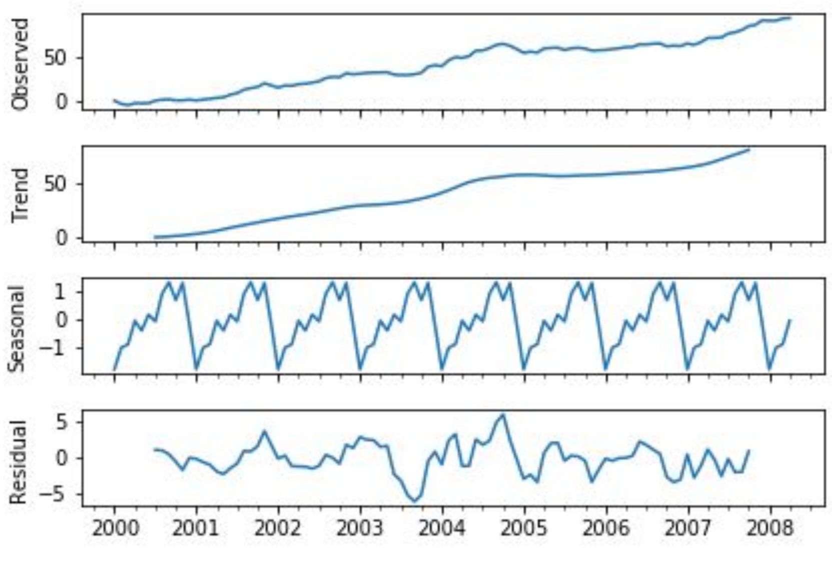
- Variables macroeconómicas (PBI, inflación, etc).
- Movimientos comerciales.
- Datos de producción.
- Consumo energético.
- Variables sociales (mortalidad infantil, pobreza...).

Componentes de las series temporales

- <u>Tendencia</u> (Trend): es el componente permanente, el efecto persistente en el tiempo.
- <u>Estacionalidad</u> (Seasonality): es un patrón estacional que se repite con regularidad.
- <u>Componente aleatoria</u> (Residual): son shocks que no presentan un efecto duradero. Se los define también como ruido o movimientos random.
- <u>Ciclos</u> (Cycle): otro tipo de dinámica no capturada por la tendencia o estacionalidad.

Descomposición de una serie de tiempo

El módulo statsmodels.api, con su método tsa.seasonal_decompose genera gráficos con: la serie de tiempo, la tendencia, la estacionalidad y el componente aleatorio.



Podemos pensar en descomponer una serie temporal en sus componentes. Básicamente, existen dos formas:

Descomposición aditiva:

Yt = S + C + e + T

Descomposición multiplicativa:

Yt = S*C*e*T

donde T es la tendencia, S es la estacionalidad, C es el ciclo y e es el error aleatorio

Descomposición de una serie de tiempo

<u>Descomposición aditiva:</u>

Yt = S + C + e + T

La descomposición aditiva es útil cuando la variación estacional se *mantiene relativamente constante* (lineal). Digamos que la altura de sus picos es constante.

<u>Descomposición multiplicativa:</u>

Yt = S*C*e*T

La descomposición multiplicativa es útil cuando la *tendencia crece* y la *amplitud de la variación estacional aumenta* (exponencial). O cuando la tendencia y la amplitud decrece.

Descomposición de una serie de tiempo

Estudiamos modelos aplicados a los componentes de las series para generar predicciones.

Para cada modelo:

- Lo *definimos*.
- Lo *ajustamos* a los datos de train.
- Lo *evaluamos* usando RMSE (podemos usar también MAPE o WAPE) sobre los datos de test.
- Lo *comparamos* con otros modelos.

Recordemos que el **error cuadrático medio raíz (RMSE)** compara los valores predichos con los valores observados.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

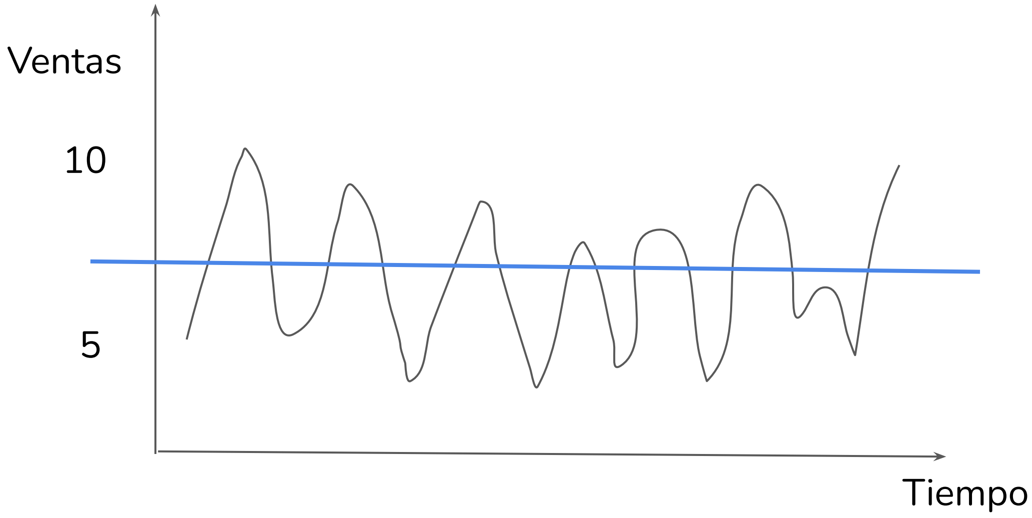
Alerta

Cuando trabajamos con series temporales, OJO!!!
Dividimos el dataset en train y test igual que siempre, pero al ser una serie temporal, tenemos que poner *shuffle=False* (que no mezcle las observaciones), creando una continuidad entre los sets de entrenamiento y de testeo.

Media constante

La <u>media constante</u> es el modelo más básico e ingenuo de todos.

Consiste simplemente en tomar *la media del dataset de train* y usarlo para predecir.



Transformación logarítmica

En algunos casos, la varianza de la serie aumenta con el paso del tiempo.

Hacer una transformación logarítmica de la serie puede ayudar a estabilizar la varianza. Para esto haremos lo siguiente:

- Generamos el modelo sobre el logaritmo de las ventas(variable y), tanto en train como en test.
- Generamos el modelo de regresión lineal usando el logaritmo de las ventas. Teniendo en cuenta train y test.
- Volvemos a transformar las predicciones a un valor de ventas, con una función exponencial.
- Este último valor será la predicción final, el cual evaluaremos utilizando cierta métrica.

Suavizado exponencial simple

En el caso de suavizamiento exponencial simple, se le da más peso a las observaciones recientes y menos peso a las antiguas. ¿ Cómo se hace esto ?

$$\hat{Y}_t = \alpha \, Y_{t-1} + (1 - \alpha) \, \hat{Y}_{t-1}$$

Donde $\hat{Y}_t$ es el valor predicho para el tiempo t, que es un promedio ponderado entre el valor previo predicho $\hat{Y}_{t-1}$ y el valor actual Y_{t-1}, α se conoce como smooth parameter y toma valores entre 0 y 1.

- Si α es igual a 1, todas las predicciones *son iguales al último valor observado*. Se le llama un **método ingenuo (naive)**.

Suavizado exponencial simple

La expresión generalizada de Single Exponential Smoothing:

$\hat{Y}_t = \alpha\, Y_{t-1} + \alpha\,(1-\alpha)\, Y_{t-2} + \alpha\,(1-\alpha)^2\, Y_{t-3} + \dots$

- Muestra que el peso de las observaciones <u>decrece en forma exponencial</u> a medida que son más antiguas.
- Y α nos indica *el ratio de caída*. Más cerca de uno, decae más rápido.

Random Walk

Decimos que un proceso es un random walk (que sigue una trayectoria aleatoria) si:

$Y_t = Y_{t-1} + \epsilon$

donde ϵ es un ruido blanco conocido como white noise.

- Este ruido conforma la parte de la serie que no se puede predecir a partir de ciertos enunciados. Piensen por ejemplo en las criptomonedas.
- Si el modelo de random walk sigue a la tendencia podemos agregar algo más y representarlo con una constante que normalmente tiene la letra ***d*** la que llamaremos drift o en inglés deriva.

$Yt = Yt\text{-}1 + \epsilon + d$

Random Walk

La serie de white noise et debe cumplir con lo siguiente:

- Tiene media igual a cero y varianza constante.
- Es completamente aleatoria.
- Los puntos tienen cero correlación entre ellos.

Tendencia Lineal

Podemos modelar la tendencia de la serie de diferentes maneras dependiendo de su comportamiento.

- La idea con este modelo es poder predecir utilizando el siguiente modelo lineal:

  $T_t = \beta_0 + \beta_1 \, \mathbf{TIEMPO}_t + \epsilon_t$

donde TIME es una variable dummy de tiempo (secuencia representando el tiempo)

Por lo que ajustamos una regresión lineal.

Tendencia Cuadrática

La idea con el modelo cuadrático es poder predecir utilizando el siguiente modelo cuadrático que solo es añadir la variable dummy de tiempo al cuadrado y ajustar :

$$T_t = \beta_0+\beta_1\ \mathbf{TIEMPO}_t + \epsilon_t + \beta_2\ (\mathbf{TIEMPO}_t\ )^2$$

Conclusiones

Una serie de tiempo tiene los componentes:

- Tendencia: el componente "permanente", el efecto que persiste en el largo plazo.
- Estacionalidad: los movimientos periódicos de la serie.
- Componente aleatorio: son cosas que no presentan un efecto duradero.
- Ciclos: se entiende por cualquier tipo de dinámica no capturada por la tendencia o la estacionalidad.

Tip: Para modelar la tendencia y la estacionalidad podemos usar *dummies de tiempo y estacionales*.

Existen varios *modelos para predecir* los nuevos valores de la serie: En la clase de hoy solo vimos unos pocos pero hay muchos más por lo que recomendamos realizar el curso de series temporales para poder entender más sobre la temática.

AHORA A PROGRAMAR :D

Bibliografía

- Elements of Forecasting (2001) - Libro Completo