



Aprendizaje Supervisado - Series de Tiempo

Clase sincrónica

Implementar modelos de aprendizaje automático por medio de técnicas estadísticas, adecuando los diferentes algoritmos debidamente a la situación y requerimientos necesarios

- Unidad 1: Introducción al Machine Learning
- Unidad 2: Aprendizaje Supervisado y No Supervisado
(Parte I: No supervisado)
(Parte II: Clasificación)
(Parte III: Clasificación)
(Parte IV: Regresión)
(Parte V: Series de tiempo)
- Unidad 3: Aplicando lo aprendido
(Parte I: Preprocesamiento de datos)
(Parte II: Modelamiento)



Te encuentras aquí



¿Qué aprenderás en esta sesión?

En esta sesión aprenderás sobre las series de tiempo, como se definen, como se componen las series de tiempo y que modelos son útiles para su aproximación.

¿Qué ventajas y
desventajas tiene la
regresión lineal?



¿Para qué se utiliza la
regularización?



¿Cómo se diferencian las
regularizaciones lasso,
ridge y elastic net?



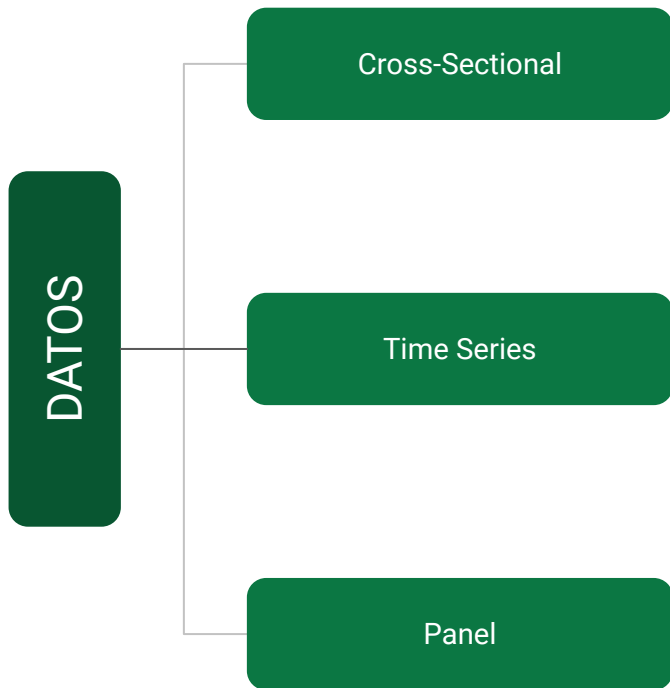
¿Qué métricas se utilizan
para evaluar las
regresiones lineales?



/* Categorías de datos*/

Categorías de datos

¿En qué consisten?



1. Datos de un periodo específico.
2. Múltiples entidades
3. Análisis Comparativo

1. Intervalos regulares de tiempo
2. Datos en orden cronológico
3. Datos con tendencias y estacionalidad.

1. Múltiples entidades en varios periodos de tiempo
2. Evolución temporales de las entidades
3. Alto volumen de generación de datos

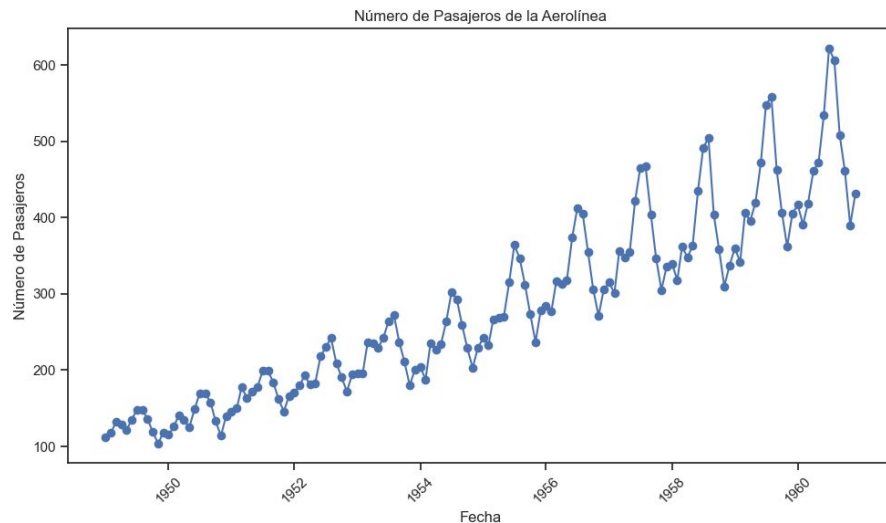
/* Series de tiempo*/

Series de tiempo

¿Qué son?

Una secuencia de puntos de datos organizados en el tiempo y ordenados:

- La secuencia captura datos a la misma puntos espaciados en el tiempo.
- Los datos recopilados irregularmente no son considerada una serie de tiempo, pero pueden trabajarse en intervalos regulares.



Series de tiempo

Aplicaciones

- **Economía y Finanzas:** Análisis de fluctuaciones en los precios de acciones, tasas de interés y datos económicos.
- **Meteorología:** Pronósticos climáticos y seguimiento de patrones climáticos a lo largo del tiempo.
- **Planificación Empresarial:** Predicción de ventas, demanda de productos y rendimiento de inventario.

¡Manos a la obra!

Tipos de datos y manejo con Python



Manos a la obra

Tipos de datos y manejo con Python

A continuación, veremos algunas características de los tipos de datos con Python, y algunas herramientas para manipular aquellos que están formateados específicamente como “tiempo”. Para ello, observa la presentación que te mostrará tu profesor, y replica los pasos en tu archivo de Jupyter Notebook si lo ves necesario. En la presentación aprenderemos:

- Categorías de datos
- Time series
 - Manejo de datos con python
 - Manejo de datos con pandas



/* Componentes de Series de Tiempo */

Componentes de las series de tiempo

Las series de tiempo tienen 3 componentes principales:

1. **Tendencia:** Dirección de movimiento a largo plazo
2. **Estacionalidad:** Comportamiento periodico.
3. **Residuos:** Fluctuaciones irregulares.

Componentes de las series de tiempo

Tendencia, estacionalidad y residuos

Tendencia

Captura la dirección general de la serie temporal

- Por ejemplo, aumentar el crecimiento del empleo año tras año a pesar de la estacionalidad.
- Puede ser creciente, decreciente o constante.
- Puede aumentar o disminuir de diferentes maneras (linealmente, exponencialmente o en otras maneras)

Estacionalidad

Captura los efectos que ocurren con una frecuencia específica

- Eventos que ocurren naturalmente, como las fluctuaciones climáticas causadas por la época del año.
- Comportamiento social o cultural, como días festivos o celebraciones religiosas.
- Fluctuaciones debidas a eventos del calendario, como el número de lunes por mes para el comercio o días festivos que cambian de un año a otro

Residuos

Fluctuaciones aleatorias que quedan después de que se remueve la tendencia y la estacionalidad.

- No se deberían encontrar patrones de tendencia o estacionalidad en los residuos.
- Representan fluctuaciones a corto plazo.
- Son aleatorios o una parte de los componentes de tendencia o estacionalidad se perdió en la descomposición.

¡Manos a la obra!

Componentes de Series de Tiempo



Manos a la obra

Componentes de series de tiempo

A continuación, veremos los componentes de las series de tiempo con Python, y cómo descomponerlas. Para ello, observa la presentación que te mostrará tu profesor, y replica los pasos en tu archivo de Jupyter Notebook si lo ves necesario.



/* Modelos de series de tiempo*/

Modelos de series de de tiempo

Forecasting

La forecasting o pronóstico de series de tiempo es una técnica utilizada en estadísticas y análisis de datos para predecir valores futuros en una secuencia de datos ordenados en función del tiempo. Estos datos pueden representar cualquier cosa que cambie con el tiempo, como ventas mensuales, temperaturas diarias, precios de acciones, tasas de interés, entre otros.

El objetivo principal de la forecasting de series de tiempo es identificar patrones, tendencias y estacionalidades en los datos históricos para hacer predicciones precisas sobre cómo se comportará la serie en el futuro

Modelos de series de tiempo

Función de autocorrelación (ACF)

La Función de Autocorrelación (ACF) muestra la correlación entre un valor en un punto de tiempo y los valores anteriores en intervalos de tiempo diferentes. Ayuda a identificar la cantidad de retrasos pasados que son relevantes para predecir el valor actual.

Matemáticamente se calcula como

$$ACF(k) = \frac{corr(X_t, X_{t-k})}{Var(X_t)}$$

Modelos de series de tiempo

Función de Autocorrelación Parcial (PACF)

La Función de Autocorrelación Parcial (PACF) muestra la correlación directa entre un valor en un punto de tiempo y los valores anteriores en un solo intervalo de tiempo. La PACF ayuda a identificar la influencia directa de los valores pasados sin la interferencia de otros valores intermedios.

$$PACF(k) = corr(X_t, X_{t-k})$$

Modelos de series de tiempo

Series estacionarias

Una serie de tiempo se considera estacionaria cuando sus propiedades estadísticas no cambian con el tiempo, es decir, exhibe ciertas características constantes a lo largo del tiempo.

Esto es fundamental en el análisis de series temporales porque muchas de las técnicas estadísticas y modelos de predicción asumen que la serie es estacionaria o que puede ser transformada en una serie estacionaria antes de aplicar el análisis.

Modelos de series de tiempo

Series estacionarias

Para que una serie de tiempo se considere estacionaria, debe cumplir con tres propiedades clave:

- **Media constante:** la media de la serie debe ser aproximadamente constante en el tiempo. Puede haber fluctuaciones alrededor de la media, pero estas fluctuaciones deben ser aleatorias y no seguir una tendencia clara.
- **Varianza constante:** la dispersión de los valores en la serie no cambia a lo largo del tiempo.
- **Autocorrelación constante:** la estructura de dependencia temporal en la serie no cambia con el tiempo.

Modelos de series de tiempo

Series estacionarias, ACF y PACF

	Estacionaria	No estacionaria
ACF	generalmente mostrará una rápida caída después de un pequeño número de rezagos (lags) y fluctuará alrededor de cero, sin patrones sistemáticos.	es común que la ACF muestre una decaída lenta y gradual, o incluso puede no mostrar un patrón claro de decaimiento. Pueden aparecer correlaciones a varios rezagos.
PACF	generalmente se reducirá a cero después de un número finito de rezagos, lo que significa que solo los valores en los primeros rezagos son significativos	puede mostrar patrones más complejos y no necesariamente caer rápidamente a cero. Puede haber correlaciones significativas incluso en rezagos más lejanos.

Modelos de series de tiempo

Modelos Autorregresivos (AR)

Un modelo AR es una regresión lineal del valor actual de la serie contra uno o más valores previos de la serie. Utiliza estimadores de máxima verosimilitud para determinar coeficientes en lugar de mínimos cuadrados. Formalmente, un modelo AR de orden p se define como:

$$y_t = c + \Phi_1 \cdot y_{t-1} + \Phi_2 \cdot y_{t-2} + \dots + \Phi_p \cdot y_{t-p} + \epsilon_t$$

c es una constante

ϕ son los coeficientes de autorregresión

y_t corresponde al valor de la serie en el tiempo t , y_{t-i} son valores anteriores

ϵ corresponde al error

Modelos de series de tiempo

Modelos de medias móviles

Los modelos de medias móviles (MA, Moving Average) son especialmente útiles para capturar patrones temporales y fluctuaciones aleatorias en los datos. En ellos, la variable dependiente se relaciona con los errores pasados de predicción y los términos de error actuales. Un modelo MA de orden "q" se define como:

$$y_t = c + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q}$$

c es una constante

ϵ_t es el término de error en el período t

θ_i son los coeficientes de autorregresión

ϵ_i son los valores anteriores en la serie de tiempo

Modelos de series de tiempo

Modelos de medias móviles

- Ajustar las estimaciones MA es más complicado que los modelos AR porque los términos de error no son observables.
- Como resultado, es necesario utilizar procedimientos de ajuste iterativos no lineales.
- Los modelos MA son menos interpretables que los modelos AR.

Modelos de series de tiempo

Test de Dickey Fuller

El test de Dickey-Fuller es una prueba estadística utilizada en el análisis de series temporales para determinar si una serie de datos es estacionaria o no.

Muchas técnicas estadísticas asumen que los datos siguen siendo constantes a lo largo del tiempo, pero las series no estacionarias muestran tendencias, patrones cíclicos o cambios en la varianza lo que puede dificultar la aplicación de modelos y análisis estadísticos adecuados.

El test de Dickey-Fuller se basa en la idea de que si una serie de tiempo es no estacionaria, su valor presente depende en gran medida de sus valores pasados. El test busca determinar si existe una raíz unitaria en la serie, es decir, que la serie muestra una tendencia o una estructura que causa cambios sostenidos en la media a medida que avanza en el tiempo. Si el test rechaza la hipótesis de una raíz unitaria, se considera que la serie es estacionaria.

Modelos de series de tiempo

Test de Dickey Fuller

El procedimiento básico del test de Dickey-Fuller implica los siguientes pasos:

- Se plantea una hipótesis nula (H_0) que asume la presencia de una raíz unitaria en la serie, lo que implica no estacionariedad.
- Se estima un modelo de regresión que incluye la serie de tiempo original y sus valores rezagados, para ver si los coeficientes de los valores rezagados son estadísticamente significativos.
- Se realiza una prueba estadística, como el estadístico t de Dickey-Fuller, que compara los coeficientes estimados con los valores críticos para determinar si se rechaza la hipótesis nula.
- Si el estadístico de prueba es menor que el valor crítico correspondiente, se rechaza la hipótesis nula y se concluye que la serie es estacionaria. En caso contrario, no se puede rechazar la hipótesis nula, lo que sugiere que la serie es no estacionaria.

Modelos de series de tiempo

Diferenciación

La diferenciación es un proceso para transformar una serie temporal no estacionaria en una serie temporal estacionaria. Implica calcular las diferencias entre los valores sucesivos de la serie temporal, para eliminar o reducir las tendencias y patrones no estacionarios en los datos. A menudo se realizan una o más diferenciaciones hasta que se obtenga una serie estacionaria.

Modelos de series de tiempo

Diferenciación - Pasos

- Se calcula la diferencia entre cada valor de la serie y el valor inmediatamente anterior. Esto puede ayudar a eliminar una tendencia lineal en los datos.
- En caso de que la primera diferencia no logre estacionarizar la serie, se repite el proceso de diferencia, calculando la diferencia entre cada valor de la primera diferencia y su valor inmediatamente anterior.
- Si la serie todavía no se vuelve estacionaria después de la segunda diferencia, puedes continuar con terceras, cuartas diferencias, y así sucesivamente, hasta que la serie se vuelva estacionaria.

Modelos de series de tiempo

ARIMA

ARIMA significa "AutoRegressive Integrated Moving Average." Combina los modelos que hemos visto (AR y MA) considerando además el componente de integración, es decir, los pasos que se deben seguir para que la serie sea estacionaria

El modelo ARIMA se denota como $ARIMA(p, d, q)$, donde:

- "p" es el orden del componente AR (AutoRegressive).
- "d" es el orden del componente de integración (Integrated).
- "q" es el orden del componente MA (Moving Average).

Series de tiempo

Métricas de evaluación

Métrica	Fórmula
Error Cuadrático Medio (MSE)	$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Raíz del Error Cuadrático Medio (RMSE)	$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$
Desviación Absoluta Media (MAE)	$\text{MAE} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Error Absoluto Porcentual Medio (MAPE)	$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100\%$
Coefficiente de determinación (R2)	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

¡Manos a la obra!

Modelos de Series de Tiempo

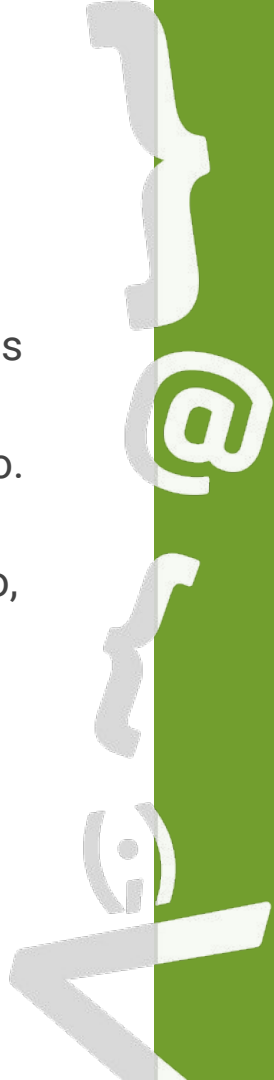


Manos a la obra

Modelos de Series de tiempo

A continuación, veremos cómo aplicar las diferentes herramientas asociadas a las series de tiempo en Python. Para ello observa la presentación que te mostrará tu profesor, y replica los pasos en tu archivo de Jupyter Notebook si lo ves necesario.

En la presentación aprenderemos a implementar los modelos de series de tiempo, compararlos y evaluarlos con enfoque de Machine Learning



Desafío - Prediciendo las ventas de una tienda



Desafío

"Prediciendo las ventas de una tienda"

- Descarga el archivo "Prediciendo las ventas de una tienda".
- Tiempo de desarrollo asincrónico: desde 4 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪



Ideas fuerza



Las **Serie de Tiempo** son datos secuenciales y ordenados en el tiempo.



Las principales componentes de la serie de tiempo son **tendencia, estacionalidad y residuos.**



Algunos modelos útiles para el trabajo de series de tiempos son **AR, MA y ARIMA.**

¿Qué conceptos no te
quedaron claros o quieres
reforzar?



Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio
- Desafío “Prediciendo las ventas de una tienda”





Próxima sesión...

Aplicar herramientas y metodologías diversas y adecuadas para implementar modelos que resuelvan problemas de negocios.

{desafío}
latam_

*Academia de
talentos digitales*

