

IBV
100

escuela
británica de
artes creativas
y tecnología

Buenas prácticas



Clase 01

Buenas prácticas

A continuación te compartimos algunas buenas prácticas sobre sobre Modelos de Regresión Lineal y Series de Tiempo:

1. Entender la naturaleza de las series de tiempo: Antes de aplicar cualquier modelo estadístico o de machine learning, es crucial comprender los componentes de las series de tiempo como la tendencia, estacionalidad y ciclo. Esto permite seleccionar el método de análisis más adecuado.
2. Uso de métodos de descomposición: Aplicar métodos de descomposición clásica para identificar y separar los componentes de una serie de tiempo. Esto es útil para entender mejor los datos y para mejorar la precisión de los pronósticos.
3. Verificar la estacionariedad: Utilizar pruebas como la de Dickey-Fuller para determinar si una serie de tiempo es estacionaria o sigue un proceso de caminata aleatoria. Los modelos de series de tiempo asumen estacionariedad, por lo que es un paso previo esencial.

Clase 01

Buenas prácticas

4. Análisis exploratorio de datos (EDA): Realizar un análisis exploratorio, incluyendo gráficos y funciones de autocorrelación, para identificar patrones y correlaciones en los datos.

5. Comparación y correlación entre series: Analizar la relación entre diferentes series de tiempo, como los precios de acciones de diferentes empresas, para encontrar correlaciones significativas que puedan ser de interés para la toma de decisiones en el mercado.

6. Preparación de datos para entrenamiento y prueba: Dividir el conjunto de datos en entrenamiento y prueba para evaluar la precisión de los modelos de pronóstico y evitar el sobreajuste.

7. Documentación y presentación de resultados: Preparar una documentación adecuada de los análisis realizados, incluyendo Jupyter Notebooks y archivos PDF con capturas de pantalla y comentarios pertinentes, para facilitar la comprensión y replicabilidad de los estudios.

Clase 01

Buenas prácticas

Ejemplo de utilización en el mercado laboral:

En el ámbito financiero, los analistas de datos pueden utilizar series de tiempo para pronosticar precios de acciones y tomar decisiones de inversión. Por ejemplo, podrían aplicar modelos de promedios móviles para predecir el precio de las acciones de IBM y Walmart, como se sugiere en la tarea asignada en la clase.

Código ejecutable de ejemplo (suponiendo que ya se ha instalado el paquete `yfinance` y se han importado las librerías necesarias):

```
python
import yfinance as yf
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.stattools import adfuller

# Descargar datos históricos de acciones de Apple
apple_data = yf.download('AAPL', start='2020-01-01',
end='2021-01-01')
```

Clase 01

Buenas prácticas

```
# Visualizar los precios de cierre  
apple_data['Close'].plot(title='Apple Stock Prices')  
plt.show()
```

```
# Realizar prueba de Dickey-Fuller para comprobar  
estacionariedad  
result = adfuller(apple_data['Close'])  
print(f'ADF Statistic: {result[0]}')  
print(f'p-value: {result[1]}')
```

```
# Si el p-value es menor que 0.05, rechazamos la hipótesis  
nula (la serie es estacionaria)  
'''
```

En este código, se descargan los datos históricos de las acciones de Apple, se visualizan los precios de cierre y se realiza la prueba de Dickey-Fuller para verificar si la serie de tiempo es estacionaria. Cada línea de código está comentada para explicar su función.

Clase 02

Buenas prácticas

A continuación te compartimos algunas buenas prácticas sobre la Nomenclatura y componentes de las series de tiempo:

1. Utilizar la correlación para identificar relaciones entre variables financieras: Es importante calcular la correlación entre activos financieros para entender su relación y cómo se mueven juntos en el mercado. Esto es útil para la diversificación de carteras y la gestión de riesgos.

Ejemplo de mercado laboral: En la gestión de carteras de inversión, los analistas utilizan la correlación para construir carteras que minimicen el riesgo a través de la diversificación, seleccionando activos que no estén perfectamente correlacionados.

2. Visualización de datos para interpretar relaciones: El uso de herramientas de visualización como Matplotlib para crear diagramas de dispersión ayuda a interpretar visualmente la relación entre dos series de datos.

Ejemplo de código:

```
python
import matplotlib.pyplot as plt
```

Clase 02

Buenas prácticas

```
# Suponiendo que 'returns' es un DataFrame con las columnas
'AAPL' y 'MSFT'
plt.scatter(returns['AAPL'], returns['MSFT'])
plt.title('Diagrama de dispersión entre AAPL y MSFT')
plt.xlabel('Rendimientos AAPL')
plt.ylabel('Rendimientos MSFT')
plt.show()
```

Cada línea de este código importa la biblioteca Matplotlib, crea un diagrama de dispersión con los rendimientos de AAPL y MSFT, y luego muestra el gráfico con títulos y etiquetas adecuados.

3. Modelado estadístico para análisis de regresión: La librería `statsmodels.api` es una herramienta poderosa para realizar regresiones y otros análisis estadísticos. Añadir una columna de unos al DataFrame permite manejar el término constante en la regresión.

Ejemplo de mercado laboral: Los analistas financieros utilizan la regresión para estimar la relación entre el rendimiento de un activo y factores de riesgo del mercado, como el rendimiento del mercado en general o factores económicos específicos.

Clase 02

Buenas prácticas

4. Análisis de autocorrelación para series temporales: El cálculo de la autocorrelación y la función de autocorrelación (ACF) es crucial para entender la dependencia temporal en series financieras y para la construcción de modelos predictivos como ARIMA.

Ejemplo de mercado laboral: Los economistas y analistas de mercado utilizan la autocorrelación para identificar patrones en series temporales económicas y financieras, lo que puede ser útil para pronosticar tendencias futuras.

5. Pruebas de estacionariedad como Dickey-Fuller: Realizar pruebas de estacionariedad permite determinar si una serie temporal es estacionaria o sigue un proceso de caminata aleatoria, lo cual es fundamental para la modelización de series temporales.

Ejemplo de mercado laboral: Los traders cuantitativos aplican pruebas de estacionariedad para validar si las estrategias de trading basadas en patrones históricos son viables o si los precios siguen un proceso aleatorio.

Clase 02

Buenas prácticas

6. Descomposición de series temporales: La descomposición en componentes cíclicos, estacionales y de tendencia permite un análisis más detallado y la identificación de patrones subyacentes en los datos.

Ejemplo de mercado laboral: En el análisis de ventas, la descomposición de series temporales ayuda a identificar patrones estacionales y tendencias a largo plazo, lo que es útil para la planificación de inventario y la estrategia de marketing.

Estas prácticas reflejan la importancia de la estadística y el análisis de datos en el campo financiero y cómo los Científicos de Datos aplican estas técnicas para obtener insights y tomar decisiones basadas en datos.

Clase 03

Buenas prácticas

A continuación te compartimos algunas buenas prácticas sobre la Nomenclatura y componentes de las series de tiempo:

1. Visualización detallada: Antes de realizar cualquier análisis, es crucial observar la serie de tiempo en detalle para identificar patrones, tendencias y anomalías. En el mercado laboral, esto puede ayudar a detectar periodos de recesión o crecimiento en indicadores económicos, ventas o métricas de rendimiento.
2. Descomposición de series de tiempo: Utilizar técnicas de descomposición para entender los componentes de una serie de tiempo (tendencia, estacionalidad y residuos) es una práctica recomendada. Por ejemplo, en el análisis financiero, esto permite a los analistas distinguir entre movimientos de mercado normales y eventos extraordinarios.
3. División en conjuntos de entrenamiento y prueba: Al predecir series de tiempo, es importante dividir los datos en conjuntos de entrenamiento y prueba para validar la precisión de las predicciones. En el ámbito laboral, esto se utiliza para evitar el sobreajuste y garantizar que los modelos sean generalizables a datos no vistos.

Clase 03

Buenas prácticas

4. Uso de promedios móviles para predicciones: Los promedios móviles son una herramienta simple pero poderosa para realizar predicciones en series de tiempo. En el mercado laboral, se utilizan para pronosticar demanda de productos, precios de acciones, entre otros.

5. Evaluación de la precisión de las predicciones: Medir el rendimiento de las predicciones con métricas como RMSE y MAPE es esencial para entender la calidad del modelo predictivo. En la práctica profesional, estas métricas ayudan a comparar diferentes modelos y elegir el más adecuado para la toma de decisiones.

6. Entrega de resultados reproducibles: Se recomienda entregar resultados en formatos que permitan la reproducibilidad del análisis, como Jupyter Notebooks, y documentar adecuadamente el proceso en informes o archivos PDF. En el mercado laboral, esto facilita la colaboración y revisión de los análisis realizados.

Clase 03

Buenas prácticas

Ejemplo de código ejecutable para calcular la media móvil de 30 días y graficarla junto con los datos originales:

```
python
import pandas as pd
import matplotlib.pyplot as plt

# Cargar datos de precios de acciones (reemplazar
'ruta_al_archivo.csv' con la ruta real)
data = pd.read_csv('ruta_al_archivo.csv', index_col='Date',
parse_dates=True)

# Calcular la media móvil de 30 días
data['30_day_MA'] = data['Close'].rolling(window=30).mean()

# Graficar los precios de cierre y la media móvil
plt.figure(figsize=(10,5))
plt.plot(data['Close'], label='Precio de Cierre')
plt.plot(data['30_day_MA'], label='Media Móvil 30 días',
color='orange')
plt.title('Precio de Cierre y Media Móvil de 30 días de Apple')
plt.legend()
plt.show()
```

Clase 03

Buenas prácticas

Explicación del código:

- Se importan las librerías ``pandas`` para manejo de datos y ``matplotlib.pyplot`` para visualización.
- Se carga el conjunto de datos de precios de acciones desde un archivo CSV, utilizando la fecha como índice.
- Se calcula la media móvil de 30 días utilizando el método ``rolling`` y ``mean`` sobre la columna 'Close'.
- Se crea una figura y se grafican tanto los precios de cierre como la media móvil de 30 días.
- Se añade un título y una leyenda al gráfico.

Este código es un ejemplo de cómo los Científicos de Datos pueden visualizar y analizar series de tiempo en el contexto de los precios de las acciones, una aplicación común en el mercado financiero.

¡Mucho éxito en tus estudios!