Nombre alumno: Esteban Javier Berumen Nieto

Correo alumno: esteban.berumen@iteso.mx

Instituto tecnológico de Estudios Superiores de Occidente

Repaso para primer examen de la clase Series de Tiempo

Docente: Daniel Nuño, danielnuno@iteso.mx

Otoño 2024

Fecha de entrega: miércoles 18 de septiembre, 2024

El repaso está diseñado en base a los conocimientos teóricos adquiridos en la clase y tomando como referencia el libro Forecasting: Principles and Practice (3rd ed), capítulos 1 a 7.

La entrega es en digital en un archivo de Word. Las preguntas y respuestas deben diferenciarse claramente.

En tus palabras, o subrayando si se presentan opciones, responde las siguientes preguntas:

1. Los componentes de una serie de tiempo descompuesta son:
   1. Tendencia-ciclo, estacionalidad y residuos.
   2. Media y varianza cero.
   3. Autocorrelación y normalidad.
2. Las ventajas de usar la descomposición Seasonal and Trend decomposition using Loess (STL) son:
   1. STL manejará cualquier tipo de estacionalidad, no solo datos mensuales y trimestrales.
   2. El componente estacional puede cambiar con el tiempo y se puede controlar la tasa de cambio.
   3. Se puede controlar la suavidad del ciclo de tendencia.
   4. Es un modelo robusto ante valores atípicos.
   5. Puede manejar fechas faltantes (la serie no contempla fines de semana).
   6. Todas las anteriores.
3. ¿Qué mide la autocorrelación?
   1. Mide la relación lineal entre una variable dependiente y una independiente.
   2. Mide la relación lineal entre valores rezagados de una serie de tiempo.
   3. Mide la tendencia de la serie de tiempo.
   4. Mide la normalidad de la serie de tiempo.
4. ¿Para qué se utilizan la función de autocorrelación?

Ayuda a identificar la dependencia entre rezagos de la serie de tiempo nos puede ayudar a detectar patrones repetitivos.

1. En las pruebas de hipótesis, ¿qué es el valor-p?

El valor-p es la probabilidad de obtener resultados al menos tan extremos como los observados en los datos, asumiendo que la hipótesis nula es verdadera. Un valor-p bajo indica evidencia fuerte en contra de la hipótesis nula.

1. ¿Para qué sirve la prueba Ljung–Box? ¿qué significa aceptar la hipótesis nula y cuando aceptas la hipótesis nula?

Se utiliza para evaluar si hay autocorrelación significativa en los residuos de un modelo de series de tiempo. Aceptar la hipótesis nula significa que no hay autocorrelación significativa en los residuos.

1. ¿Para qué sirve la prueba Shapiro-Wilk? ¿qué significa aceptar la hipótesis nula y cuando aceptas la hipótesis nula?

Evalúa si una muestra sigue una distribución normal. Aceptar la hipótesis nula significa que los datos parecen seguir una distribución normal.

1. ¿Para qué sirve la prueba Breusch–Pagan o White? ¿qué significa aceptar la hipótesis nula y cuando aceptas la hipótesis nula?

Estas pruebas se utilizan para detectar heterocedasticidad, es decir, si los residuos de un modelo tienen varianza constante. Aceptar la hipótesis nula significa que los residuos son homocedásticos (tienen varianza constante).

1. ¿Qué es heterocedasticidad y homocedasticidad?

Heterocedasticidad: Se refiere a que los residuos del modelo presentan una varianza constante

Homocedasticidad: Nos indica lo contrario una varianza no constante en los residuales

1. ¿En qué casos cree usted que sea útil realizar un pronóstico a partir de una descomposición?

Cuando se desea entender cómo se comportan individualmente los ciclos de estacionalidad o las tendencias a largo plazo, y prever su comportamiento futuro. También es útil cuando la estacionalidad cambia con el tiempo, o cuando se quiere eliminar el ruido de los residuos para hacer pronósticos más precisos.

1. En general cuando se desean producir pronósticos, ¿cuándo cree usted que debería trabajar sobre datos desestacionalizados y cuándo con la serie original? ¿Qué diferencias prácticas o en la interpretación de los resultados pudiera llegar a observar al hacerlo de una u otra forma?

Se deben usar datos desestacionalizados cuando se busca eliminar el componente estacional y enfocarse en la tendencia y los patrones subyacentes. Esto es útil cuando la estacionalidad es regular y se conoce bien.

Por otro lado, se usa la serie original cuando la estacionalidad es relevante para los pronósticos, como en ventas estacionales o series con ciclos conocidos. La diferencia práctica es que en los datos desestacionalizados se facilita la identificación de tendencias subyacentes y su interpretación es más directa al aislar la estacionalidad.

1. ¿Cuáles son otras formas de modificar tus datos para mejorar tu ajuste y pronóstico?
   1. Transformación matemática
   2. Trabajar sobre las diferencias y retornos
   3. Modificar observaciones atípicast
   4. Ajustes de incremento poblacional o inflación
   5. Recortar tus datos
2. ¿Por qué harías una transformación matemática?

para estabilizar la varianza, hacer que los datos sean más lineales y reducir el efecto de valores extremos o la heterocedasticidad. Por ejemplo, una transformación logarítmica puede ser útil cuando los datos tienen una tendencia exponencial.

1. Cómo identificas los datos atípicos?

Podemos hacerlo gráficamente con un box plot o de manera mas precisa viendo que datos se encuentran fuera de los limites inferior o superior, que seria Q1 – 1.5 X IQR o Q3 +1.5 X IQR

1. Brevemente describa intuitivamente los modelos base y cuando los usaría: Media, Ingenuo, Ingenuo Estacional, Deriva.

Media: Pronostica el valor medio de los datos históricos. Se usa cuando no hay una tendencia o estacionalidad clara.

Ingenuo: Usa el último valor observado como pronóstico. Es útil en series sin tendencia ni estacionalidad, o como un modelo base.

Ingenuo estacional: Usa el último valor del mismo periodo del año anterior como pronóstico. Se utiliza cuando hay estacionalidad clara y fuerte.

Deriva: Pronostica un cambio lineal en función de la tendencia observada. Es útil cuando hay una tendencia lineal constante a lo largo del tiempo.

1. Brevemente enliste posibles ventajas/desventajas de las métricas de error RMSE, MAE, MAPE, MASE.

RMSE:

Ventaja: Penaliza más los grandes errores, lo que lo hace útil para detectar grandes desviaciones.

Desventaja: Sensible a valores atípicos.

MAE:

Ventaja: Más fácil de interpretar porque es el promedio de los errores absolutos.

Desventaja: No penaliza tanto los errores grandes como el RMSE.

MAPE:

Ventaja: Proporciona una medida relativa, fácil de interpretar en términos porcentuales.

Desventaja: Puede ser problemático con valores cercanos a cero, lo que genera errores extremadamente altos

MASE:

Ventaja: Comparativa entre modelos porque está escalada por la escala del error de un modelo base.

Desventaja: Es menos intuitiva de interpretar que el MAE.

1. Un buen método de pronóstico producirá errores con las siguientes propiedades:
   1. Media es diferente a cero, están correlacionadas, varianza no es continua.
   2. Están correlacionados, la media es cero, la varianza no es constante a través del tiempo.
   3. No están correlacionados, media es cero, varianza continua a través del tiempo
2. ¿Cuál es la afectación en un modelo o sus pronósticos, cuando los errores no se distribuyen de manera normal?

Si los errores no son normales, puede ser más difícil confiar en los intervalos de confianza y las pruebas de hipótesis basadas en la distribución normal. Esto podría afectar la capacidad de hacer inferencias estadísticas precisas y suponer que el modelo está correctamente especificado.

1. ¿Cuál es la afectación en un modelo o sus pronósticos, cuando los errores son heterocedasticos?

La heterocedasticidad indica que los errores no tienen varianza constante, lo que puede llevar a que las predicciones y los intervalos de confianza sean inexactos. Los errores grandes en ciertos momentos pueden influir más en el modelo, lo que puede resultar en un ajuste incorrecto.

1. Si los residuos parecen estar auto correlacionados, ¿qué alternativas tiene para intentar mejorar el modelo?

Algunas opciones incluyen:

* + Añadir términos de autocorrelación al modelo, como en los modelos ARIMA o SARIMA.
  + Incluir más variables explicativas o rezagadas de la variable dependiente.
  + Descomponer la serie de tiempo en componentes como tendencia y estacionalidad para mejorar la precisión.
  + Ajustar los datos utilizando diferencias para eliminar la autocorrelación.

1. De acuerdo con el libro, ¿cuál es el proceso iterativo de un flujo de pronóstico?

El flujo iterativo de un pronóstico típicamente incluye los siguientes pasos:

* Recolección y preparación de los datos.
* Exploración y análisis de los datos.
* Selección del modelo.
* Entrenamiento del modelo.
* Evaluación del modelo con datos de prueba o validación.
* Ajuste del modelo si es necesario.
* Implementación del pronóstico final.
* Monitoreo y ajustes continuos basados en nuevos datos.

1. ¿Cuál es el propósito de separar tus datos entre entrenamiento y prueba?

El propósito es evaluar el rendimiento del modelo de manera imparcial. Los datos de entrenamiento se utilizan para ajustar el modelo, mientras que los datos de prueba permiten verificar si el modelo generaliza bien a datos nuevos que no ha visto antes.

1. ¿Cuál es la proporción recomendada para dividir los datos entre entrenamiento y prueba?

La proporción más comúnmente recomendada es 70% para entrenamiento y 30% para prueba. Sin embargo, también se utilizan divisiones de 80/20 o 60/40 dependiendo del tamaño del conjunto de datos y la naturaleza del problema.

1. ¿Para qué sirve proceso de cross-validaion (o back-testing)?

El cross-validation o back-testing sirve para evaluar la capacidad de generalización de un modelo. Consiste en dividir los datos en varios subconjuntos (folds), entrenar el modelo en uno de ellos y probarlo en los otros, lo que ayuda a reducir el sesgo y la varianza en la estimación del rendimiento del modelo.

1. Si estuvieras ante un caso en el cual las métricas de error del pronóstico de un proceso de cross-validation (o back-testing) apuntaran a distintos modelos, ¿qué haría para decidir cuál modelo utilizar?

Si las métricas no son consistentes, podrías:

Evaluar el contexto y el uso final del modelo: un modelo con menor error en métricas como MAPE podría ser más útil si la precisión relativa es importante.

Considerar la simplicidad del modelo: si dos modelos tienen rendimientos similares, es recomendable elegir el más simple (según la navaja de Ockham).

Repetir el proceso con diferentes divisiones de datos o usar un conjunto de validación adicional.

Evaluar el comportamiento del modelo en casos extremos o datos atípicos para ver cuál maneja mejor esos casos.

1. ¿Qué es una regresión lineal?

La regresión lineal es un modelo estadístico que busca predecir el valor de una variable dependiente (Y) en función de una o más variables independientes (X). El objetivo es encontrar la relación lineal entre las variables

1. ¿Qué es el método de mínimos cuadrados?

El método de mínimos cuadrados es una técnica que busca minimizar la suma de los cuadrados de los errores (la diferencia entre los valores predichos por el modelo y los valores reales). Al minimizar esta suma, se obtiene la mejor línea recta que ajusta los datos en una regresión lineal.

1. ¿Cuál es la función objetivo y pasos para obtener los parámetros beta β de una regresión lineal simple ?

La **función objetivo** es minimizar la suma de los cuadrados de los errores: S(β0,β1)=∑i=1n(yi−(β0+β1xi))2S(\beta\_0, \beta\_1) = \sum\_{i=1}^{n} \left( y\_i - (\beta\_0 + \beta\_1 x\_i) \right)^2S(β0​,β1​)=i=1∑n​(yi​−(β0​+β1​xi​))2

**Pasos**:

1. Calcular los derivados parciales de la función objetivo respecto a β0\beta\_0β0​ y β1\beta\_1β1​.
2. Igualar las derivadas a cero para encontrar los valores de β0\beta\_0β0​ y β1\beta\_1β1​.
3. Resolver el sistema de ecuaciones resultante para obtener los valores óptimos de β0\beta\_0β0​ y β1\beta\_1β1​.
4. ¿Cuáles son los supuestos de regresión lineal simple? Describe los correspondientes supuestos para los datos y para los errores.

Supuestos para los datos:

Relación lineal entre X e Y.

Los valores de X son fijos o controlados (sin errores).

Supuestos para los errores:

Independencia: Los errores son independientes entre sí.

Normalidad: Los errores siguen una distribución normal.

Homocedasticidad: La varianza de los errores es constante en todos los niveles de X.

Media cero: La media de los errores es cero.

1. ¿Qué es R2?

Es el coeficiente de determinación que mide qué proporción de la variación total de la variable dependiente es explicada por las variables independientes en el modelo. Su valor oscila entre 0 y 1, donde 1 indica que el modelo explica perfectamente la variación de Y.

1. ¿Cuál es la diferencia entre regresión lineal simple y regresión lineal múltiple y escribe la función en forma matricial?

En la regresión lineal simple, hay solo una variable independiente xxx, mientras que en la regresión lineal múltiple, hay varias variables independientes x1,x2,…,xn​.

Forma matricial de la regresión lineal múltiple:

y=Xβ+ε

Donde:

* y es el vector de la variable dependiente.
* X es la matriz de las variables independientes (incluyendo una columna de unos para el intercepto).
* Β es el vector de coeficientes.
* ε es el vector de errores.

1. Describe la función objetivo y los pasos para obtener el vector de parámetros β

Función objetivo:

S(β)=(y−Xβ)^⊤(y−Xβ)

Pasos:

* Derivar la función objetivo respecto a β
* Igualar la derivada a cero: X⊤Xβ=X⊤y
* Resolver para β: β=(X^⊤X)^{−1}X^⊤y

1. ¿Qué es la multicolinealidad y por qué no es buena para la regresión lineal múltiple?

La multicolinealidad ocurre cuando dos o más variables independientes están altamente correlacionadas entre sí. Esto no es deseable porque dificulta la estimación precisa de los coeficientes 𝛽, aumenta la varianza de los estimadores y puede hacer que los resultados del modelo sean inestables o no interpretables.

1. ¿Qué son las métricas AIC, BIC y R2 ajustada para seleccionar los predictores (variables exógenas, X’s, features)

AIC (Criterio de Información de Akaike): Evalúa el ajuste del modelo penalizando por el número de parámetros, buscando el balance entre ajuste y simplicidad.

BIC (Criterio de Información Bayesiano): Similar al AIC, pero penaliza más el número de parámetros, favoreciendo modelos más simples.

R^2 ajustada: Similar al R^2, pero ajusta por el número de predictores en el modelo, penalizando la inclusión de variables irrelevantes.

1. Describe las técnicas para selección de predictores (Stepwise, Lasso, Ridge, Elastic Net)

Stepwise: Procedimiento automático que añade o elimina predictores basándose en criterios estadísticos como AIC, BIC o ppp-valores.

Lasso: Aplica una penalización L1, lo que puede llevar a que algunos coeficientes se reduzcan a cero, seleccionando así un subconjunto de predictores.

Ridge: Utiliza una penalización L2 que no elimina predictores, pero reduce el tamaño de los coeficientes para evitar sobreajuste.

Elastic Net: Combina las penalizaciones L1(Lasso) y L2(Ridge) para obtener un balance entre selección de variables y regularización del modelo.