Análisis de Series Temporales

Clase 3 - Introducción a Pronósticos

Rodrigo Del Rosso
RDelRosso-ext@austral.edu.ar

16 de Octubre de 2021

UNIVERSIDAD AUSTRAL | INGENIERÍA

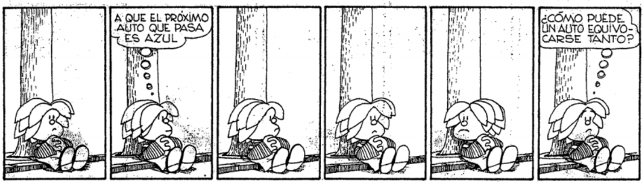
CIENCIA DE DATOS
Maestría en Ciencia de Datos

Agenda

- Intro a pronósticos con modelos $ARMA(p, q)$
- Intervalos de predicción
- Algunos métodos simples
- Medidas de performance

Predicción con Modelos ARMA(p,q)

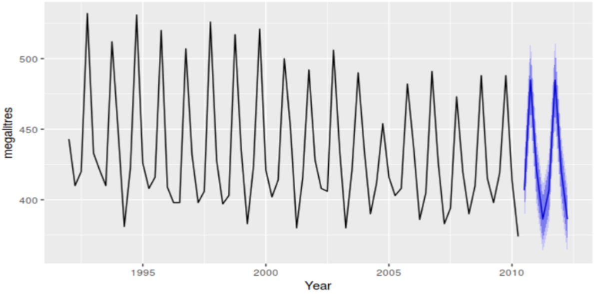
Kolmogorov-Wiener. . .



Estacionariedad

¿Por qué es bueno que las series sean estacionarias?

- Con series estacionarias podemos obtener predicciones fácilmente.
- Como la media es constante, podemos estimarla con todos los datos, y utilizar este valor para predecir una nueva observación.
- También se pueden obtener intervalos de predicción (confianza) para las predicciones asumiendo que Y_t sigue una distribución conocida, por ejemplo, normal.



Introducción a pronósticos

Los pronósticos puntuales se pueden calcular utilizando los siguientes tres pasos,

1. Expandir la ecuación ARIMA para que Y_t está en el lado izquierdo y todos los demás términos están en el derecho.
2. Reescribir la ecuación reemplazando t con $t + l$.
3. En el lado derecho de la ecuación, reemplazar las observaciones futuras con sus pronósticos, los errores futuros con cero y los errores pasados con los residuos correspondientes.

Empezando con $l = 1$, estos pasos se repiten para $l = 2, 3, \ldots$ hasta que se hayan calculado todos los pronósticos.

Introducción a pronósticos con $ARMA(p,q)$

Sea $\{Y_t\}: ARMA(p,q)$ un proceso estocástico estacionario de 2do orden $\{Y_t\}: I(0)$ discreto en el dominio en el tiempo, donde

$\{Y_t\}$: en variables centradas $\quad \{\epsilon_t\}: WN \sim N(0, \sigma_\epsilon)$

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p\right) Y_t = \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) \epsilon_t$$

$$\Phi(B)\, Y_t = \Theta(B) \epsilon_t$$

La ecuación $\Phi(B)\, Y_t = 0$ produce p raíces $G_i \quad (i = 1, 2, \ldots, p)$.

Condición: $|G_i| > 1$ ó $\left|G_i^{-1}\right| < 1$

Es decir, que las p raíces sean todas, en módulo, mayores que 1.

Introducción a pronósticos con $ARMA(p, q)$

Se cuenta con la siguiente secuencia de valores $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$ que representa la trayectoria asumida por el proceso generador de datos. A partir de dicha secuencia de valores se intentará predecir o pronosticar con el modelo entrenado.

Algunas cuestiones a tener en cuenta:

Los procesos AR son de memoria corta, por lo tanto la FAC decrece rápidamente de forma exponencial. Por lo tanto su utilidad es para predicción de períodos muy cortos.

Por ejemplo, un $AR(1)$ tendrá una capacidad de predicción para 1 período y un $AR(2)$ para 2 períodos y así sucesivamente.

Ahora bien, con los modelos $ARMA(p, q)$ se utilizará la parte AR con fines de predicción.

Introducción a pronósticos con $ARMA(p,q)$

El problema a resolver es el siguiente,

$$Y_{t+l} = ? \qquad (l = 1, 2, \ldots)$$

Con la serie observada $\{Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\}$ se intentará predecir Y_{t+l}, que representa el parámetro a estimar.

A l se lo denomina **"horizonte de predicción"** y el estimador del predictor se denota de la forma siguiente,

$$\hat{Y}_{t+l/t} \equiv \ f(t,l)(Y_{t-n}, Y_{t-n+1}, \ldots, Y_t)$$

El predictor es una función de la serie cronológica, donde $\hat{Y}_{t+l/t}$ representa el valor que asume el predictor, es decir lo que se denomina "la realización" (estimación).

El predictor es aquel que minimiza la varianza del error de predicción, forma parte de la familia de predictores L_2.

Introducción a pronósticos con $ARMA(p,q)$

El valor desconocido que se intenta predecir se puede descomponer como la suma de dos términos,

$$Y_{t+l} = \hat{Y}_{t+l/t} + e_{l/t}$$

Donde,

Y_{t+l} *es el verdadero valor*

$\hat{Y}_{t+l/t}$ *es la predicción condicionada al último dato en el momento t*

$e_{l/t} \equiv e(l,t)$ *es el error de predicción*

El predictor es aquel que minimiza la varianza del error de predicción, forma parte de la familia de predictores L_2. El predictor minimiza la siguiente expresión,

$$min\ \sigma^2 \left[e\left(l,t\right) \right] = min\ E\left\{ \left[Y_{t+l} - \hat{Y}_{t+l/t} \right]^2 \right\}$$

Introducción a pronósticos con $ARMA(p, q)$

A partir de la demostración (buscarla en la literatura). Se llega a demostrar lo siguiente,

$$\hat{Y}_{t+l/t} = E(Y_{t+l}/Y_{t-n}, Y_{t-n+1}, \ldots, Y_t)$$

Es aplicable solamente si se conoce la función de probabilidad conjunta de la variable multidimensional $\left\{ Y_{t-n}, Y_{t-n+1}, \ldots, Y_t \right\}$ o si se cuenta con una representación para $\{Y_t\}$ que involucre una cantidad finita de coeficientes.

$$\hat{Y}_{t+l/t} = \phi_1 \hat{Y}_{t+l-1/t} + \ldots + \phi_p \hat{Y}_{t+l-p/t} + E(\epsilon_{t+l}/Y_t) - \theta_1 E(\epsilon_{t+l-1}/Y_t) - \ldots - \theta_q E(\epsilon_{t+l-q}/Y_t)$$

$$E\left(\epsilon_{t+j}/Y_t\right) = \begin{cases} 0 & \text{para } \; j > 0 \\ \\ \epsilon_{t+j} & \text{para } \; j \leq 0 \end{cases}$$

El valor esperado condicionado es lineal en $\left\{ Y_{t-n}, Y_{t-n+1}, \ldots, Y_t \right\}$.

Por lo tanto,

$$\hat{Y}_{t+l/t} = a_0 Y_t + a_1 Y_{t-1} + \ldots + a_n Y_{t-n}$$

El problema es encontrar el valor de los a_i que **minimice la varianza del error de predicción**.

Introducción a pronósticos con $ARMA(p, q)$

Si se cumple la condición de invertibilidad y estacionariedad se verifica que,

$$a_i = f\left(\phi_i, \theta_i\right)$$

Cada modelo que se constituye tiene una única manera de predictor.

Ejemplo: $\{Y_t\} : ARMA(1, 0) \Rightarrow Y_t = \phi_1 Y_{t-1} + \epsilon_t$

$Y_{t+1} = \phi_1 Y_t + \epsilon_{t+1}$

$\hat{Y}_{t+1/t} = E\left(Y_{t+1}/Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\right) = E\left(\phi_1 Y_t + \epsilon_{t+1}/Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\right)$

$$\hat{Y}_{t+1/t} = E\left(\phi_1 Y_t/Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\right) + E\left(\epsilon_{t+1}/Y_{t-n}, Y_{t-n+1}, \ldots, Y_t\right)$$

$$\hat{Y}_{t+1/t} = \phi_1 Y_t$$

$$e_{1/t} \equiv \; e\left(1, t\right) = Y_{t+1} - \hat{Y}_{t+1/t} = \epsilon_{t+1}$$

$$\sigma^2\left[e\left(1, t\right)\right] = \sigma_\epsilon^2$$

Toda expresión que posea un subíndice mayor a t es una variable aleatoria.

$$Y_{t+2} = \phi_1 Y_{t+1} + \epsilon_{t+2}$$

$$Y_{t+2} = \phi_1 \left(\phi_1 Y_t + \epsilon_{t+1}\right) + \epsilon_{t+2} = \phi_1^2 Y_t + \phi_1 \epsilon_{t+1} + \epsilon_{t+2}$$

El predictor y el error de predicción para un horizonte de predicción igual a $l = 2$,

$$\hat{Y}_{t+2/t} = \phi_1^2 Y_t$$

$$e_{2/t} \equiv \; e\left(2, t\right) = \phi_1 \epsilon_{t+1} + \epsilon_{t+2}$$

La varianza del error de predicción para $l = 2$ es mayor que para $l = 1$,

$$\sigma^2 \left[e\left(2, t\right)\right] = \left(1 + \phi_1^2\right) \sigma_\epsilon^2$$

Para un horizonte genérico l por recursividad se llega a lo siguiente,

$$Y_{t+l} = \phi_1^l Y_t + \sum_{j=0}^{l-1} \phi_1^j \epsilon_{t+l-j}$$

El predictor y el error de predicción para el horizonte de predicción genérico l, $\hat{Y}_{t+l/t} = \phi_1^l Y_t$

$$e_{l/t} \equiv \; e\left(l, t\right) = \sum_{j=0}^{l-1} \phi_1^j \epsilon_{t+l-j}$$

La varianza del error de predicción para el período l crece en progresión geométrica a una razón ϕ_1^2,

$$\sigma^2 \left[e\left(l, t\right)\right] = \sigma_\epsilon^2 \sum_{j=0}^{l-1} \phi_1^{2j}$$

Comentarios:

$\hat{Y}_{t+l/t} = \phi_1^l Y_t$ es una función exponencialmente decreciente, dado que si cumple con la condición de Estacionariedad, entonces $|\phi_1| < 1$ y las predicciones se acercan a cero muy rápidamente si el modelo está expresado en variables centradas, en caso contrario se acerca rápidamente al valor esperado $m(Y)$.

Respecto a la varianza del error de predicción (condicionada), la misma converge en el límite a la varianza incondicionada de la serie. Al crecer el horizonte de predicción, la calidad de la predicción empeora, dado que aumenta la incertidumbre respecto a períodos futuros.

$$\sigma^2\left[e\left(l,t\right)\right] = \sigma_\epsilon^2 \sum_{j=0}^{l-1} \phi_1^{2j} = \sigma_\epsilon^2 \left(1 + \phi_1^2 + \phi_1^4 + \ldots + \phi_1^{2(l-1)}\right) = \sigma_\epsilon^2 \frac{1-\phi_1^{2l}}{1-\phi_1^2}$$

$$\lim_{l\to\infty} \sigma^2\left[e\left(l,t\right)\right] = \lim_{l\to\infty} \sigma_\epsilon^2 \frac{1-\phi_1^{2l}}{1-\phi_1^2} = \frac{\sigma_\epsilon^2}{1-\phi_1^2} = \sigma_Y^2$$

Si $\{\epsilon_t\} \sim N$ se demuestra que,

$$\hat{Y}_{t+l/t} \;\sim N\left(Y_{t+l}, \sigma\left[e\left(l, t\right)\right]\right)$$

$$\hat{Y}_{t+l/t} = \phi_1^l Y_t$$

$$\lim_{l \to \infty} \phi_1^l Y_t = 0 = E\left[Y_t\right]$$

El predictor converge a la Esperanza Matemática Incondicionada.

En un $AR(1)$ la velocidad de retorno al valor esperado se mide por el coeficiente conocido como Vida Media. Formalmente,

$$l^* = \frac{1}{\ln(2|\phi_1|)}$$

EJEMPLO

Ejemplo: $MA(q)$

$$Y_t = \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \ldots - \theta_q\epsilon_{t-q}$$

$\{Y_t\}$: en variables centradas

$\{\epsilon_t\}$: $WN \sim N(0, \sigma_\epsilon)$

Para un horizonte de predicción $l = 1$ se obtiene lo siguiente,

$$Y_{t+1} = \epsilon_{t+1} - \theta_1\epsilon_t - \theta_2\epsilon_{t-1} - \ldots - \theta_q\epsilon_{t-q+1}$$

$$\hat{Y}_{t+1/t} = -\theta_1\epsilon_t - \theta_2\epsilon_{t-1} - \ldots - \theta_q\epsilon_{t-q+1}$$

$$e_{1/t} \equiv \; e(1,t) = Y_{t+1} - \hat{Y}_{t+1/t} = \epsilon_{t+1}$$

$$\sigma^2\left[e(1,t)\right] = \sigma_\epsilon^2$$

EJEMPLO - continuación

Para un horizonte de predicción $l = 2$ se obtiene lo siguiente,

$$Y_{t+2} = \epsilon_{t+2} - \theta_1 \epsilon_{t+1} - \theta_2 \epsilon_t - \ldots - \theta_q \epsilon_{t-q+2}$$

$$\hat{Y}_{t+2/t} = -\theta_2 \epsilon_t - \ldots - \theta_q \epsilon_{t-q+2}$$

$$e_{2/t} \equiv \; e(2,t) = Y_{t+2} - \hat{Y}_{t+2/t} = \epsilon_{t+2} - \theta_1 \epsilon_{t+1} \Rightarrow \sigma^2 \left[e(2,t) \right] = \left(1 + \theta_1^2\right) \sigma_\epsilon^2$$

Para un horizonte de predicción $l = q$ se obtiene lo siguiente,

$$Y_{t+q} = \epsilon_{t+q} - \theta_1 \epsilon_{t+q-1} - \theta_2 \epsilon_{t+q-2} - \ldots - \theta_q \epsilon_t$$

$$\hat{Y}_{t+q/t} = -\theta_q \epsilon_t$$

$$e_{q/t} \equiv \; e(q,t) = \sum_{j=0}^{q-1} \theta_j \epsilon_{t+q-j} \; \Rightarrow \sigma^2 \left[e(q,t) \right] = \left(1 + \theta_1^2 + \ldots + \theta_{q-1}^2\right) \sigma_\epsilon^2$$

EJEMPLO - continuación

Para un horizonte de predicción $l > q$ la función de predicción coincide con la Esperanza Matemática no condicionada de Y_t (tarda q períodos en retornar al valor esperado) y la varianza del error de predicción coincide con la varianza no condicionada de Y_t.

$$Y_{t+l} = \epsilon_{t+l} - \theta_1\epsilon_{t+l-1} - \theta_2\epsilon_{t+l-2} - \ldots - \theta_q\epsilon_{t+l-q}$$

$$\hat{Y}_{t+l/t} = \begin{cases} -\theta_l\epsilon_t - \theta_{l+1}\epsilon_{t-1} - \ldots - \theta_q\epsilon_{t-q+l} & \text{para} \quad l \leq q \\ \\ 0 & \text{para} \quad l > q \end{cases}$$

continuando...

$$e_{l/t} \equiv \; e\left(l, t\right) = \begin{cases} \displaystyle\sum_{j=0}^{l-1} \theta_j \epsilon_{t+l-j} & \text{para} \quad l \leq q \\ \\ \displaystyle\sum_{j=0}^{q} \theta_j \epsilon_{t+l-j} & \text{para} \quad l > q \end{cases}$$

$$\sigma^2 \left[e\left(q, t\right)\right] = \begin{cases} \left(1 + \theta_1^2 + \ldots + \theta_{l-1}^2\right) \sigma_\epsilon^2 & \text{para} \quad l \leq q \\ \\ \left(1 + \theta_1^2 + \ldots + \theta_q^2\right) \sigma_\epsilon^2 & \text{para} \quad l > q \end{cases}$$

OTRO EJEMPLO

Ejemplo: $ARMA(1,1)$

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}$$

Para un horizonte de predicción $l = 1$ se obtiene lo siguiente,

$$Y_{t+1} = \phi_1 Y_t + \epsilon_{t+1} - \theta_1 \epsilon_t$$

$$\hat{Y}_{t+1/t} = \phi_1 Y_t - \theta_1 \epsilon_t$$

$$e_{1/t} \equiv \; e\left(1,t\right) = Y_{t+1} - \hat{Y}_{t+1/t} = \epsilon_{t+1} \Rightarrow \sigma^2\left[e\left(1,t\right)\right] = \sigma_\epsilon^2$$

Para un horizonte de predicción $l = 2$ se obtiene lo siguiente,

$$Y_{t+2} = \phi_1 Y_{t+1} + \epsilon_{t+2} - \theta_1 \epsilon_{t+1} = \phi_1 \left(\phi_1 Y_t + \epsilon_{t+1} - \theta_1 \epsilon_t\right) + \epsilon_{t+2} - \theta_1 \epsilon_{t+1}$$

$$\hat{Y}_{t+2/t} = \phi_1 \left(\phi_1 Y_t - \theta_1 \epsilon_t\right)$$

$$e_{2/t} \equiv \; e\left(2,t\right) = \epsilon_{t+2} + \left(\phi_1 - \theta_1\right) \epsilon_{t+1} \Rightarrow \sigma^2\left[e\left(2,t\right)\right] = \left(1 + \left(\phi_1 - \theta_1\right)^2\right) \sigma_\epsilon^2$$

OTRO EJEMPLO - continuación

Para un horizonte de predicción l se obtiene lo siguiente,

$$Y_{t+l} = \phi_1 Y_{t+l-1} + \epsilon_{t+l} - \theta_1 \epsilon_{t+l-1}$$

$$\hat{Y}_{t+l/t} = \phi_1^{l-1} \left(\phi_1 Y_t - \theta_1 \epsilon_t\right)$$

$$e_{l/t} \equiv \; e\left(l, t\right) = \epsilon_{t+l} + \left(\phi_1 - \theta_1\right) \sum_{j=1}^{l-1} \phi_1^{j-1} \epsilon_{t+l-j}$$

$$\sigma^2 \left[e\left(l, t\right)\right] = \left(1 + \left(\phi_1 - \theta_1\right)^2 \sum_{j=1}^{l-1} \phi_1^{2(j-1)}\right) \sigma_\epsilon^2$$

$$\lim_{l \to \infty} \sigma^2 \left[e\left(l, t\right)\right] = \left[1 + \frac{\left(\phi_1 - \theta_1\right)^2}{1 - \phi_1^2}\right] \sigma_\epsilon^2 = \sigma_Y^2$$

INTERVALOS DE PREDICCIÓN

El cálculo de los intervalos de predicción de ARIMA es más difícil, y los detalles están más allá del alcance de esta presentación. Solo daremos algunos ejemplos simples.

Si $\{\epsilon_t\} \sim N$ se demuestra que,

$$\hat{Y}_{t+l/t} \sim N\left(Y_{t+l}, \sigma\left[e\left(l,t\right)\right]\right)$$

Puede ser de interés construir un intervalo de predicción de la forma siguiente,

$$P\left(A \leq Y_{t+l} \leq B\right) = 1 - \alpha$$

$$P\left(Z_{\frac{\alpha}{2}} \leq \frac{\hat{Y}_{t+l/t} - Y_{t+l}}{\sigma[e(l,t)]} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(Z_{\frac{\alpha}{2}}\sigma\left[e\left(l,t\right)\right] \leq \hat{Y}_{t+l/t} - Y_{t+l} \leq Z_{1-\frac{\alpha}{2}}\sigma\left[e\left(l,t\right)\right]\right) = 1 - \alpha$$

INTERVALOS DE PREDICCIÓN - continuación...

$$P\left(\hat{Y}_{t+l/t} - Z_{1-\frac{\alpha}{2}}\sigma\left[e\left(l,t\right)\right] \leq Y_{t+l} \leq \hat{Y}_{t+l/t} + Z_{1-\frac{\alpha}{2}}\sigma\left[e\left(l,t\right)\right]\right) = 1-\alpha$$

La calidad del modelo depende de $\sigma\left[e\left(l,t\right)\right]$.

La amplitud del intervalo de predicción se determina de la forma siguiente,

$$L = 2\ Z_{1-\frac{\alpha}{2}}\sigma\left[e\left(l,t\right)\right]$$

Resultados más generales y otros casos especiales de intervalos de predicción de varios pasos para un $ARIMA(p,d,q)$, se dan en libros de texto más avanzados como **Brockwell & Davis (2016)**.

INTERVALOS DE PREDICCIÓN

Importante

- Los intervalos de predicción para los modelos *ARIMA* se basan en suposiciones de que los residuos no están correlacionados y normalmente están distribuidos.
- Si alguno de estos supuestos no se cumple, entonces los intervalos de predicción pueden ser incorrectos.
- Por esta razón, siempre es conveniente graficar la FAC y el histograma de los residuos para verificar los supuestos antes de producir intervalos de predicción.
- En general, los intervalos de predicción de los modelos *ARIMA* aumentan a medida que aumenta el horizonte de pronóstico.
- Para modelos estacionarios (es decir, con $d = 0$) convergerán, de modo que los intervalos de predicción para horizontes largos son esencialmente iguales.
- Para $d \geq 1$, los intervalos de predicción continuarán creciendo en el futuro.

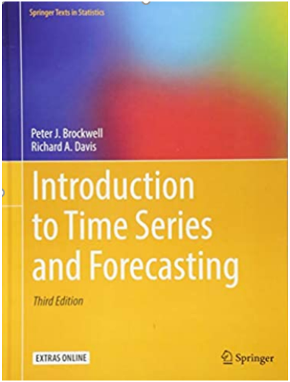
INTERVALOS DE PREDICCIÓN

Como con la mayoría de los cálculos de intervalos de predicción, los intervalos basados en *ARIMA* tienden a ser demasiado estrechos.

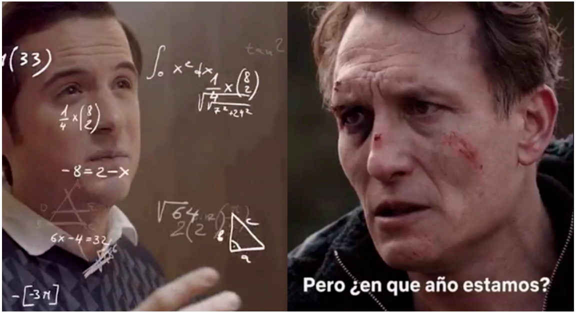
Esto ocurre porque solo se ha tenido en cuenta la variación en los errores.

También hay una variación en las estimaciones de los parámetros, y en el orden del modelo, que no se ha incluido en el cálculo.

Además, el cálculo supone que los patrones históricos que se han modelado continuarán en el período de pronóstico.



Algunos Métodos Simples de *Forecasting*



Algunos Métodos Simples de *Forecasting*

Algunos métodos de pronóstico son extremadamente simples y sorprendentemente efectivos.

1. Average Method
2. Naïve Method
3. Seasonal Naïve Method
4. Drift Method

Algunos Métodos Simples de *Forecasting*

Average Method

Los pronósticos de todos los valores futuros son iguales al promedio (o "media") de los datos históricos. Si los datos históricos se denoten por $y_1, y_2, \ldots, y_t$ entonces puede escribir los pronósticos como,

$$\hat{y}_{t+h/t} = \sum_{i=1}^{t} \frac{y_i}{t}$$

Naive Method

Configuramos todos los pronósticos para que sean el valor del último observación. Es decir,

$$\hat{y}_{t+h/t} = y_t$$

Un pronóstico ingenuo es óptimo cuando los datos siguen Random Walk (Pronósticos de Random Walk)

Algunos Métodos Simples de *Forecasting*

Seasonal Naive Method

Un método similar es útil para datos altamente estacionales. En este caso, establecemos cada pronóstico para ser igual al último valor observado de la misma estación del año (por ejemplo, el mismo mes del año anterior). Formalmente, el pronóstico del tiempo $t+h$ se escribe como,

$$\hat{y}_{t+h/t} = \hat{y}_{t+h-m(k+1)}$$

Donde m es el período estacional, y k es la parte entera de $\frac{h-1}{m}$ (es decir, el número de años completos en el período de pronóstico anterior al tiempo $t+h$).

Esto parece más complicado de lo que realmente es. Por ejemplo, con datos mensuales, el pronóstico para todos los valores futuros de febrero es igual al último valor de febrero observado. Con datos trimestrales, el pronóstico de todos los valores futuros de $Q2$ es igual al último valor observado de $Q2$.

Se pueden formalizar reglas similares para otros meses y trimestres, y para otros períodos estacionales.

Algunos Métodos Simples de *Forecasting*

Drift Method

Una variación del método ingenuo es permitir que los pronósticos aumenten o disminuyan con el tiempo, donde se establece la cantidad de cambio a lo largo del tiempo (llamada deriva). El cambio promedio visto en los datos históricos.

Así, el pronóstico del tiempo es dado por

$$\hat{y}_{t+h/t} = y_t + \frac{h}{t-1}\sum_{j=2}^{t}(y_j - y_{j-1}) = y_t + h\left(\frac{y_t - y_1}{t-1}\right)$$

Esto es equivalente a dibujar una línea entre la primera y la última observación, y extrapolando hacia el futuro.

Evaluación de la Performance del Pronóstico

Evaluación de la Performance del Pronóstico

Conjuntos de entrenamiento y prueba

Es importante evaluar la precisión del pronóstico utilizando pronósticos genuinos. En consecuencia, **el tamaño de los residuos no es una indicación confiable de cuán grandes pueden ser los errores de pronóstico reales**.

La precisión de los pronósticos solo se puede determinar considerando qué tan bien se desempeña un modelo con los nuevos datos que no se utilizaron al ajustar el modelo.

Al elegir modelos, es una práctica común separar los datos disponibles en dos partes, **datos de entrenamiento y prueba**, donde los datos de entrenamiento se usan para estimar cualquier parámetro de un método de pronóstico y los datos de prueba se usan para evaluar su precisión. Debido a que los datos de la prueba no se usan para determinar los pronósticos, deberían proporcionar una indicación confiable de qué tan bien es probable que el modelo pronostique los nuevos datos.

Evaluación de la Performance del Pronóstico

El tamaño del conjunto de prueba generalmente es de aproximadamente el **20% de la muestra total**, aunque este valor depende de cuánto dura la muestra y qué tan lejos desea pronosticar.

El conjunto de prueba idealmente debería ser al menos tan grande como el horizonte de pronóstico máximo requerido. Deben tenerse en cuenta los siguientes puntos:

- Un modelo que se ajuste bien a los datos de entrenamiento no necesariamente pronosticará bien.
- Siempre se puede obtener un ajuste perfecto utilizando un modelo con **suficientes parámetros**. El ajuste excesivo de un modelo a los datos es tan malo como no identificar un patrón sistemático en los datos.

Importante

Algunas referencias describen el conjunto de prueba como el "conjunto de retención" ($hold - out$ set) porque estos datos están "retenidos" ("*heldout*") de los datos utilizados para el ajuste.

Otras referencias llaman al conjunto de entrenamiento los "datos dentro de la muestra" y la prueba establece los "datos fuera de la muestra" ("$out - of - sample$ data"). En este curso utilizaremos "datos de entrenamiento" y "datos de prueba".

Evaluación de la Performance del Pronóstico

Errores de previsión

Podemos medir la precisión del pronóstico resumiendo los errores de pronóstico de diferentes maneras.

Errores dependientes de la escala

Los errores de pronóstico están en la misma escala que los datos. Medidas de precisión que se basan solo en $e(l,t) = e_l$ son, por lo tanto, dependientes de la escala y no pueden usarse para hacer comparaciones entre series que involucran unidades diferentes. Las dos medidas dependientes de la escala más comúnmente utilizadas se basan en los errores absolutos o los errores al cuadrado,

$$Mean\ Absolute\ Error\ \ (MAE) = \sum_{i=1}^{n} \frac{|e_i|}{n}$$

$$Root\ Mean\ Squared\ Error\ \ (RMSE) = \sqrt{\sum_{i=1}^{n} \frac{e_i^2}{n}}$$

MAE vs RMSE

Continuando...

$$Mean\ Absolute\ Error\ \ (MAE) = \sum_{i=1}^{n} \frac{|e_i|}{n}$$

$$Root\ Mean\ Squared\ Error\ \ (RMSE) = \sqrt{\sum_{i=1}^{n} \frac{e_i^2}{n}}$$

Al comparar los métodos de pronóstico aplicados a una sola serie de tiempo, o a varias series de tiempo con las mismas unidades, el **MAE es popular ya que es fácil de entender y calcular**.

Un método de pronóstico que minimiza el **MAE conducirá a pronósticos de la mediana, mientras que minimizar el RMSE conducirá a pronósticos de la media**. En consecuencia, el RMSE también se usa ampliamente, a pesar de ser más difícil de interpretar.

Evaluación de la Performance del Pronóstico

Porcentaje de errores

Los errores porcentuales tienen la ventaja de estar libres de unidades y, por lo tanto, se usan con frecuencia para comparar los rendimientos de pronóstico entre conjuntos de datos. La medida más utilizada es:

$$Mean\ Absolute\ Percentage\ Error\ \ (MAPE) = \sum_{i=1}^{n} \frac{|p_i|}{n} = \sum_{i=1}^{n} \frac{|100e_i/Y_i|}{n}$$

Las medidas basadas en errores porcentuales **tienen la desventaja de ser infinitas o indefinidas si** $Y_t = 0$ para cualquier período t de interés, y tener valores extremos si Y_t está cerca de cero.

Otro problema con los errores porcentuales que a menudo se pasa por alto es que suponen que la unidad de medida tiene un cero significativo. Por ejemplo, un error porcentual no tiene sentido al medir la precisión de los pronósticos de temperatura en las escalas Fahrenheit o Celsius, porque la temperatura tiene un punto cero arbitrario..

Evaluación de la Performance del Pronóstico

Validación Cruzada de Series de Tiempo (*Cross Validation Time Series*)

Una versión más sofisticada de los conjuntos de entrenamiento / prueba es la validación cruzada de series temporales. En este procedimiento, hay una serie de conjuntos de prueba, cada uno de los cuales consta de una sola observación.

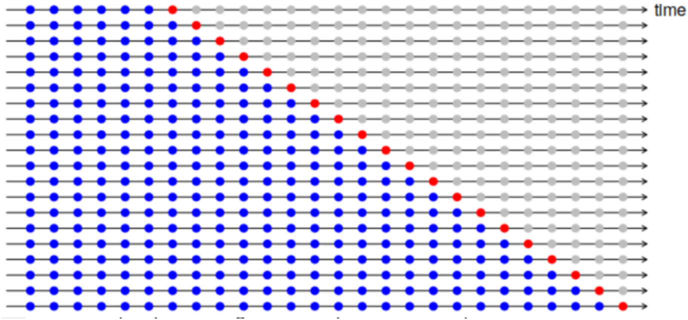
El conjunto de entrenamiento correspondiente consta solo de observaciones que ocurrieron antes de la observación que forma el conjunto de prueba.

Por lo tanto, no se pueden utilizar observaciones futuras para construir el pronóstico.

Dado que no es posible obtener un pronóstico confiable basado en un pequeño conjunto de entrenamiento, **las primeras observaciones no se consideran conjuntos de prueba**.

Validación Cruzada de Series de Tiempo

El siguiente diagrama ilustra la serie de conjuntos de entrenamiento y prueba, donde las observaciones azules forman los conjuntos de entrenamiento y las observaciones rojas forman los conjuntos de prueba.

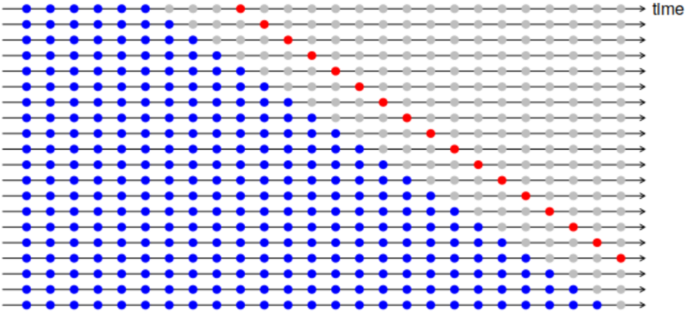


La precisión del pronóstico se calcula promediando los conjuntos de prueba. **Este procedimiento a veces se conoce como "evaluación de un origen de pronóstico continuo"** porque el "origen" en el que se basa el pronóstico avanza en el tiempo.

Validación Cruzada de Series de Tiempo

Con el pronóstico de series de tiempo, los pronósticos de un paso pueden no ser tan relevantes como los **pronósticos de varios pasos**. En este caso, el procedimiento de validación cruzada basado en un origen de pronóstico continuo puede modificarse para permitir el uso de errores de varios pasos.

Supongamos que estamos interesados en modelos que **produzcan pronósticos de buen paso adelante**. Entonces el diagrama correspondiente se muestra a continuación.



Validación Cruzada de Series de Tiempo

El RMSE de los residuos es menor, dado que los "pronósticos“ correspondientes se basan en un modelo **ajustado a todo el conjunto de datos**, en lugar de ser los pronósticos verdaderos.

Una buena manera de elegir el mejor modelo de pronóstico es encontrar el modelo con el RMSE más pequeño calculado utilizando la validación cruzada de series temporales.

Bibliografía sugerida

Hyndman, R. J., y Koehler, A. B. (2006). Another look at measures of forecast accuracy.*International Journal of Forecasting*, 22, 679–688.
https://robjhyndman.com/publications/another-look-at-measures-of-forecast-accuracy/

Lecturas

Brockwell et al.. Introduction to Time Series and Forecasting. Third Edition. Springer. Leer **Capítulo 5** sobre modelización y forecasting con modelos ARMA.

Chatfield et al.. The Analysis of Time Series. An Introduction with R. Seventh Edition. CRC Press. Leer **Capítulo 5** destinado a Forecasting.

Hyndman. Forecasting: Principles and Practice. Second Edition. Monash University, Australia. Leer **Capítulo 3** destinado a herramientas de Forecasting y medidas de performance.

Uriel. Análisis de Series Temporales. Modelos ARIMA. Leer **Capítulo 7** destinado a Predicción.

Peña. Análisis de Series Temporales. Alianza Editorial. Leer **Capítulos 8** destinado a predicción con modelos ARIMA.

Shmueli et al.. Practical Time Series Forecasting with R. A Hands-On Guide. Second Edition. Axelrod Schnall Publishers. Leer **Capítulos 3 - 4** destinados a evaluación de la performance y un resumen de los métodos de forecasting.

Fin