Análisis de Series de Tiempo

Carrera de Especialización en Inteligencia Artificial

Práctica 2

Ing. Magdalena Bouza, Ing. Carlos German Carreño Romano

Criterios de Bondad de modelos

Criterios de bondad de modelo

Uno de los objetivos cuando analizamos series de tiempo es poder modelarlas. Esto incluye tanto elegir la familia de modelos correcta como hallar todos los parámetros de la misma.

Es necesario entonces poder contar con algún criterio que nos permita saber cuán bueno es nuestro modelo, es decir cuán cerca se encuentra la distribución especificada por el modelo a la distribución verdadera de los datos.

Divergencia de Kullback-Leibler (DK-L)

Supongamos que los datos se encuentra generados por un proceso desconocido $g(y)$, y que consideramos un modelo $f(y)$ para aproximarnos a $g(y)$.

Si conociéramos $f(y)$, podríamos utilizar la divergencia de Kullback-Leibler

$$D_{\mathrm{KL}}(g \parallel f) = \int_{-\infty}^{\infty} g(y) \log\left(\frac{g(y)}{f(y)}\right) dy = \mathbb{E}\left[\log\left(\frac{g(Y)}{f(Y)}\right)\right]$$

que nos da una medida de la pérdida de información que tenemos al aproximar $f(y)$ por $g(y)$. Cuanto más pequeño sea el valor de la divergencia K-L, más cerca estarán $g(y)$ y $f(y)$. El mejor modelo será entonces aquel minimice la DK-L.

Estimación de la divergencia K-L

Dado que en general no se conoce el valor verdadero de la distribución g(y), debemos estimar el valor de $D_{\mathrm{KL}}(g \parallel f)$, por lo cual debemos estimar su valor a partir de las muestras $y_1, \ldots, y_n$. Para ello, se asume que las muestras son observadas de forma independiente de g(y).

En primer lugar, vemos que podemos reescribir

$$D_{\mathrm{KL}}(g \parallel f) = \mathbb{E}\left[\log\left(\frac{g(Y)}{f(Y)}\right)\right] = \mathbb{E}[\log(g(Y))] - \mathbb{E}[\log(f(Y))]$$

El primer término no puede ser calculado, pero si lo que nos interesa es comparar modelos, podemos descartarlo pues será una constante común a todos.

Estimación de la divergencia K-L

El término $\mathbb{E}[\log(f(Y))] = \int \log(f(y))g(y)dy$ y tampoco se puede calcular. Sin embargo la ley de los grandes números garantiza que

$$\frac{1}{n}\sum_{i=1}^{n} \log(f(y_i)) \to \mathbb{E}[\log(f(Y))]$$

Recordemos que $\ell = \sum_{i=1}^{n} \log(f(y_i)) = \log(L)$ es el logaritmo de la verosimilitud (*log-likelihood*), donde L es la función de verosimilitud.

Obs: Minimizar$D_{\mathrm{KL}}(g \parallel f)$ equivale a maximizar la log-verosimiliud. Por lo tanto una forma natural de hallar los mejores parámetros para un modelo dado es utilizando el estimador de máxima verosimilitud.

Criterio de Información de Akaike (AIC)

El modelo de información de Akaike sirve para comparar la bondad de dos modelos propuestos, y se basa en los resultados analizados previamente.

$$\text{AIC} = 2\boxed{k} - 2\ell(\boxed{\hat{\theta}})$$

cantidad de parámetros   parámetros estimados por MV

Se elegirá entonces el modelo que alcance el menor AIC.

Observaciones:

- El resultado de AIC es asintótico pues se basa en la LGN
- AIC no nos dice cuán bueno es cada modelo sino cuál es el mejor de todos (podría ser que sean todos malos)

Criterio de información Bayesiano

Similar a la definición de AIC, el criterio de información Bayesiano también se basa en la log-verosimilitud, pero penaliza más una mayor cantidad de parámetros.

$$\mathrm{BIC} = k\log(n) - 2\ell(\hat{\theta})$$

Nuevamente, se preferirán los modelos con BIC más bajo.

Si bien la expresión es similar a AIC, la derivación de BIC es totalmente distinta. AIC aparece como consecuencia de un enfoque Bayesiano, donde lo que se busca es el modelo f que maximice $\mathbb{P}(f|y_1,\ldots,y_n)$.

Para una explicación más detallada ver “Elements of Statistical Learning”, Trevor Hastie, Robert Tibshirani, Jerome Friedman.

Modelos para series de tiempo estacionarias

Proceso lineal general

Sean $\{Y_t\}$ la serie de tiempo observada y $\{e_t\}$ una serie de ruido blanco no observable. $\{e_t\}$ es una secuencia de v.a. i.i.d, con media cero y varianza σ_e^2.

Un proceso lineal general es aquel que puede representarse como una combinación lineal de términos presentes y pasados del proceso de ruido blanco:

$$Y_t = e_t + a_1 e_{t-1} + a_2 e_{t-2} + \ldots$$

Si la cantidad de términos a sumar es infinita se pide que $\sum_{i=1}^{\infty} a_i^2 < \infty$.

Se puede demostrar que:

- $\mathbb{E}[Y_t] = 0$
- $C_k = cov(Y_t, Y_{t-k}) = \sigma_e^2 \sum_{i=1}^{\infty} a_i a_{i+k}, \quad k \geq 0$

Modelo de promedios móviles

Un **modelo de promedio móvil (MA)** es un modelo lineal general donde existe una cantidad finita términos $a_i \neq 0$.

Diremos que $\{Y_t\}$ es un modelo de promedio móvil de orden q (MA(q)) si

$$Y_t = e_t - a_1 e_{t-1} - \ldots - a_q e_{t-q}$$

Los parámetros de este modelo son los pesos $a_1, \ldots, a_q$.

Para estos modelos,

$$C_0 = (1 + a_1^2 + \ldots + a_q^2)\sigma_e^2$$

$$R_k = \begin{cases} \frac{-a_k + a_1 a_{k+1} + a_2 a_{k+2} + \ldots + a_{1-k} a_q}{1 + a_1^2 + \ldots + a_q^2} & k = 1, 2, \ldots q \\ 0 & k > q \end{cases}$$

Modelo autoregresivo

Los modelos autoregresivos incluyen regresiones sobre sí mismos.

Diremos que Y_t es un modelo autoregresivo de orden p (AR(p)) si cumple que

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \boxed{e_t} \rightarrow \text{innovación}$$

Suponemos entonces que para cada t, e_t es independiente de $Y_{t-1}, Y_{t-2}, \ldots$

El modelo AR(p) tiene asociado su polinomio característico definido como:

$$\phi(x) = 1 - a_1 x + a_2 x^2 + \ldots + a_p x^p$$

Se puede demostrar que el proceso es AR(q) es estacionario si las raíces $\phi(x)$ se encuentran fuera del círculo unitario (módulo mayor a 1)

Modelo AR(1)

Si p=1, tenemos que $Y_t = a_1 Y_{t-1} + e_t$, luego $\mathbb{E}[Y_t] = 0$.

Puedo tomar la varianza miembro para obtener

$$var(Y_t) = a_1^2 var(Y_{t-1}) + \sigma_e^2 \rightarrow C_0 = a_1^2 C_0 + \sigma_e^2 \Rightarrow C_0 = \frac{\sigma_e^2}{1-a_1^2}$$

$$|a_1| < 1$$

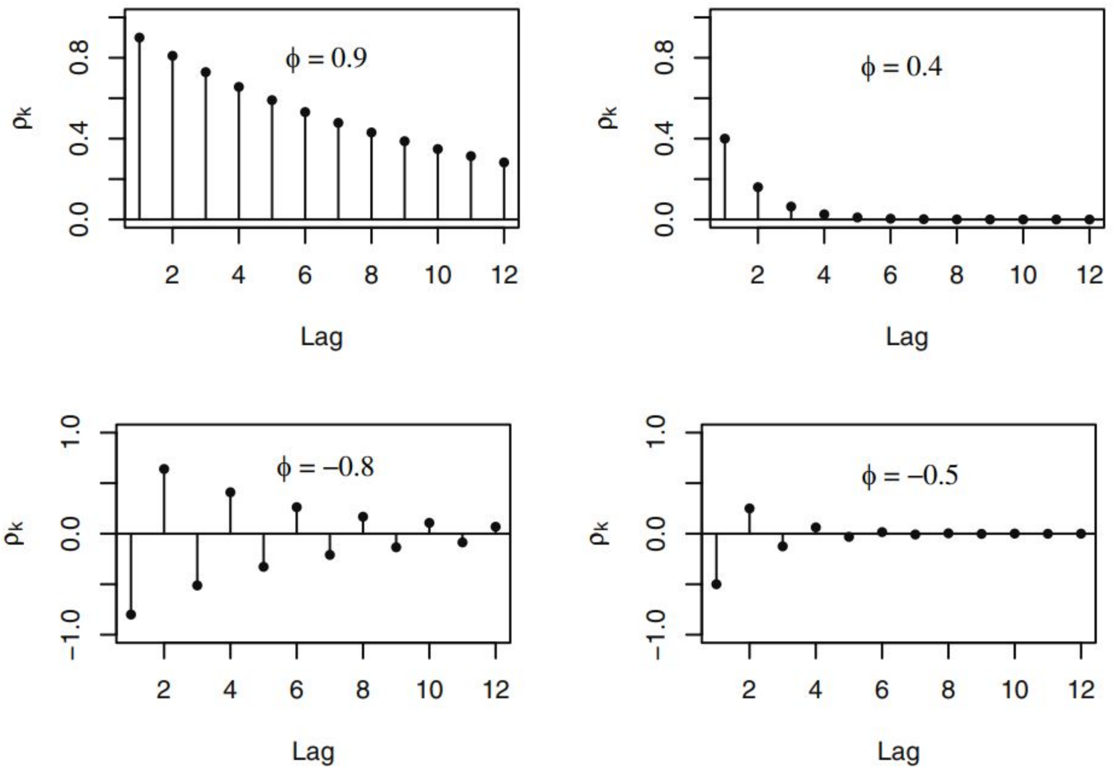
Además,

$$\begin{aligned} C_k = cov(Y_{t-k}, Y_t) &= cov(Y_{t-k}, a_1 Y_{t-1} + e_t) \\ &= a_1 \underbrace{cov(Y_{t-k}, Y_{t-1})}_{C_{k-1}} + \underbrace{cov(Y_{t-k}, e_t)}_{0} \\ &= a_1 C_{k-1} \end{aligned}$$

A partir del valor semilla de C_0 obtenemos de forma recursiva que

$$C_k = a_1^k \frac{\sigma_e^2}{1-a_1^2}$$

Modelo AR(1)



Modelo AR(1) como un proceso lineal general

Para comprender el modelo, la expresión dada para el modelo AR(1) es muy útil, sin embargo para muchas otras cosas es necesario llevarlo a la forma de }un proceso general lineal.

Para eso comenzamos reemplazando Y_{t-1} por $a_1 Y_{t-2} + e_{t-1}$:

$$Y_t = a_1(a_1 Y_{t-2} + e_{t-1}) + e_t = a_1^2 Y_{t-2} + a_1 e_{t-1} + e_t$$

aplicando la misma idea (k-1) veces:

$$Y_t = a_1^k e_{t-k} + a_1^{k-1} e_{t-k+1} + \ldots + a_1 e_{t-1} + e_t$$

Asumiendo $|a_1| < \infty$ e incrementando k sin límite tenemos que

$$Y_t = e_t + a_1 e_{t-1} + a_1^2 e_{t-2} + a_1^3 e_{t-3} + \ldots$$

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \psi_3 e_{t-3} + \ldots$$

Modelo AR(p)

Se puede mostrar que una condición necesaria (pero no suficiente) para que el proceso sea estacionario es que $a_1 + a_2 + \ldots + a_p < 1$ y $|a_p| < 1$.

Al igual que como hicimos para el AR(1), podemos calcular la función de autocovarianza y autocorrelación:

$$\begin{aligned} C_k = cov(Y_t, Y_{t-k}) &= cov(a_1 Y_{t-1}, Y_{t-k}) + \ldots + cov(a_p Y_{t-p}, Y_{t-k}) + cov(e_t, Y_{t-k}) \\ &= a_1 C_{k-1} + \ldots + a_p C_{k-p} \end{aligned}$$

$$R_k = \frac{cov(Y_t, Y_{t-k})}{C_0} = a_1 R_{k-1} + \ldots + a_p R_{k-p}$$

Modelo AR(p)

Evaluando para k=1,...,p y recordando que $R_0 = 1$ y $R_k = R_{-k}$ obtenemos las **ecuaciones de Yule-Walker (Y-W)**:

$$\begin{cases} R_1 = a_1 + a_2 R_1 + \ldots + a_p R_{p-1} \\ \quad \vdots \\ R_p = a_1 R_{p-1} + a_2 R_{p-2} + \ldots + a_p R_p \end{cases}$$

Dado los valores de $a_1, \ldots, a_p$, se puede resolver el sistema de ecuaciones para hallar $R_1, \ldots, R_p$

Finalmente, podemos usar estos valores para hallar $C_0 = a_1 C_1 + \ldots + a_p C_p + \sigma_e^2$ observando que $C_0 = a_1 R_1 C_0 + \ldots + a_p R_p C_0 + \sigma_e^2 \Rightarrow C_0 = \frac{\sigma_e^2}{1 - a_1 R_1 - \ldots - a_p R_p}$

Modelo ARMA

El modelo arma es una combinación de un proceso AR con un MA. Diremos que $\{Y_t\}$ sigue un modelo ARMA(p,q) si

$$Y_t = a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + e_t - b_1 e_{t-1} - \ldots - b_q e_{t-q}$$

Si se satisfacen las condiciones de estacionariedad, el modelo ARMA(p,q) puede reescribirse como un proceso lineal general con coeficientes $\psi_1, \psi_2, \ldots$ dados por:

$$\begin{cases} \psi_0 = 1 \\ \psi_1 = -b_1 + a_1 \\ \psi_2 = -b_2 + a_2 + a_1 \psi_1 \\ \vdots \\ \psi_j = -b_j + a_p \psi_{j-p} + a_{p-1} \psi_{j-p+1} + \ldots + a_1 \psi_{j-1} \end{cases} \qquad \psi_j = 0, si\ j < 0\ y\ b_j = 0\ si\ j > q$$

ARMA(p,q)

Se puede ver que la función de autocorrelación está dada por:

$$\begin{cases} C_0 = a_1 C_1 + a_2 C_2 + \ldots - \sigma_e^2 (b_0 + b_1 \psi_1 + \ldots + b_q \psi_q) \\ C_1 = a_1 C_0 + a_2 C_1 + \ldots + a_p C_{p-1} - \sigma e^2 (b_1 + b_2 \psi_1 + \ldots + b_1 \psi_{q-1}) \\ \quad \vdots \\ C_p = a_1 C_{p-1} + a_2 C_{p-2} + \ldots + a_p C_0 - \sigma_e^2 (b_p + b_{p+1} \psi_1 + \ldots + b_q \psi_{q-p}) \end{cases}$$

Si k>q entonces la expresión puede simplificarse como:

$$C_k = a_1 C_{k-1} + a_2 C_{k-2} + \ldots + a_p C_{k-p}$$

Cómo estimar los parámetros de un modelo AR?

En general no conocemos los valores de $a_1, \cdots a_p$ y debemos estimarlos a partir de las observaciones de la serie de tiempo.

Existen distintos enfoques:

1. Usando las ecuaciones de Yule-Walker
2. Basado en el método de máxima verosimilitud.
3. Cuadrados mínimos

Estimación del modelo AR mediante las ecs. de Y-W

Usando la expresión de las ecs. de Y-W, podemos reemplazar los valores de R_i por $\hat{R}_i$ y resolver el sistema de ecuaciones para $a_1, \ldots, a_p$:

$$\begin{cases} \hat{R}_1 = a_1 + a_2\hat{R}_1 + \ldots + a_p\hat{R}_{p-1} \\ \quad \vdots \\ \hat{R}_p = a_1\hat{R}_{p-1} + a_2\hat{R}_{p-2} + \ldots + a_p\hat{R}_p \end{cases}$$

Se puede demostrar que los estimadores de Y-W minimizan el ECM.

Una forma eficiente de hacer este cálculo es mediante el algoritmo de Levinson:

1. Set $\hat{\sigma}_0^2 = \hat{C}_0$ and $\mathrm{AIC}_0 = N(\log 2\pi\hat{\sigma}_0^2 + 1) + 2$
2. For $m = 1, \cdots, M$, repeat the following steps

   (a) $\hat{a}_m^m = \left(\hat{C}_m - \sum_{j=1}^{m-1} \hat{a}_j^{m-1}\hat{C}_{m-j}\right)\left(\hat{\sigma}_{m-1}^2\right)^{-1}$,

   (b) $\hat{a}_i^m = \hat{a}_i^{m-1} - \hat{a}_m^m \hat{a}_{m-i}^{m-1}$ for $i = 1, \ldots, m-1$,

   (c) $\hat{\sigma}_m^2 = \hat{\sigma}_{m-1}^2\{1 - (\hat{a}_m^m)^2\}$,

   (d) $\mathrm{AIC}_m = N(\log 2\pi\hat{\sigma}_m^2 + 1) + 2(m+1)$.

Estimación del modelo AR por MV

Bajo la suposición de que el ruido $e_i \sim \mathcal{N}(0, \sigma_e^2)$, la serie de tiempos

$$\{Y_1, \dots Y_n\} \sim \mathcal{N}(0, \Sigma),\ \Sigma = \begin{bmatrix} C_0 & C_1 & \dots & C_{n-1} \\ C_1 & C_0 & \dots & C_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n-1} & C_{n-2} & \dots & C_0 \end{bmatrix}$$

Hallar el EMV puedes costoso computacionalmente pues hay que calcular Σ^{-1}

Una alternativa es descomponer $L(\theta) = f_{Y_1,\dots,Y_n}(y_1, \dots, y_n; \theta) = \prod_{i=1}^{n} f(y_i | y_1, \dots, y_{i-1}; \theta)$

y usar un filtro de Kalman para hallar de forma eficiente cada término de la productoria.

Modelo de estados

Para las series de tiempo, un modelo de estados está dado por

$$x_t = F_t x_{t-1} + G_t e_t$$
$$y_t = H_t x_t + w_t$$

Les recuerda a algo?

donde xn es un vector **no** observable (estado), $e_t \sim \mathcal{N}(0, Q_t)$ (blanco) es el ruido de del sistema, y $w_t \sim \mathcal{N}(0, R_t)$ (blanco) es el ruido de medición.

Filtro de Kalman

$$\hat{x}_{t|t-1} = F_{t-1}\hat{x}_{t-1|t-1}$$
$$P_{t|t-1} = F_{t-1}P_{t-1|t-1}A_{t-1}^T + B_{t-1}Q_{t-1}B_{t-1}^T$$
$$K_t = P_{t|t-1}H_t^T(R_k + H_kP_{t|t-1}H_t^T)^{-1}$$
$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - H_t\hat{x}_{t|t-1})$$
$$P_{t|t} = (I - K_tH_t)P_{t|t-1}$$

Donde $\hat{x}_{t|t-1} = \mathbb{E}[x_t Y_{t-1}]$, $\hat{x}_{t|t} = \mathbb{E}[x_t Y_t]$, $P_{t|t-1} = \mathbb{E}[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})^T]$ y

$P_{t|t} = \mathbb{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^T]$

Modelo de estados y MV

Recuerdo que puedo descomponer la función de densidad conjunta como

$$\begin{aligned} f(y_1,\ldots,y_t) &= f(y_t|y_{t-1},\ldots,y_1)f(y_{t-1},\ldots,y_1) \\ &= f(y_t|y_{t-1},\ldots,y_1)f(y_{t-1}|y_{t-2},\ldots,y_1)f(y_{t-2},\ldots,y_1) \\ &\vdots \\ &= f(y_t|y_{t-1},\ldots,y_1)f(y_{t-1}|y_{t-2},\ldots,y_1)\ldots f(y_2|y_1)f(y_1) \end{aligned}$$

Luego,

$$\begin{aligned} L(\theta) &= f(y_t|y_{t-1},\ldots,y_1;\theta)f(y_{t-1}|y_{t-2},\ldots,y_1;\theta)\ldots f(y_2|y_1;\theta)f(y_1;\theta) \\ &= f(y_t|Y_{t-1};\theta)f(y_{t-1}|Y_{t-2};\theta)\ldots f(y_2|y_1;\theta)f(y_1;\theta) \end{aligned}$$

y

$$\log(L(\theta)) = \textstyle\sum_{i=1}^{n} \log(f(y_i|Y_{i-1};\theta))$$

Modelos de estados y MV

Además, dado que el ruido $e_t \sim \mathcal{N}(0, Q_t)$, las distribuciones condicionales son de la forma

$$y_t | Y_{t-1} \sim \mathcal{N}(y_{n|n-1}, d_{n|n-1})$$

donde

$$y_{t|t-1} = H_t \hat{x}_{t|t-1} \text{ y } d_{n|n-1} = H_t P_{t|t-1} H_t^T + Q_t$$

Luego

$$\log(L(\theta)) \propto \sum_{i=1}^{n} d_{t|t-1} + \sum_{i=1}^{n} (y_t - y_{t|t-1})^T d_{t|t-1}^{n} (y_t - y_{t|t-1})$$

Cómo usamos el KF + MV para estimar los parámetros

Básicamente los pasos a seguir son:

1. Asignar un valor para θ
2. Dado θ, calcular las matrices del filtro de Kalman
3. Asignar los valores iniciales para $x_{0|0} = 0, \text{ y } P_{0|0}$
4. Para t=1,...,n evolucionar el FK y obtener $x_{t|t-1}, \text{ y } P_{t|t-1}$
5. Con estos valores hallar la expresión de MV
6. Actualizar θ y volver a 1.

Fuente: http://www.kris-nimark.net/pdf/Handout_S4.pdf

Modelo AR como modelo de estados

Si definimos $x_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})^T$, nos queda que

$$F = \begin{bmatrix} a_1 & a_2 & \ldots & a_p \\ 1 & 0 & \ldots & 0 \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Además, considerando $H = [1 \quad 0 \quad \ldots \quad 0]$, recuperamos $y_t = Hx_t$. Fijando $Q = \sigma^2$ y $R = 0$ obtenemos el modelo de estados del modelo AR.

Modelo ARMA como modelo de estados

Para el modelo ARMA vamos a definir $x_t = [y_t, y_{t-1}, \ldots, y_{t-p+1}]^T$ al igual que en el modelo AR. Las matric F y H también va a ser las mismas:

$$F = \begin{bmatrix} a_1 & a_2 & \ldots & a_p \\ 1 & 0 & \ldots & 0 \\ & \ddots & & \\ & & 1 & 0 \end{bmatrix} \qquad H = [1\ 0\ \ldots\ 0]$$

Lo que va a cambiar en el modelo ARMA es la matriz G, que debe incluir la dependencia con ruidos anteriores:

$$G = \begin{bmatrix} 1 & b_1 & \ldots & b_q \\ 0 & 0 & \ldots & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

¿Cómo estimar el orden del modelo?

Usando la función de autocorrelación y la función de autocorrelación parcial.

Si k>p, la autocorrelación parcial es nula

Si k>q la autocorrelación es nula

Función de autocorrelación parcial: define el efecto que tiene Y(t-k) sobre Y(t)

$$\alpha(1) = corr(Y_{t+1}, Y_t), \text{ for } k = 1$$

$$\alpha(k) = corr((Y_{t+k} - P_{t,k}(Y_{t+k}), Y_t - P_{t,k}(Y_t)), \text{ para } k \geq 2$$

donde $P_{t,k}(x)$ es el operador de proyección ortogonal de x sobre el espacio generado por $Y_{t+1}, \ldots, Y_{t+k-1}$