Pronóstico del Precio de Acciones Empleando Modelos Ocultos de Markov

Congreso Interamericano de Estadística 2017

Luis Damiano

damiano.luis@gmail.com

https://luisdamiano.github.io

20 de octubre de 2017

Modelos Ocultos de Markov

Especificación (1)

Representa $\mathbf{x}_t$ (observaciones, emisiones o salidas) mediante dos modelos interconectados.

- **Estados latentes**: Cadena de Markov para tiempo discreto y estados discretos $z_t \in \{1, \dots, K\}$. Transición según $p(z_t|z_{t-1})$.
- **Observaciones**: Densidad $p(\mathbf{x}_t|z_t)$. En el caso Gaussiano, $p(\mathbf{x}_t|z_t = k, \theta) = \mathcal{N}(\mathbf{x}_t|\mu_k, \sigma_k)$.

La función de densidad conjunta está dada por

$$p(\mathbf{z}_{1:T}, \mathbf{x}_{1:T}) = p(\mathbf{z}_{1:T}) p(\mathbf{x}_{1:T}|\mathbf{z}_{1:T}) = \left[p(z_1) \prod_{t=2}^{T} p(z_t|z_{t-1}) \right] \left[\prod_{t=1}^{T} p(\mathbf{x}_t|z_t) \right].$$

Especificación (2)

Cantidades:

- **No estocásticas**: la secuencia de observaciones $\mathbf{x}_t$, el largo de la secuencia T y la cantidad de estados ocultos K.
- **Parámetros**: $\theta = (\pi_1, \theta_h, \theta_o)$
    - Distribución del estado inicial: vector de dimensión K π_1.
    - Parámetros del modelo oculto: matriz $K \times K$ de transición $\theta_h = \mathbf{A} = \{a_{ij} = p(z_t = j | z_{t-1} = i)\}$.
    - Parámetros del modelo observacional: $\theta_o = \{\mu_k, \sigma_k\}, \; k \in \{1, \ldots, K\}$.

Modelo generativo

A los fines de generar un conjunto de datos $\mathbf{x}_t^{(0)}$:

1. Generar los parámetros en función de las densidades a priori $\theta^{(0)} \sim p(\theta)$.
2. Generar el camino oculto $\mathbf{z}_{1:T}^{(0)}$ en función de los $K^2 - 1$ parámetros libres.
3. Generar las observaciones en función de la densidad muestral $\mathbf{x}_t^{(0)} \sim p(\mathbf{x}_t | \mathbf{z}_{1:T}^{(0)}, \theta^{(0)})$.

Existe un gran número de cantidades ocultas y algoritmos de inferencia asociados.[1]

Table 1: Resumen de las cantidades ocultas y sus correspondientes algoritmos de inferencia. *Cuando va de izquierda hacia derecha. **Cuando la matriz es dispersa.

Nombre	Cantidad oculta	Disponibilidad	Algoritmo	Complejidad
Filtering	$p(z_t \| \mathbf{x}_{1:t})$	t (online)	Forward	$O(K^2T)$ $O(KT)$ *
Smoothing	$p(z_t \| \mathbf{x}_{1:T})$	T (offline)	Forward-backward	$O(K^2T)$ $O(KT)$ *
Fixed lag smoothing	$p(z_{t-\ell} \| \mathbf{x}_{1:t})$, $\ell \geq 1$	$t + \ell$ (lagged)		
State prediction	$p(z_{t+h} \| \mathbf{x}_{1:t})$, $h \geq 1$	t		
Observation prediction	$p(x_{t+h} \| \mathbf{x}_{1:t})$, $h \geq 1$	t		
MAP Estimation	$p(z_t \| \mathbf{x}_{1:T})$	T	Viterbi encoding	$O(K^2T)$ $O(KT)$ **
Prob. of the evidence	$p(\mathbf{x}_{1:T})$	T	Forward	$O(K^2T)$ $O(KT)$ *

[1] Leonard E Baum and Petrie (1966); Leonard E. Baum and Eagon (1967); Leonard E Baum and Sell (1968); Leonard E. Baum et al. (1970); Leonard E. Baum (1972).

Input-Output HMM (Bengio and Frasconi 1995)

- Mapea una sequencia de entradas (señal de control) hacia una secuencia de salida.
- Probabilístico.
- Admite la producción, clasificación y predicción de datos.
- A diferencia de HMM (aprendizaje no supervisado), no aprende la distribución de las salidas sino la secuencia de salidas en sí misma.
- Especialmente efectivo para modelar memoria entre subconjuntos de una misma serie con grandes separaciones de tiempo.

$$
\begin{aligned}
z_t &= f(z_{t-1}, \mathbf{u}_t) \\
\mathbf{x}_t &= g(z_t, \mathbf{u}_t).
\end{aligned}
$$

$$p(z_t | \mathbf{x}_t, \mathbf{u}_t, z_{t-1} = i) = \mathrm{softmax}^{-1}(\mathbf{u}_t \mathbf{w}_i),$$

Aplicación

Datos

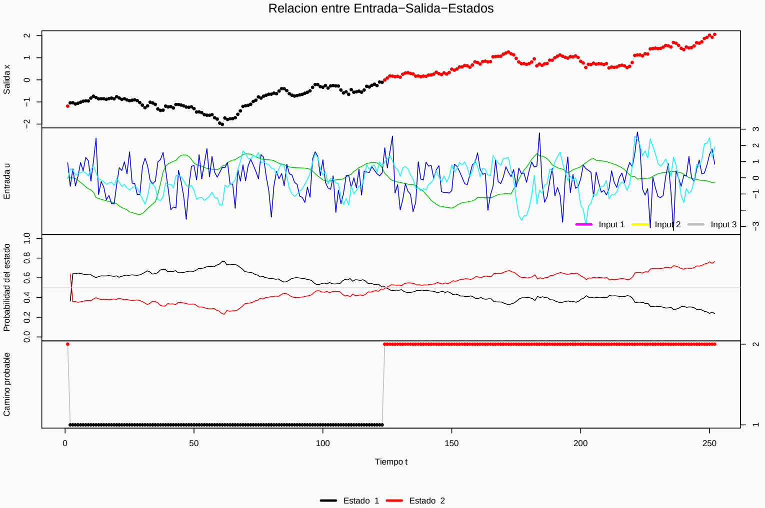
- BCBA:GGAL.
- Desde 2012-01-01 hasta 2017-09-30 (412 observaciones).
- Output: Precio de cierre.
- Inputs (lag $= 1$):
  - Tendencia en el precio de cierre.
  - Tendencia en el volumen negociado.
  - Volatilidad Parkinson (1980).

Modelo

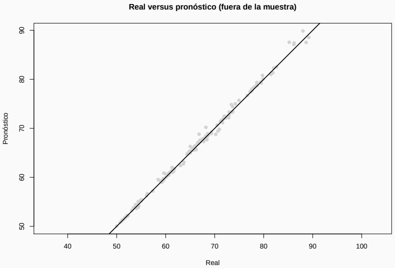
- Feature engineering: medias móviles, medias móviles exponenciales y ponderadas, transformación logarítmica, omisión de datos faltantes.
- Estados ocultos: $K = 2$.
- Estructura jerárquica en las medias (facilita convergencia).
- Prioris[2]: débilmente informativas (facilita convergencia).
- Walk-forward: pronóstico un paso hacia adelante para 160 ventanas rodantes con 252 observaciones cada una.

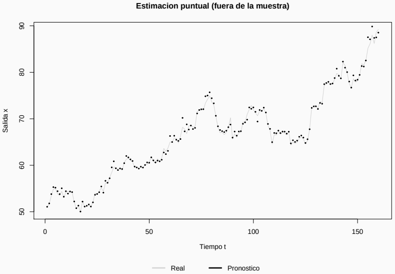
[2] Se recomienda la lectura de Prior Choice Recommendations para una guía informal pero muy práctica sobre la elección de densidades a priori.

Análisis exploratorio



Pronósticos





Estadístico	Valor
1 Razón aciertos (negativo)	0.8939
2 Razón aciertos (positivo)	0.9560
3 Razón aciertos (ambos signos)	0.9182
4 RMSE	0.5566
5 MAE	0.3888
6 MAPE	0.0057
7 R^2	0.9967

Conclusiones

- Enfoque de Machine Learning: foco en la predicción.
- Las variables de entrada incorporan información en la detección de los estados latentes.
- Los inputs admiten estados latentes persistentes, una característica difícil de observar en los HMM.
- Especialmente útil para series con memoria en observaciones muy distanciadas (ej. crisis).

¿Por qué Stan?

Stan (Carpenter et al. 2016)

- Monte Carlo Hamiltoniano.
- Muy rápido como algoritmo y además como implementación.
- Pensado para parámetros muy correlacionados (modelos jerárquicos).
- Gran conjunto de herramientas de diagnóstico.
- Excelente soporte de los desarrolladores.
- Conclusión: la vanguardia.

Agradecimientos

- Equipo de Desarrolladores de Stan (Aaron Goodman, Ben Bales y Bob Carpenter por sus opiniones en 1 y 2).
- Brian Peterson y Michael Weylandt por ser mentores del proyecto Bayesian Hierarchical Hidden Markov Models applied to financial time series bajo el programa Google Summer of Code 2017.

Referencias

Baum, Leonard E, and Ted Petrie. 1966. "Statistical Inference for Probabilistic Functions of Finite State Markov Chains." *The Annals of Mathematical Statistics* 37 (6). JSTOR: 1554–63.

Baum, Leonard E, and George Sell. 1968. "Growth Transformations for Functions on Manifolds." *Pacific Journal of Mathematics* 27 (2). Mathematical Sciences Publishers: 211–27.

Baum, Leonard E. 1972. "An Inequality and Associated Maximaization Technique in Stattistical Estimation for Probablistic Functions of Markov Process." *Inequalities* 3: 1–8.

Baum, Leonard E., and J. A. Eagon. 1967. "An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology." *Bulletin of the American Mathematical Society* 73 (3). American Mathematical Society (AMS): 360–64. doi:10.1090/s0002-9904-1967-11751-8.

Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. 1970. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *The Annals of Mathematical Statistics* 41 (1). Institute of Mathematical Statistics: 164–71. doi:10.1214/aoms/1177697196.

Bengio, Yoshua, and Paolo Frasconi. 1995. "An Input Output Hmm Architecture."

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2016. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 20.

Parkinson, Michael. 1980. "The Extreme Value Method for Estimating the Variance of the Rate of Return." *Journal of Business*. JSTOR, 61–65.