

MACHINE LEARNING & BIG DATA

Fundamentos de Series Temporales

Series Temporales

JOSÉ NELSON ZEPEDA DOÑO

Cluster de Estudio: Advanced Analytics

Este material es el resumen de muchos autores que por medio de sus libros y documentos nos ofrecen fuentes riquísimas de conocimiento sobre los temas de Big Data y Machine Learning.

Algunas citas, figuras y tablas pueden ser encontradas de forma textual tal como lo indica el autor en su material original.

Nelson Zepeda

MIP • V 1.0

San Salvador El Salvador

Phone 503 79074137 • @nelsonzepeda733

Tabla de Contenido

Conceptos de Series Temporales	1
Objetivo de una Serie Temporal	2
Elementos de una Serie Temporal.....	3
La Tendencia	5
Modelos	8
Modelos Aditivos y Multiplicativos.	9
Métricas de Error.....	9
Bibliografía	10

Conceptos de Series Temporales

El tiempo avanza de manera inexorable y no hay nada que podamos hacer para modificar su curso.

Una serie temporal o cronológica es una secuencia de datos, observaciones o valores, medidos en determinados momentos y ordenados cronológicamente en donde los datos pueden estar espaciados a intervalos iguales o desiguales¹.

Dicho lo anterior, es correcto afirmar que toda serie temporal refleja el comportamiento de una variable en el tiempo, idealmente suponemos que las observaciones se toman en intervalos regulares de tiempo y que no faltan observaciones.

Por otro lado, es fundamental tener en cuenta que el orden de los datos es de total importancia y puesto que las observaciones no son independientes podemos concluir que el pasado puede afectar al futuro.

En resumen y antes de proseguir con mayores detalles, se debe tener en consideración lo siguiente²:

- El orden de los datos importa.
- Las observaciones no son independientes.
- Al estimar relaciones se debe tener en cuenta que no son independientes.
- El tiempo tiene escala y esta escala está ordenada.
- Por tanto, debe utilizar técnicas matemáticas y estadísticas diferentes.
- Una serie temporal permite:
 - Entender el pasado

¹ https://es.wikipedia.org/wiki/Serie_temporal

² <https://economipedia.com/definiciones/serie-temporal.html>

- Entender la situación actual
- Entender el futuro

Objetivo de una Serie Temporal

El objetivo del análisis de series temporales es doble. Por un lado, se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Por otra parte, si se consigue definir ese patrón o modelo, se intentara predecir el comportamiento futuro de la misma.

Para alcanzar este doble objetivo se utiliza una metodología bastante consolidada, según la cual se admite que la serie temporal es una función del tiempo. Bajo este esquema, la serie sería una variable dependiente y el tiempo una variable independiente o explicativa. Dejando claro que el tiempo, en sí, no es una variable explicativa, sino simplemente un soporte o escenario en el que se realiza o tiene lugar una serie temporal. A esta forma de abordar el estudio de las series de tiempo se le conoce como enfoque clásico, frente al causal, según el cual, cualquier serie temporal, como variable que es, puede ser explicada por otra u otras series³.

Ahora bien, para poder explicar las variaciones en la serie, es necesario recalcar que las observaciones no son independientes. Una variable presenta dependencia temporal cuando los datos del pasado afectan al valor de la variable en el futuro. Por ejemplo, el Producto Interior Bruto (PIB) mundial a largo plazo tienen una tendencia alcista prolongada. Lo cual quiere decir que el crecimiento económico es sostenido en el tiempo. Por tanto, lo que ocurrió en el pasado tiene un efecto en el futuro. Al contrario, si tiramos un dado y apuntamos la fecha en la que lo tiramos, veremos que no existe relación ninguna entre los datos pasados y los presentes. En este último caso el pasado no afecta al futuro⁴.

La forma más sencilla de iniciar el análisis de una serie temporal es mediante su representación gráfica, para ello, en un sistema cartesiano, los valores de la serie Y_t se representan en el eje de las ordenadas y los periodos de tiempo en el eje de las abscisas. Mediante este tipo de representaciones se pueden detectar las características más sobresalientes de la serie, tales como el movimiento a largo plazo, la amplitud de las oscilaciones, la posible existencia de ciclos, los posibles puntos de ruptura, la presencia de valores atípicos, etc.

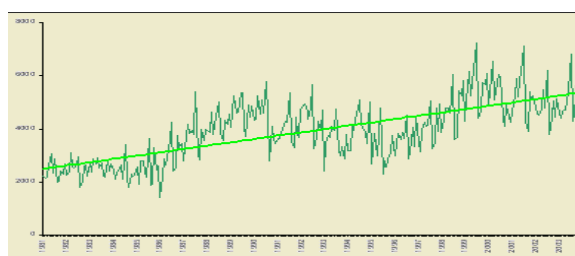


Figura 1-1 Serie Temporal

³ Tomado de Estadística Descriptiva: Series Temporales, Profesor Santiago de la Fuente Fernández

⁴ <https://economipedia.com/definiciones/serie-temporal.html>

Elementos de una Serie Temporal

Existen tres elementos básicos al momento de abordar/analizar una serie temporal: la tendencia, la estacionalidad y la aleatoriedad.

Las definiciones de cada componente son:

- Tendencia: Movimiento regular de la serie, a largo plazo. La Tendencia mide si temporalmente los valores tienen una direccionalidad hacia arriba o hacia abajo. En definitiva, capta una pendiente general de los valores. Una pendiente que puede ser positiva, si es de subida, o negativa, si es de bajada.
- Estacionalidad: Oscilaciones a corto plazo del periodo regular, mide la presencia de ciclos, de subidas y bajadas realizadas con una determinada regularidad.
- Aleatoriedad: Son fluctuaciones producidas por valores eventuales, esporádicos o imprevisibles, que no muestran una periodicidad previsible. En otras palabras, la aleatoriedad mide desvíos respecto de estos dos elementos vistos anteriormente, pequeños alejamientos de la tendencia o de la estacionalidad que se atribuirán a elementos no controlados en el modelo, a elementos incluso idiosincráticos, propios del individuo o los individuos evaluados en aquel momento.

Algunos autores agregan una cuarta componente: Componente Cíclica, la cual refleja comportamientos recurrentes, aunque no tienen por qué ser exactamente periódicos, con un periodo superior a un año. Muestran, habitualmente, cómo se suceden las etapas de bonanza económica con las de crisis, o al menos, desaceleración.

Veamos en forma gráfica la representación de los componentes.

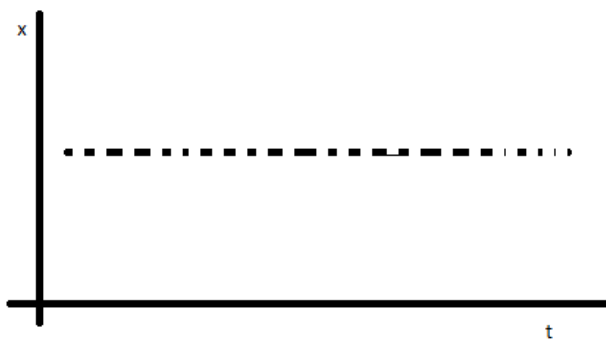


Fig 1-2 No existe tendencia, no existe estacionalidad ni aleatoriedad

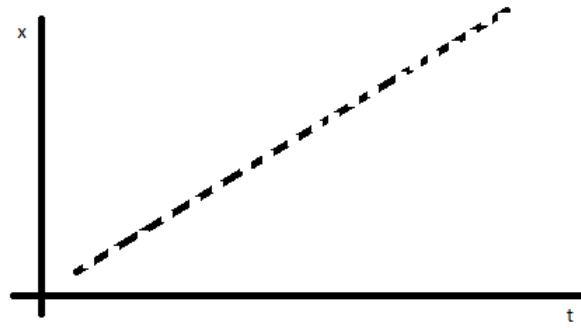


Fig 1-3 Solo existe tendencia

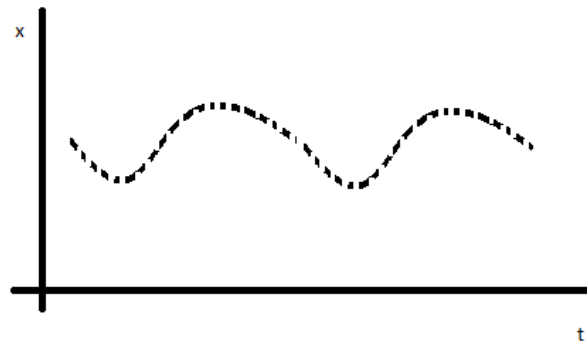


Fig 1-4 solo existe estacionalidad

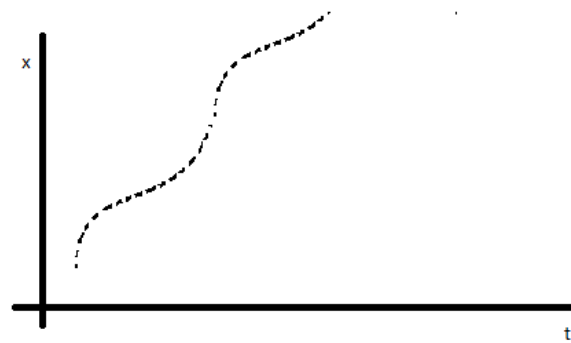


Fig 1-5 Muestra tendencia y estacionalidad

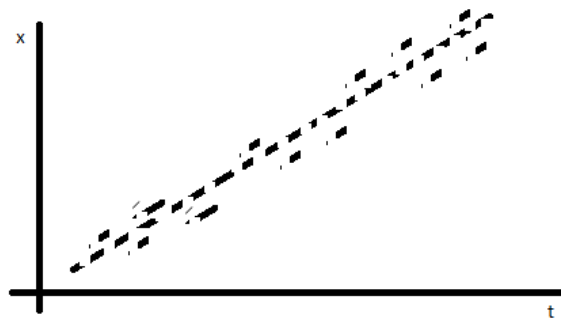


Fig 1-6 Muestra tendencia y aleatoriedad

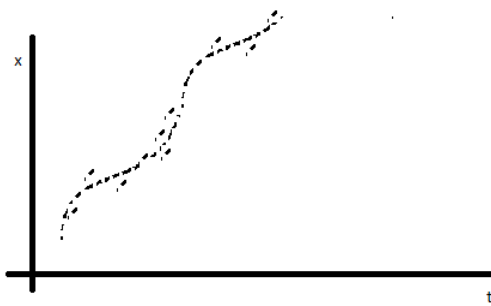


Fig 1-7 muestra tendencia, estacionalidad y aleatoriedad

La Tendencia de una serie se podrá evaluar mediante una Regresión lineal simple, a través del modelo: $X=at+b$, donde “a” es la pendiente de la recta y “b” la llamada “ordenada en el origen”. Ver el artículo dedicado a la Regresión lineal simple. Se trata, pues, de ver si, en su conjunto, entre los datos de la variable X estudiada y los valores temporales se podría ajustar una recta de regresión significativa. Evidentemente, esa Regresión lineal no nos servirá para crear una modelo general de la serie temporal pero sí para detectar y caracterizar una tendencia.

La Estacionalidad se evalúa mediante el llamado Correlograma. El correlograma consiste en un cálculo de correlaciones entre la misma muestra pero con diferentes desfases temporales. Es un original método de captar estacionalidad. Por ejemplo, si en datos mensuales se detecta una importante correlación entre los valores de cada mes con los del año siguiente hablaremos de una estacionalidad anual o de cada 12 meses.

La Tendencia

El cálculo de la tendencia requiere que tengamos una cantidad considerable de observaciones, de no ser así podrían obtenerse conclusiones erróneas.

Los métodos más utilizados para el cálculo de la tendencia son:

- Método gráfico

- Método del ajuste analítico
- Método de las medias móviles

Método Gráfico: Se trata de un método sencillo que permite obtener una línea de tendencia sin necesidad de realizar ningún cálculo.

El proceso consiste en la representación gráfica de la serie, uniendo mediante líneas los puntos altos que representan la serie, lo mismo se hace con los puntos bajos. De este modo aparecen dos líneas: la poligonal de cimas y la poligonal de fondos. Se unen luego los puntos medios de los segmentos que separan ambas poligonales obteniendo una línea mucho más suave que indica la dirección predominante o tendencia.

Este método presenta falta de objetividad, aunque en algunos casos puede resultar útil para analizar una ligera aproximación.

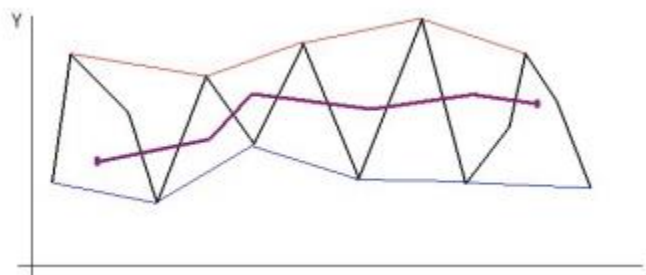


Figura 1-8 Método Gráfico

Método del Ajuste Analítico: Se realiza un ajuste por regresión de los valores de la serie a una función del tiempo que recoja de manera satisfactoria la marcha general del fenómeno representado por la serie temporal. Los tipos de ajuste más utilizados se muestran a continuación:

- Tendencia lineal: una tendencia lineal normalmente muestra que algo aumenta o disminuye a un ritmo constante. $Z(t) = a + bt$
- Tendencia logarítmica: es una línea curva muy útil cuando el índice de cambios aumenta o disminuye rápidamente y, después se estabiliza. $Z(t) = a + b \ln t$

⁵ <http://www.estadistica.net/Descriptiva/series-temporales-teoria.pdf>

⁶ <http://www.estadistica.net/Descriptiva/series-temporales-teoria.pdf>

- Tendencia Semi-logarítmica: es una variante de la tendencia logarítmica. $Z(t) = \ln(a + bt)$
- Tendencia Polinómica: es una línea curva que se utiliza cuando los datos fluctúan según la ecuación del polinomio. El orden del polinomio se puede determinar mediante el número de fluctuaciones en los datos, o en función de los máximos y mínimos que aparecen en la curva por ejemplo, una línea de tendencia polinómica de orden 2 suele tener solo un máximo o un mínimo.
 $Z(t) = a + bt + ct^2 + \dots + ct^n$
- Tendencia potencial: se utiliza con conjuntos de datos que comparan medidas que aumentan a un ritmo concreto. No es posible crear una tendencia de potencia si los datos contienen valores ceros o negativos.
 $Z(t) = at^b$
- Tendencia Exponencial: esta tendencia es muy útil cuando los valores de los datos aumentan o disminuyen a intervalos cada vez mayores, al igual que la potencial no admite valores cero o negativos.
 $Z(t) = ae^{bt}$
- Tendencia inversa o hiperbólica: es una línea curva
 $Z(t) = a + \frac{b}{t}$

Método de las medias móviles⁷: Se sustituye la serie original por una serie suavizada, que se toma como línea de tendencia.

El método de las medias móviles no sirve para hacer predicciones, dado que solo proporciona el valor de la tendencia en el intervalo de tiempo para el que se disponen los datos de la serie (excepto los valores que se pierden al inicio y al final de promediar), no para momentos futuros.

Dada una serie temporal $Y_{it}, t \equiv (t_1, t_2, \dots, t_n), i = 1, 2, \dots, k$, el método para suavizar la serie y determinar la tendencia, consiste en promediar cada valor con algunas de las observaciones que le preceden y le siguen.

El método consiste en sustituir cada Y_t por la media móvil \bar{Y}_t , la longitud K de la media móvil viene determinada por el número de sub periodos (Trimestres, cuatrimestres, semestres, etc) con lo que se eliminan las variaciones estacionales y accidentales (Si se dispone de gran cantidad de datos se puede utilizar periodos más grandes para intentar eliminar el ciclo).

Se diferencian dos casos para K :

⁷ <http://www.estadistica.net/Descriptiva/series-temporales-teoria.pdf>

- K Impar: todos los subíndices de las medias móviles serán números enteros y en consecuencia la serie de las medias móviles estará centrada, se pierden K-1 datos, la mitad al principio y la otra mitad al final.
- K es par: los subíndices no serán siempre enteros, y por tanto la serie no estará centrada. En este caso no es necesario centrarla, para lo que se calcula la media aritmética entre dos valores consecutivos de las medias móviles. Se pierden K datos.

Tal como se indicó antes, el mayor inconveniente que presenta el método de las medias móviles es que no permite hacer predicciones puesto que con el no se obtiene la expresión de una fórmula matemática que permita obtener el valor de la tendencia para un instante futuro.

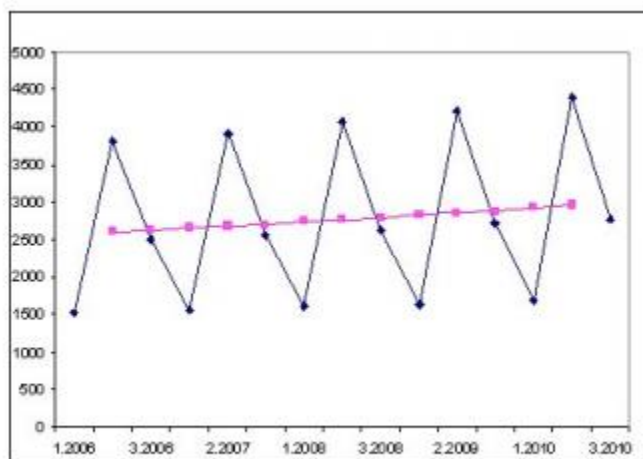


Fig 1-9 Tendencia con medias móviles.

Modelos

Existen cuatro modelos principales para series de tiempo, se trata de modelos que no son independientes sino que están contruidos de piezas para explicar situaciones progresivamente más complejas.

- AR: Modelos Autorregresivos, son modelos donde el valor de la variable X se puede poner en función de valores de la misma X pero anteriores en la serie.
- MA: Son modelos donde el valor de la variable X se puede poner en función no de valores anteriores de la variable X sino de los errores introducidos y descontrolados. Se fundamenta en el promedio móvil pero considerando el ruido blanco.
- ARMA: Son modelos donde conviene juntar un modelo AR con un modelo MA

- **ARIMA:** Son modelos no estacionarios, que tienen una tendencia y que conviene explicar esa tendencia. Son la fusión de un modelo ARMA con una modelización de la tendencia a través de un proceso de diferenciación.

Modelos Aditivos y Multiplicativos.

- **Modelo Aditivo:** Modelo de datos en el cual los efectos de factores individuales son diferenciados y agregados de manera conjunta para modelar los datos.
- **Modelo Multiplicativo:** Este modelo presupone que a medida que se incrementan los datos, también se incrementa el patrón estacional.

Se elige el modelo multiplicativo cuando la magnitud del patrón estacional en los datos depende de la magnitud de los datos.

En otras palabras, la magnitud del patrón estacional aumenta a medida que los valores de los datos se incrementan y disminuye a medida que los valores de los datos decrecen.

Se elige el modelo aditivo cuando la magnitud del patrón estacional en los datos no depende de la magnitud de los datos, es decir, la magnitud del patrón de estación no cambia cuando la serie sube o baja.

Métricas de Error

- **Bias:** Mean Forecast Error
- **MAE:** Mean Absolute Error
- **MSE:** Mean Squared Error
- **RMSE:** Root Mean Squared Error

Bibliografía

- Machine Learning for Beginners
By Ken Richards, 2017
- R Data Analysis Cookbook
by Kuntal Ganguly, 2017
- Estadística Descriptiva: Series Temporales
by Santiago de la Fuente Fernández