Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ciencias de la Computación

Semestre Primavera 2008
Tópicos en Minería de Datos

Series de Tiempo

Nombre: *Gonzalo Ríos*
Profesor: *Carlos Hurtado*
Fecha: *14 de Noviembre de 2008*

Índice de Contenidos

1 Definición Básica de Serie de Tiempo

Se llama Series de Tiempo a un conjunto de observaciones sobre valores que toma una variable (cuantitativa) en diferentes momentos del tiempo. Los datos se pueden comportar de diferentes formas a través del tiempo, puede que se presente una tendencia, un ciclo; no tener una forma definida o aleatoria, variaciones estacionales (anual, semestral, etc) [2]. Las observaciones de una serie de tiempo serán denotadas por Y_1,Y_2,...,Y_T , donde Y_t es el valor tomado por el proceso en el instante t. [3]

Los modelos de series de tiempo tienen un enfoque netamente predictivo y en ellos los pronósticos se elaborarán sólo con base al comportamiento pasado de la variable de interés. Podemos distinguir dos tipos de modelos de series de tiempo [1]:

- **Modelos deterministas**: se trata de métodos de extrapolación sencillos en los que no se hace referencia a las fuentes o naturaleza de la aleatoriedad subyacente en la serie. Su simplicidad relativa generalmente va acompañada de menor precisión. Ejemplo de modelos deterministas son los modelos de promedio móvil en los que se calcula el pronóstico de la variable a partir de un promedio de los "n" valores inmediatamente anteriores.
- **Modelos estocásticos**: se basan en la descripción simplificada del proceso aleatorio subyacente en la serie. En término sencillos, se asume que la serie observada Y_1, Y_2,...,Y_T se extrae de un grupo de variables aleatorias con una cierta distribución conjunta difícil de determinar, por lo que se construyen modelos aproximados que sean útiles para la generación de pronósticos.

La serie $\{Y_t\}_{t=1}^{T}$ podrá ser estacionaria o no estacionaria [1]:

- **Serie no estacionaria**: es aquella cuyas características de media, varianza y covarianza cambian a través del tiempo lo que dificulta su modelamiento. Sin embargo, en muchas ocasiones, si dicha serie es diferenciada una o más veces la serie resultante será estacionaria (procesos no estacionarios homogéneos).
- **Serie estacionaria**: es aquella cuya media y varianza no cambian a través del tiempo y cuya covarianza sólo es función del rezago. Gracias a estas características podremos modelar el proceso subyacente a través de una ecuación con coeficientes fijos estimados a partir de los datos pasados.
  - Media:$E(Y_t) = E(Y_{t+m})$ para todo t, m
    Varianza: $\sigma(Y_t) = \sigma(Y_{t+m})$ para todo t, m
    Covarianza: $cov(Y_t, Y_{t+k}) = cov(Y_{t+m}, Y_{t+m+k})$ para todo t, m, k

2 Aplicaciones de Series de Tiempo

Hoy en día diversas organizaciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar, prevenir,es decir, se utilizan para predecir lo que ocurrirá con una variable en el futuro a partir del comportamiento de esa variable en el pasado. En las organizaciones es de mucha utilidad en predicciones a corto y mediano plazo, por ejemplo ver que ocurriría con la demanda de un cierto producto, las ventas a futuro, decisiones sobre inventario, insumos, etc [2].

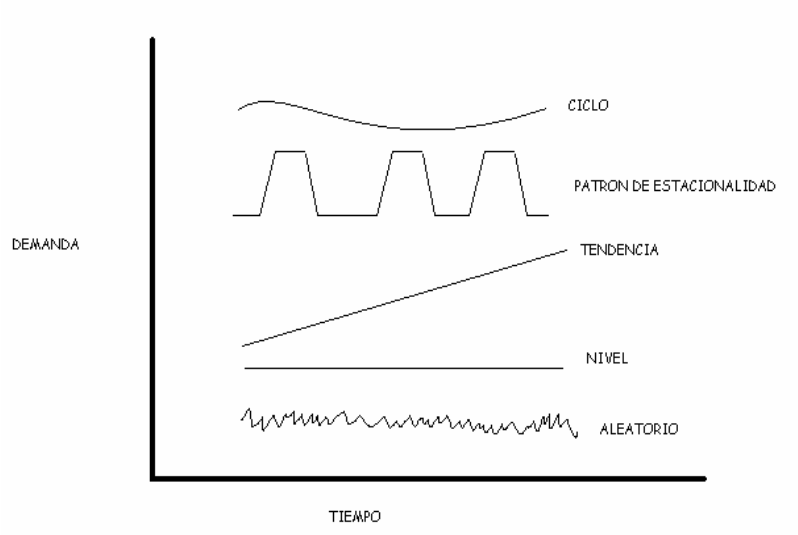
Algunas de las áreas de aplicación de Series de Tiempo son [3]:

- Economía: Precios de un articulo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.
- Meteorología: Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.
- Geofísica: Series sismológicas.
- Química: Viscosidad de un proceso, temperatura de un proceso.
- Demografía: Tasas de natalidad, tasas de mortalidad.
- Medicina: Electrocardiograma, electroencéfalograma.
- Marketing: Series de demanda, gastos, utilidades, ventas, ofertas.
- Telecomunicaciones: Análisis de señales.
- Transporte: Series de tráfico.

3 Componentes de una serie de tiempo: Enfoque clásico

Se dice que una serie de tiempo puede descomponerse en cuatro componentes (cinco si se considera una constante llamada nivel) que no son directamente observables, de los cuales únicamente se pueden obtener estimaciones. Estos cuatro componentes son [3,5]:

- **Tendencia (T)**: representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un extenso período de tiempo
- **Ciclo (C)**: caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros. Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años.
- **Estacionalidad (E)**: es un movimiento periódico que se producen dentro de un periodo corto y conocido. Este componente está determinado, por ejemplo, por factores institucionales y climáticos.
- **Aleatorio (A):** son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este componente es prácticamente impredecible. Este comportamiento representan todos los tipos de movimientos de una serie de tiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas.



Un modelo clásico de series de tiempo, supone que la serie $Y_1,...,Y_T$ puede ser expresada como suma o producto de sus componentes [3]:

- Modelo aditivo: $Y(t) = T(t)+E(t)+C(t)+A(t)$
- Modelo multiplicativo: $Y(t) = T(t)\times E(t)\times C(t)\times A(t)$

4 Aspectos Importantes en Series de Tiempo

4.1 Pronósticos dentro y fuera de la muestra

Al hablar de pronósticos, se distingue entre proyecciones dentro y fuera de muestra. En las primeras, las proyecciones realizadas se refieren a los mismos datos que se emplearon para la construcción y calibración del modelo (la muestra), mientras que en las segundas las proyecciones se refieren a datos ajenos a dicha muestra. En la búsqueda de metodologías que generen pronósticos precisos de los valores futuros de una variable, sólo son relevantes las proyecciones fuera de muestra por las siguientes razones:

- Las proyecciones fuera de muestra replican el funcionamiento de la herramienta de pronósticos en la práctica, por lo que la evaluación de su desempeño predictivo será un referente válido para los futuros errores de pronóstico.
- Los modelos de pronóstico se construyen minimizando los errores dentro de muestra por lo que los errores de pronósticos intramuestrales sobrestiman el potencial predictivo de las herramientas.
- Un modelo con buen desempeño intramuestral podría tener un muy mal desempeño en proyecciones fuera de muestra. Esto se debe a un sobreajuste (overfitting) o memorización de los datos muestrales, con lo que el modelo resultante será incapaz de responder de buena manera a nuevos valores. [1]

4.2 Pronósticos estáticos y dinámicos

Los pronósticos estáticos son aquellos que están basados en la última información efectiva disponible, por lo que están limitados a las proyecciones a un periodo hacia adelante. Los pronósticos dinámicos son caracterizados por utilizar el último pronóstico disponible como dato para el siguiente pronóstico, permitiendo la realización de proyecciones a dos y más periodos hacia delante. [1]

4.3 Alcance de los pronósticos y toma de decisiones

Todo pronóstico tiene asociado un alcance, pudiendo ser éste de corto, mediano o largo plazo. Los horizontes de tiempo correspondientes a dichos alcances dependerán de la industria bajo estudio. En cuanto al atractivo de uno u otro pronóstico, éste estará sujeto al tipo de decisión que se desea tomar o de acción en desarrollo. A modo de ejemplo, en la industria del cobre, alcances convencionales y decisiones comunes en el mercado y la industria son:

Tipo de pronóstico	**Alcance**	**Decisiones**
Cortísimo Plazo	Minutos, horas	Operaciones especulativas
Corto Plazo	Días, semanas, meses, un año	Operaciones especulativas, de cobertura y de gestión comercial
Mediano Plazo	Uno a seis años	Evaluación y control de los resultados de la gestión y de los negocios de una empresa
Largo Plazo	6 a 50 años	Planificación de la producción y evaluación de proyectos

Relacionado con el alcance de un pronóstico está su nivel de incertidumbre. A mayor alcance del pronóstico, mayor es el nivel de incertidumbre que se debe enfrentar. Esta consideración no debe olvidarse al momento de tomar decisiones basadas en datos proyectados.[1]

4.4 Conjuntos de entrenamiento y evaluación

Existen dos formas de evaluar la precisión de los pronósticos fuera de muestra:

- Esperar hasta que se cuente con los valores reales para los periodos pronosticados. Por ejemplo: si, durante el año 2006, se pronostica el precio del año siguiente, esperar hasta conocer el valor efectivo del año 2007.
- Evaluar la precisión sobre la base de un conjunto de datos que previamente se separó de la muestra disponible y que no participó de la construcción del modelo. Al conjunto de datos empleados para la construcción del

modelo se le denomina conjunto de entrenamiento, mientras que el resto de los datos conforma el conjunto de evaluación.

La división de los datos muestrales es una decisión trascendental en la generación de pronósticos ya que determina la cantidad de datos para la construcción del modelo y la

cantidad de pronósticos fuera de muestra que se podrán evaluar.

La definición del tamaño y composición de los conjuntos de entrenamiento y evaluación deberá considerar factores tales como:

- Tamaño total de la muestra: en muestras pequeñas, grandes conjuntos de evaluación podrían comprometer la calidad del modelo construido, si es que el conjunto de entrenamiento no consigue un tamaño que lo haga representativo.
- Tipo de metodología de pronóstico a emplear: distintas metodologías demandan conjuntos de entrenamiento más o menos numerosos.
- Representatividad: los componentes del conjunto de entrenamiento deben ser diversos para asegurar que el modelo pueda captar los diversos patrones de comportamiento de la serie bajo estudio (por ejemplo: precios en fases depresivas, precios en fases expansivas) [1]

4.5 Origen fijo versos origen móvil de los pronósticos

Dados los conjuntos de entrenamiento y evaluación, se define como el origen de los pronósticos al índice T correspondiente al último dato del conjunto de entrenamiento y se define como N al tamaño del conjunto de evaluación.

Los pronósticos de origen fijo, predicen la variable de interés a partir del dato T, esto es para los periodos $T+1, T+2, ..., T+N$. De este modo, para un origen fijo sólo se calcularán N pronósticos y sólo un pronóstico para cada alcance (un pronóstico a un periodo, un pronóstico a dos periodos, etc.), lo que es insuficiente para evaluar el desempeño de una metodología.

Por el contrario, en los pronósticos de origen móvil, se actualiza sucesivamente el origen de los pronósticos, lo que incrementa el número de proyecciones para cada alcance. Así en la situación recién descrita, una vez que se proyectaron los valores a partir de T, se calculan los pronósticos a partir de$T+1$ $(T+2, T+3, ..., T+N)$, a partir de $T+2$ $(T+3, T+4, ..., T+N)$ y así sucesivamente. El total de pronósticos calculados será:

$$N \times (N+1)/2$$

Alcance del pronóstico: $H \implies$ Número de evaluaciones: $N - H + 1$

Esta última relación entre el alcance de los pronósticos y el número de evaluaciones, nos permite dimensionar el tamaño absoluto del conjunto de evaluación. Sea H el máximo alcance de los pronósticos que se desea evaluar y sea M el número mínimo de evaluaciones que se desea realizar a dicho alcance, el tamaño del conjunto de evaluación estará dado por:

$$N \geq M + H - 1$$

Por otra parte, el uso de origen móvil disminuye la influencia de un determinado origen en los resultados (por ejemplo, fase depresiva de un ciclo económico). Las ventajas del origen móvil por sobre el origen fijo hacen que el origen móvil sea la técnica preferida en evaluaciones fuera de muestra.

El empleo de origen móvil plantea la posibilidad de reestimar el modelo de pronóstico en cada actualización. Este procedimiento es el más usado ya que disminuye la influencia del conjunto de entrenamiento original, aunque esto signifique un aumento de los cálculos necesarios.[1]

4.6 Conjunto de entrenamiento de tamaño creciente versus conjunto de entrenamiento de tamaño constante

Al utilizar la técnica de origen móvil, cada nueva evaluación significa la adición de un nuevo dato al conjunto de entrenamiento, por lo que se puede optar entre la realización de proyecciones sobre la base de un conjunto de

entrenamiento de tamaño creciente o de tamaño constante (fixed size rolling window). El uso de un conjunto de entrenamiento de tamaño constante implicaría que al agregar un nuevo dato, se descarte la observación más antigua (pruning), lo que parece recomendable si la trayectoria de precios a través del tiempo sigue un patrón notoriamente distinto al del pasado, situación que parece no aplicar al caso del cobre. [1]

4.7 Metodología de Box-Jenkins

El enfoque de Box-Jenkins es una de las metodologías de uso más amplio para el modelamiento estocástico de series de tiempo. Es popular debido a su generalidad, ya que puede manejar cualquier serie, estacionaria o no estacionaria, y por haber sido implementado en numerosos programas computacionales.

Los pasos básicos de la metodología de Box-Jenkins son [1]:

1. Verificar la estacionariedad de la serie. Si ésta no es estacionaria, diferenciarla hasta alcanzar estacionariedad.
2. Identificar un modelo tentativo.
3. Estimar el modelo.
4. Verificar el diagnóstico (si este no es adecuado, volver al paso 2).
5. Usar el modelo para pronosticar.

Lo que se trata es de identificar el proceso estocástico que ha generado los datos, estimar los parámetros que caracterizan dicho proceso, verificar que se cumplan las hipótesis que han permitido la estimación de dichos parámetros. Si dichos supuestos no se cumplieran, la fase de verificación sirve como retroalimentación para una nueva fase de identificación. Cuando se satisfagan las condiciones de partida, se puede utilizar el modelo para pronosticar.[5]

4.8 Estimación de parámetros

Para estimar los parámetros del modelo se utiliza un algoritmo de mínimos cuadrados de Gauss Marquatt para minimizar la suma de cuadrados de los residuos. Este algoritmo trata de minimizar la suma de cuadrados de los residuos, comenzando con algún valor de los parámetros del modelo. El algoritmo busca si otro vector de parámetros mejora el valor de la función objetivo y se produce un proceso de iteración hasta que se alcanza u cierto criterio de convergencia. [5]

5 Evaluación de Modelos de Series de Tiempo

5.1 Evalución del desempeño predictivo: Medición del error

Para la evaluación del desempeño predictivo se emplean diferentes indicadores que cuantifican qué tan cerca está la variable pronosticada de su serie de datos correspondiente. Una de las medidas más utilizadas es el Promedio del Error Porcentual Absoluto (MAPE)

$$MAPE = \frac{1}{T}\left(\sum_{t=1}^{T} APE_t\right) = \frac{1}{T}\left(\sum_{t=1}^{T} \frac{|Y_t^s - Y_t^a|}{Y_t^a}\right) \times 100$$

donde
APE :error porcentual absoluto.
Y_t^a :valor pronosticado de Y_t.
Y_t^s :valor real de Y_t.
T: número de periodos.
El MAPE mide el valor medio del error absoluto en términos porcentuales al valor real de la variable[1].

En lugar de considerar el promedio de error porcentual absoluto, MAX_MAPE indica el valor máximo del error del modelo respecto a la serie real, en términos porcentuales y absolutos [6] :

$$MAX\_MAPE = MAX_t \frac{|Y_t^s - Y_t^a|}{Y_t^a} \times 100$$

Para evaluar la dispersión de los errores se puede calcular el Desvío Estándar del Error porcentual absoluto (APE).

$$Desvío\ Estándar\ APE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(APE_t - MAPE)^2}$$

Otra medida del error de pronóstico comúnmente empleada es la Raíz Cuadrática Media del Error (RMSE):

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y_t^s - Y_t^a)^2}$$

donde
Y_t^a :valor pronosticado de Y_t.
Y_t^s :valor real de Y_t.
T: número de periodos.
El RMSE mide la dispersión de la variable simulada en el curso del tiempo, penalizando fuertemente los errores grandes al elevarlos al cuadrado. Esta característica hace que el RMSE se recomiende cuando el costo de cometer un error es aproximadamente proporcional al cuadrado de dicho error.

No siempre el modelo que genere pronósticos con un menor MAPE generará los pronósticos con el menor RMSE y viceversa, por lo que en la selección de los mejores modelos de pronóstico se hace necesario establecer la medida de error a utilizar para la elaboración del ranking de desempeño.

Dado que una mala estimación del precio fututo del cobre se traduce en una pérdida de ingresos proporcional al tamaño del error, el MAPE, y no el RMSE, parece ser la medida de desempeño más adecuada. A esto se suma la ventaja práctica del MAPE de no requerir ser acompañado por la media para dimensionar la magnitud del error. Luego, la medida de error que se empleará para identificar los modelos de mejor desempeño será el MAPE.[1]

6 Estimación de la Tendencia

Hay varios métodos para estimar la tendencia T(t), uno de ellos es utilizar un modelo de regresión lineal. Se pueden utilizar otros tipos de regresiones, como regresión cuadrática, logística, exponencial, entre otros.

Una forma de visualizar la tendencia, es mediante suavizamiento de la serie. La idea central es definir a partir de la serie observada una nueva serie que filtra o suaviza los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios), de manera que podamos visualizar la tendencia. [3]

6.1 Promedio Móvil

Este método de suavizamiento es uno de los más usados para describir la tendencia. Consiste en fijar un número k, preferentemente impar, como 3, 5, 7, etc., y calcular los promedios de todos los grupos de k términos consecutivos de la serie. Se obtiene una nueva serie suavizada por promedios móviles de orden k. De este modo se tienden a anular las variaciones aleatorias. La formula está dada por

$$Z(t) = \frac{Y(t-k) + Y(t-k+1) + ... + Y(t) + Y(t+1) + ... + Y(t+k)}{2 * k + 1}$$

El suavizamiento de media móvil es muy fácil de aplicar, permite visualizar la tendencia de la serie. Pero tiene dos inconvenientes: No es posible obtener estimaciones de la tendencia en extremos y no entrega un medio para hacer predicciones. Si la serie presenta un efecto estacional de período k, es conveniente aplicar un suavizamiento de media móvil de orden k. En tal caso se elimina el efecto estacional, junto con la variación aleatoria, observándose solamente la tendencia.[3]

6.2 Suavizamiento exponencial

Este modelo se basa en que una observación suavizada, en tiempo t, es un promedio ponderado entre el valor actual de la serie original y el valor de la serie suavizada, en el tiempo inmediatamente anterior. Si $Y(t)$ representa la serie de tiempo original, y $Z(t)$ la serie de tiempo suavizada, entonces lo anterior se puede escribir

$$Z(t) = \alpha Y(t) + (1 - \alpha) Z(t-1)$$

en donde α es un número entre 0 y 1.

Si α es cercano a 1, la serie suavizada pondera más fuertemente el valor original, luego ambas se parecen, y en consecuencia, el suavizamiento es poco.

Si α se acerca a 1/2, se ponderan moderadamente la serie original y la suavizada, por lo que el suavizamiento es moderado.

Si α es cercano a cero, (1-α) es cercano a 1, y la serie suavizada pondera más fuertemente el valor suavizado inmediatamente anterior, por lo que el suavizado es importante.

Consecuencia de la fórmula anterior es que la serie suavizada se puede expresar como

$$Z(t) = \alpha Y(t) + \alpha(1-\alpha) Y(t-1) + \alpha(1-\alpha)^2 Y(t-2) + ... + \alpha(1-\alpha)^{t-1} Y(1)$$

Es decir, cada término suavizado es un promedio ponderado de todos los términos históricos de la serie original.

Como α está entre 0 y 1, estos números se van achicando a medida que avanzan. Eso significa que a medida que nos alejamos hacia el pasado, los término van influyendo cada vez menos en el término presente. La rapidez con que disminuye la influencia es mayor mientras más grande (cercano a 1) es α.

Si la serie varía lentamente, por lo general se eligen valores de α cercanos a 0 (valor típico $\alpha = 0.3$). En cambio, si varía bruscamente, se eligen valores de α cercanos a 1 (valor típico $\alpha = 0.7$). [3]

7 Transformada Discreta de Fourier: Enfoque Espectral

La mayoría de los métodos en series de tiempo se basan en el espacio del tiempo. Otro enfoque muy poderoso es en el espacio de frecuencia, en donde la transformada discreta de fourier tiene un papel primordial.

7.1 Definicion Matemática

Sea $L_p^2(a) = \{f : \mathbb{R} \to \mathbb{C} \ \mid f \text{ es } a-\textit{periódica} \text{ y } \int_0^a |f(t)|^2 dt < \infty\}$. Sea $f \in L_p^2(a)$ y denotamos por $S[f]$ a su Serie de Fourier, la que viene dada por:

$$S[f](t) = \sum_{n=-\infty}^{n=\infty} c_n e^{2\pi i n \frac{t}{a}}$$

,donde los $\{c_n\}_{n\in\mathbb{N}}$ son sus "coeficientes de Fourier", que vienen dados por:

$$c_n = \frac{1}{a}\int_0^a f(t) e^{-2\pi i n \frac{t}{a}} dt$$

Ahora, sean $S^N[f]$ las sumas parciales de la serie anterior, esto es:

$$S^N[f](t) = \sum_{n=-N}^{n=N} c_n e^{2\pi i n \frac{t}{a}}$$

Es bien sabido que las sumas parciales de una función $f \in L_p^2(a)$ convergen a ella en la norma de $L^2(0,a)$.

Aún más, un teorema debido a Dirichlet señala que si además para un punto $t_o \in (0,a)$ los límites $f(t_o^+)$ y $f(t_o^-)$ existen, al igual que las derivadas laterales en ese punto, entonces:

$$S^N[f](t_o) \to S[f](t) = \frac{1}{2}[f(t_o^+) + f(t_o^-)]$$

Luego conocer la Serie de Fourier de una función suficientemente buena puede otorgar bastante información sobre esta. Sin embargo, uno precisaría conocer una infinidad de las constantes $\{c_n\}$(lo que a parte de ser costoso, salvo casos muy particulares, es imposible). En la práctica, no es necesario conocer "demasiadas" de estas constantes, por dos razones:

- La serie será evaluada numéricamente, por lo que se considerará solo una suma parcial de esta.
- Se tiene que para $f \in L_p^2(a)$, $\sum_{n=-\infty}^{n=\infty} |c_n|^2 < \infty$, de donde $c_n \to 0$ si $|n| \to \infty$. Luego los únicos término de interés (numérico) son los de índice no demasiado grande.

Pese a lo anterior, puede ser necesario conocer muchos c_n, lo que implicaría evaluar muchas integrales. Por lo tanto, para efectos prácticos, será necesario integrar numéricamente las expresiones para estos coeficientes.

Esta es la motivación original para la Transformada de Fourier Discreta (DFT), vale decir, obtener una expresión aproximada para la Serie de Fourier de una función a partir de un "sampleo" de datos conocidos sobre la función. Se verá más adelante que el espectro de problemas en que esta herramienta ha resultado útil es mucho más amplio.

7.1.1 Calculando los Coeficientes de Fourier

Suponer que se tiene una función $a-\textit{periódica}$ $f : \mathbb{R} \to \mathbb{C}$ sobre la cual solo conocemos N valores equiespaciados en $(0,a)$, vale decir:

$$f\left(k\frac{a}{N}\right) = y_k \ , \ k = 0, 1, ..., N-1$$

La idea es estimar N coeficientes de la Serie de Fourier de f, la que se asumirá converge puntualmente. Por simplicidad se supondrá que N es par. Así, se estimará c_n para $n = -N/2, ..., N/2-1$. De la fórmula para

los coeficientes de Fourier, se puede integrar mediante el **Método del Trapecio**, lo que entrega la siguiente aproximación para c_n :

$$c_n \approx Y_n := \frac{1}{N}\sum_{k=0}^{N-1} y_k w_N^{-nk} \text{ , con } w_N = e^{\frac{2\pi i}{N}},\ n \in \mathbb{N}$$

Además, de la periodicidad de la exponencial, se verifica que $\frac{1}{N}\sum_{k=0}^{N-1} y_k w_N^{-nk} = Y_{n+N}$, si $-\frac{N}{2} \leq n < 0$, de donde

$$c_n \approx Y_{j(n)} \text{ , con } j(n) = n + (N\ \boldsymbol{\chi}_{n<0}) \text{ , } n = -N/2, ..., N/2 - 1$$

y además se verifica que:

$$y_k = \sum_{n=0}^{N-1} Y_n w_N^{nk}, k = 0, .., N-1$$

Así, queda definido un isomorfismo:

$$F_N : \mathbb{C}^N \to \mathbb{C}^N, con F_N(\{y_k\}_{k=0}^{N-1}) = \{Y_n\}_{n=0}^{N-1}$$

, el que se llama la **Transformada de Fourier Discreta de orden N**.

7.2 Algunas Propiedades de F_N

Si $\{u_k\}_{k\in\mathbb{Z}} \subset \mathbb{C}$ satisface $u_n = u_{n+kN}\ \forall n, k \in \mathbb{Z}$, se dirá que es una secuencia $N-$*periódica*. Con esta definición, es claro que dado $\{y_k\}_{k=0}^{N-1}$ puede extenderse como una secuencia periódica, al igual que $\{Y\}_{k=0}^{N-1} = F_N(\{y_k\}_{k=0}^{N-1})$. Entre las muchas propiedades que aparecen, destacan:

1. Sean $\{x_k\}$ y $\{y_k\}$ dos secuencias N-periódicas, y sean $\{X_n\}$ y $\{Y_n\}$ las secuencias asociadas a sus transformadas discretas. Entonces:

- La secuencia definida por su **Convolución Circular,** $z_k = \sum_{q=0}^{N-1} x_q y_{k-q}$ con $k \in \mathbb{Z}$, tiene por transformada a $Z_n = NX_nY_n$ (producto término a término).
- La transformada de la secuencia $\{p_k = x_k y_k\}$ es $\left(P_n = \sum_{q=0}^{N-1} X_q Y_{n-q}\right)_n$

2. Si $(Y_n) = F_N(y_k)$, entonces $\sum\limits_{k=0}^{N-1} |y_k|^2 = N \sum\limits_{n=0}^{N-1} |Y_n|^2$

7.3 Relación entre los coeficientes de Fourier exactos y aproximados

Suponiendo que una función a-periódica f puede ser expresada como

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{2i\pi n \frac{t}{a}}$$

, y que esta serie es absolutamente convergente, para cada t es posible reordenar sus términos sin alterar su convergencia. Por ejemplo, se puede primero sumar sobre todos los índices iguales a 0 mod N, luego los iguales a 1 mod N, etc. Así, en particular:

$$f\left(k\frac{a}{N}\right) = y_k = \sum_{m=-\infty}^{\infty} c_m w_N^{mk} = \sum_{n=0}^{N-1}\left(\sum_{q=-\infty}^{\infty} c_{n+qN}\right) w_N^{nk}$$

De esto se deduce que $Y_n = \sum_{q=-\infty}^{\infty} c_{n+qN}$, y luego

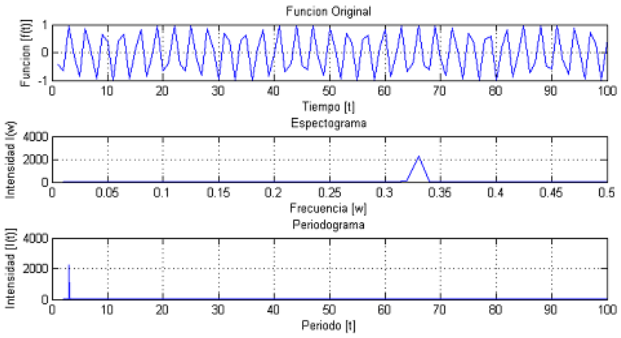
$$Y_n - c_n = \sum_{q \neq 0} c_{n+qN}$$

Con esto se obtiene que entre más fuertemente decaigan a cero los coeficientes c_n, mejor es la aproximación. Esto ocurre, por ejemplo, entre más suave sea la función.

7.4 Aplicación a Series de Tiempo

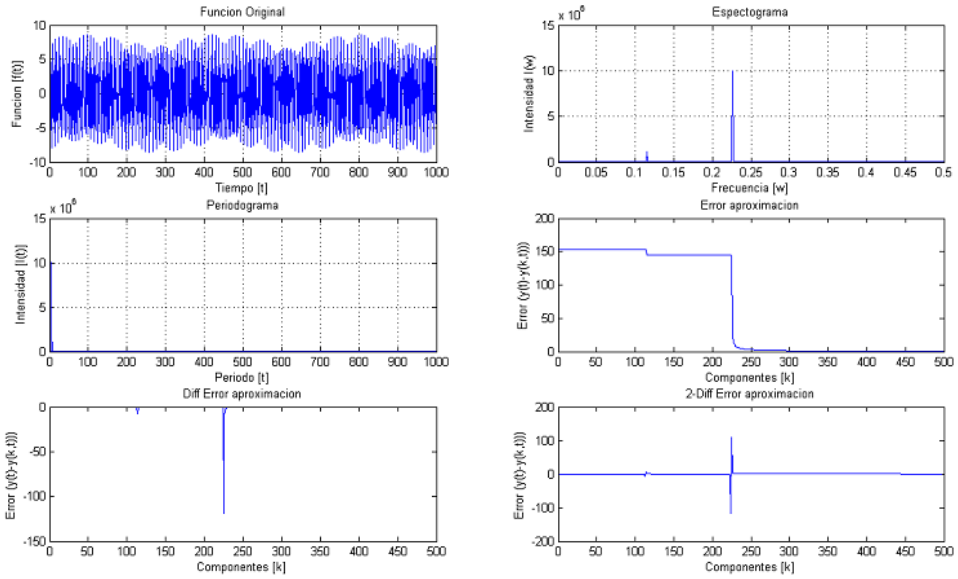
El uso de la transformada discreta de fourier en series de tiempo se basa en la idea de observar las frecuencias más importantes, y graficar la intensidad de cada frecuencia. Este gráfico se llama espectograma. Si se grafica la intensidad versus el periodo, que es el inverso de la frecuencia, entonces se obtiene el periodograma. Este gráfico nos mostrará la presencia de componentes estacionales importantes en la serie de tiempo, y así separar la componente estacionaria de la serie.

Por ejemplo, la transformada de fourier de la función y=cos(2*x) en el intervalo [0,100], con un paso igual a 1, obtenemos el siguiente gráfico:

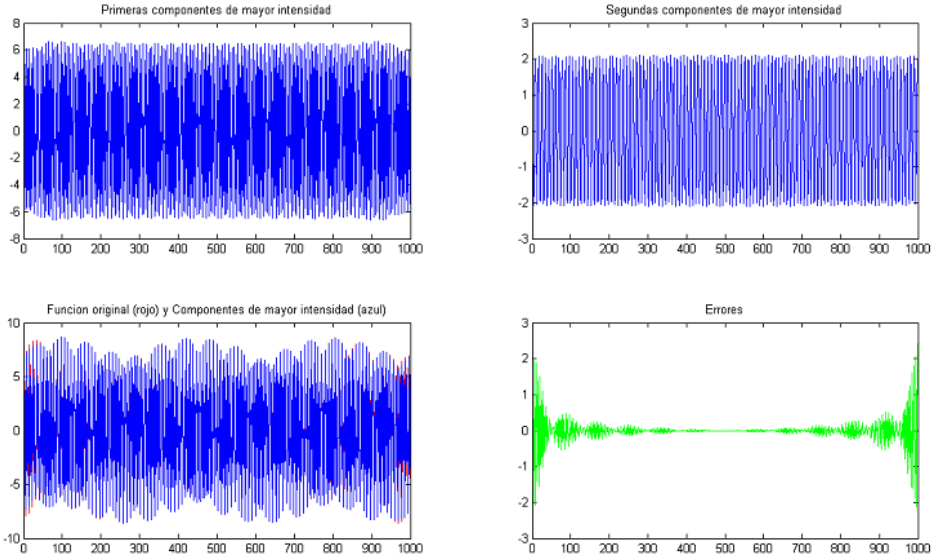


Podemos ver que la única frecuencia importante se encuentra adentro del intervalo [0.3,0.35], y si observamos que cos(2*x)=cos($\frac{2*\pi}{\pi}$x) y $\frac{1}{\pi} = 0.318$, luego podemos reconstruir la función con solo conocer este valor.

Otro gráfico interesante es el error de aproximación versus el número de coeficientes de la transformada de fourier que se ocupan para la aproximación. Lo que a nosotros nos interesa ver de este gráfico son los cambio bruscos, por lo que además se grafica su primera y segunda derivada. Por ejemplo:



Se puede observar que cerca de la frecuencia 0.225 hay una gran intensidad, y otra de menor magnitud alrededor de la frecuancia 0.115. Estas corresponden a las componentes número 226 y 115, respectivamente. Luego, si filtramos una vecindad de estas frecuencias y las demás las eliminamos, obtenemos el siguiente resultado:



Luego, podemos observar que recuperamos casi toda la información de la función original, excepto en los bordes, que es un fenómeno numérico que ocurre con la transformada de fourier. Con estas componentes aisladas, se pueden analizar los diferentes fenómenos estacionarios que la suma dan la serie original, por lo que se pueden identificar diferentes fuentes explicativas de la serie de tiempo.

8 Modelos ARIMA: Enfoque Moderno

Los modelos multivariantes o econométricos tratan de explicar el comportamiento de una o más variables en función de la evolución de otras variables que se consideran explicativas. Las variables explicadas por el modelo se denominan endógenas, mientras que las variables explicativas del modelo, pero no explicadas por él, se denominan predeterminadas. Entre las variables predeterminadas se distinguen dos grupos: exógenas y endógenas retardadas, estas últimas no son explicadas por el modelo en el momento t, pero han sido explicadas por él en un momento anterior, por su parte, las exógenas son variables que no son explicadas por el modelo en ningún momento.

Los modelos econométricos contemplan de forma explícita la información que aportan las variables causales del fenómeno de interés de acuerdo con una determinada teoría económica. Una ventaja de este modelo consiste en que los resultados que se generan son más eficientes y poseen mayor poder explicativo que los modelos univariantes. Sin embargo, en estos modelos, cuando se desea realizar predicciones, el desconocimiento de los valores de las variables explicativas en el futuro determina la necesidad de utilizar predicciones para éstas, lo cual incrementa el nivel de incertidumbre con que se realiza la predicción econométrica. Por otro parte, cuando el futuro puede suponer una alteración de tendencias de comportamiento respecto al pasado reciente, es recomendable utilizar estos modelos causales para predecir a mediano plazo (1 a 5 años).

Los modelos univariantes o de series de tiempo no necesitan conocer ninguna relación de causalidad, explicativa del comportamiento de la variable endógena, ni en su defecto, ninguna información relativa al comportamiento de otras variable explicativas, ya que en este caso no existe este tipo de variables. Es suficiente con conocer una serie temporal de la variable en estudio, para estimar el modelo que se utilizará para predecir.

La predicción univariante se utiliza, en problemas económicos, principalmente con dos objetivos:

- La predicción de algunas variables explicativas de un modelo causal, cuando se espera que en el futuro conserven algunas de las características de su evolución en el pasado.
- La predicción a corto plazo, debido a su gran capacidad para recoger la dinámica en el comportamiento de la variable estudiada. Además, en condiciones normales, cuando no existen bruscas alteraciones respecto a la experiencia reciente de la variable, estos métodos pueden proporcionar buenas predicciones.

Entre las técnicas univariantes existen algunas muy sencillas, tales como el modelo autorregresivo de primer orden, el modelo de tendencia lineal o exponencial, entre otros. Las técnicas más rigurosas para la predicción univariante son las denominadas técnicas o modelos Box-Jenkins, o más concretamente modelos ARIMA, pues las técnicas Box-Jenkins constituyen un conjunto más amplio, dentro del cual los modelos ARIMA univariantes son sólo una parte. [5]

En la generación de proyecciones de corto y mediano plazo, existe evidencia de un mejor desempeño de los modelos de series de tiempo. Llama la atención el que los modelos de series de tiempo sean más precisos en sus proyecciones que los modelos econométricos más complejos, algunos de los cuales tienen múltiples ecuaciones y decenas de variables. Entre las razones que explican este fenómeno está la dificultad asociada a la selección de las variables explicativas de un modelo estructural y la dificultad que conlleva el pronóstico de las mismas, problema que podría ser aún más difícil que el pronóstico de la variable de interés.[1]

8.1 Modelamiento de series no estacionarias

8.1.1 Caminata aleatoria

$$Y_t = \delta + Y_{t-1} + \varepsilon_t$$

donde, δ es una constante (drift) y cada perturbación ε_t (error) es una variable aleatoria con distribución normal con media cero, varianza constante y covarianza cero (el proceso $\varepsilon_1, \varepsilon_2, ...$ se denomina ruido blanco). En el modelo de caminata aleatoria más simple (sin drift), el pronóstico para Y_t es su valor más reciente. La inclusión de un drift intenta reproducir una tendencia existente en la variable de interés.

En un modelo de caminata aleatoria, la varianza de Y_t aumenta a través del tiempo, lo que es propio de un proceso no estacionario. Cuando una serie se comporta como una caminata aleatoria se dice que ésta presenta raíz unitaria. [1]

8.2 Modelamiento de series estacionarias

Box y Jenkins han desarrollado modelos estadísticos que tienen en cuenta la dependencia existente entre los datos. Cada observación en un momento dado es modelada en función de los valores anteriores. Se modela a través de ARIMA (Autorregresive Integrate Moving Average). Alguna de las características de este modelo son [2]:

- Tiene solamente en cuenta la pauta de serie serie de tiempo en el pasado.
- Ignora la información de variables causales.
- Procedimiento técnicamente sofisticado de predicción de una variable.
- Utiliza la observación más reciente como valor inicial.
- Permite examinar el modelo más adecuado
- Analiza errores recientes de pronósticos para seleccionar el ajuste apropiado para periodos futuros.
- Box-Jenkins es más apropiado para predicciones a largo plazo que para corto plazo.
- Extrae mucha información de la serie de tiempo, más que cualquier otro método.

8.2.1 Modelos de Media Móvil, MA(q)

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

En los modelos de media móvil, el proceso se representa como una suma ponderada de errores actuales y anteriores. El número de rezagos del error considerados (q) determina el orden del modelo de media móvil. [1]

8.2.2 Modelos Autorregresivos, AR(p)

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t$$

En los modelos autorregresivos, el proceso se representa como una suma ponderada de observaciones pasadas de la variable. El número de rezagos (p) determina el orden del modelo autorregresivo. [1]

8.2.3 Modelos Mixtos Autorregresivos – Media Móvil, ARMA(p,q)

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

En estos modelos, el proceso se representa en función de observaciones pasadas de la variable y de los valores actuales y rezagados del error. El número de rezagos de la variable de interés (p) y el número de rezagos del error (q) determinan el orden del modelo mixto. [1]

8.2.4 Modelos Autorregresivos Integrados de Promedio Móvil, ARIMA(p,d,q)

Muchas series de tiempo no son estacionarias, por ejemplo el Producto Nacional Bruto o la Producción Industrial. Un tipo especial de series no estacionarias, son las no estacionarias homogéneas que se caracterizan porque, al ser diferenciadas una o más veces, se vuelven estacionarias.

La serie Y_t será no estacionaria homogénea de orden d si $W_t = \Delta^d Y_t$ es estacionaria, donde:

- $\Delta Y_t = Y_t - Y_{t-1}$
- $\Delta^{n+1} Y_t = \Delta^n Y_t - \Delta^n Y_{t-1}$

Si después de haber diferenciado la serie Y_t se consigue una serie estacionaria W_t, y dicha serie obedece a un proceso ARMA(p,q), se dice que Y_t responde a un proceso ARIMA(p,d,q):

$$W_t = \delta + \phi_1 W_{t-1} + ... + \phi_p W_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

Para la correcta identificación del modelo ARIMA representativo de una serie se hace necesario [1]:

- **Determinar el grado de homogeneidad u orden de integración de la serie**

  Para determinar el orden de integración se utilizan herramientas como el correlograma y tests de raíz unitaria. Cabe mencionar que los tests de raíz unitaria, como el test de Dickey-Fuller Aumentado (ADF), tienen baja potencia presentando, ante situaciones de difícil discriminación, un sesgo al no rechazo de la hipótesis nula de presencia de raíz unitaria.

- **Determinar el orden de las partes de promedio móvil y autorregresivas del modelo**

  El examen de las funciones de autocorrelación total y parcial ayuda en esta tarea, aunque habitualmente la selección correcta no será clara, por lo que se recomienda probar distintas formulaciones guiándose por el conocimiento que se tenga del fenómeno analizado.

- **Evaluar los distintos modelos construidos**

  Se descartan las estructuras que arrojen coeficientes no significativos y/o que fueron mal evaluadas de acuerdo a indicadores como el Criterio de Información de Schwarz (SIC). Un buen modelo tendrá un buen ajuste (coeficiente de determinación cercano a la unidad) y arrojará residuos que se comportarán como ruido blanco.

8.3 Modelos ARIMA con variables de intervención

Existen los modelos ARIMA con variables de intervención, en los cuales las series económicas son afectadas por fenómenos externos, tales como cambios tecnológicos, huelgas, cambios en medidas de política o económicas, cambios en la legislación o escala de algún impuesto, cambios metodológicos en la medición de las estadísticas, etc. Estos fenómenos son llamados intervenciones ya que interfieren en el comportamiento original de la serie, por lo tanto se debe evaluar su efecto e incorporarlo al modelo ARIMA a través de variables artificiales binarias (análisis de intervención).

Se recurre a variables que explican la presencia de fenómenos exógenos en la serie de tiempo. Se incorporan como variables dummy en la forma de impulsos y escalones que se utilizan para representar cambios temporales o permanentes en el nivel de las series debidos a eventos especiales. La no-incorporación de variables artificiales conduce a sesgos en las estimaciones de los parámetros, a elevar el error estándar residual y en ocasiones a errores en la especificación del modelo ARIMA.

La mayoría de veces a priori no se conoce los fenómenos exógenos que afectan la serie de tiempo y más bien se utiliza una primera aproximación del modelo ARIMA para determinar la presencia de valores anómalos que son posteriormente incorporados al modelo.

A continuación se describen las principales variables de intervención [5]:

- Variables Impulso: Recoge el efecto de fenómenos que intervienen en la serie en un único momento T0. Esto se traduce en una variable que contiene un uno en T0 y ceros en el resto. Afecta el componente irregular de la serie.

- Variable escalón: Recoge el efecto de un cambio en el nivel en la serie, es decir, que contienen ceros hasta el momento T0 y unos en adelante. Afecta el componente tendencia de la serie.

- Variable tendencia o rampa: Estas contienen ceros en un tramo de la serie hasta un momento T0, a partir del cual empieza a crecer en forma ascendente. Afecta la tendencia de la serie.

- Efecto calendario: Este efecto se refiere al hecho de que cabe esperar un mayor nivel de actividad en aquellos meses con mayor número de días laborales, por lo cual hay que tomar en cuenta no solo el número de días de cada mes, sino también su diferente composición porcentual en términos de lunes, martes, etc., en cada mes.

- Efecto de la semana santa o pascua (Easter effect): Con este efecto se intenta representar la influencia de la festividad móvil de semana santa ejerce sobre la actividad económica en los meses de marzo y abril.

- Días de comercio (Trading-Days): Consiste en el ciclo semanal que se presenta cuando los días de la semana tienen un nivel de actividad distinto, unido a la distinta longitud de los meses; de tal modo que por ejemplo, un mes en particular podría tener un nivel de ventas superior a otro, debido únicamente a que posee un mayor número de días.

8.4 Modelos Autorregresivos con Promedio Móvil y Entradas Exógenas, ARMAX(p,q,n)

Además de componenter autorregresivas y de medias móviles, se pueden incorporar al modelo variables externas $\{X_i\}$ como regresores. Dichas variables son "externas" en el sentido de la información que contienen proviene de una fuente distinta a la serie de tiempo que se desea pronosticar, y los modelos resultantes se denominan modelos ARMAX(p,q,n), donde p es la cantidad de componentes autorregresivas, q componentes de medias móviles y n variables regresoras externas [6].

$$Y_t = \delta + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \beta_1 X_{1,t} + ... + \beta_n X_{n,t}$$

8.5 Modelos con varianza cambiante

En los modelos de series de tiempo, podemos distinguir dos tipos de pronósticos, los condicionales, que están condicionados a la información disponible hasta el momento, y los no condicionales. Por ejemplo, en el modelo AR(1), el pronóstico condicional de y_{t+1} es $E(y_{t+1}) = \delta + \phi_1 y_{t-1}$. Este valor contrasta con el pronóstico no condicional que es, simplemente, la media de largo plazo de la serie, $\frac{\delta}{1-\phi_1}$. Se puede demostrar que la varianza del error de pronóstico condicional es menor a la varianza del error de pronóstico no condicional, por lo que los modelos ARIMA presentados se utilizarán para la generación de pronósticos condicionales.

En muchas aplicaciones (modelamiento de la inflación, tasas de interés y rendimientos de acciones), la varianza condicional de una variable cambia a través, dependiendo de la magnitud de los errores del pasado. Se observa un " agrupamiento" de errores, esto es, periodos de alta volatilidad (y grandes errores) seguidos de periodos de baja volatilidad (y errores menores).

La adecuada representación de las variables que muestran este tipo de heterocedasticidad requiere la definición de un modelo para su varianza condicional. Una vez hecho esto se procede a la estimación simultánea de los modelos de la media y de la varianza condicional. [1]

8.5.1 Modelos de Heterocedasticidad Condicional Autorregresiva, ARCH(p)

Se basan en la existencia de una relación entre la varianza del error y los rezagos del error al cuadrado. La cantidad de rezagos utilizada determina el orden del proceso ARCH [1]:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_p \varepsilon_{t-p}^2$$

8.5.2 Modelos de Heterocedasticidad Condicional Autorregresiva Generalizado, GARCH(p,q)

Suponen una relación entre la varianza del error, los rezagos del error al cuadrado y los rezagos de la varianza. La cantidad de rezagos utilizada determina el orden del proceso GARCH:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_p \varepsilon_{t-p}^2 + \lambda_1 \sigma_{t-1}^2 + ... + \lambda_q^2 \sigma_{t-q}^2$$

Para el adecuado modelamiento de la varianza del error se deberá recurrir a herramientas como el correlograma de los residuos al cuadrado y tests para detectar la presencia de heterocedasticidad condicional autorregresiva.[1]

8.6 Verificación en el modelo ARIMA

Una vez estimado el modelo ARIMA y dado que el modelo va a ser utilizado para predecir, se debe verificar que se cumplen las hipótesis de partida. El análisis principal se centra en los residuos, pero tampoco se debe descuidar el análisis de la bondad del ajuste del modelo estimado y el análisis de los parámetros del modelo. A continuación se citan algunos de los indicadores que se deben analizar [5]:

- Análisis de los parámetros
  - Valores de los parámetros
    - $|\theta| < 1$ condición de invertivilidad (coef. de medias moviles)

    - $|\phi| < 1$ condición de estacionariedad (coef. de autocorrelacion)
  - Significancia de los parámetros (t-Student)

- Bondad del ajuste
  - Error estándar de los residuos
  - Estadístico BIC

- Análisis de los residuos (ruido blanco)
  - Análisis gráfico
  - Histograma
  - Correlograma de los residuos
  - Estadístico Q de Box-Pierce:

$$Q = T \times \sum r_k^2$$

    Este valor se compara con el valor tabular de la χ^2 con k grados de libertad. Si el valor calculado es mayor que el valor tabular se rechaza la hipótesis de estacionariedad.

9 Autocorrelación

Si se pretende establecer un modelo para una serie estacionaria un paso usual, luego de eliminar componentes estacionales y tendencias es estudiar la correlación entre una observación de la serie y las observaciones previas. La presencia de correlaciones altas entre observaciones de la serie (autocorrelaciones) puede ser consecuencia de un comportamiento lineal del fenómeno a través del tiempo y nos da idea del tipo de modelo apropiado.

Una forma visual de estudiar las autocorrelaciones es a través de correlogramas. Este tipo de gráfica nos muestra la correlación entre observaciones separadas por q intervalos de tiempo o "lags". El proceso para calcular la autocorrelación particiona las observaciones de la serie en dos grupos: $\{Y_1, Y_2, ..., Y_{t-q}\}$y $\{Y_{1+q}, Y_{2+q}, ..., Y_t\}$. La correlación es computada entre los dos conjuntos.

9.1 Definición

Dado un proceso estocástico (X_t), se define la función de Auto covarianza como la función que relaciona los valores a diferentes instantes:

$$\gamma(s,t) = Cov(X_s, X_t) = E[(X_s - E(X_s))(X_t - E(X_t))]$$

Un proceso estocástico se dirá Estrictamente Estacionario cuando la distribución conjunta es invariante ante traslaciones, o sea la función de distribución conjunta de cualquier subconjunto de variables es invariante respecto a un desplazamiento en el tiempo.

Un proceso estocástico (X_t) se dirá estacionario si se cumplen las siguientes condiciones:

1. $\gamma(s,t) = \gamma(s+r, t+r)$
2. $E(X_t) = C$
3. $E(X_t^2) < \infty$

Cuando el proceso (X_t), es estacionario, es común definir la auto covarianza como una función del desplazamiento h:

$$\gamma(h) = Cov(X_t, X_{t+h})$$

La autocorrelación corresponde a la auto covarianza normalizada por la varianza de X_t :

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

La función de autocorrelación resulta de gran utilidad para encontrar patrones repetitivos dentro de una señal, como por ejemplo, la periodicidad de una señal enmascarada bajo el ruido o para identificar la frecuencia fundamental de una señal que no contiene dicha componente, pero aparecen numerosas frecuencias armónicas de ésta. La autocorrelación muestra la asociación entre valores de la misma variable en diferentes periodos de tiempo (no aleatoria). La altura de la líneas en el correlograma representa la correlación entre las observaciones que están separadas por la cantidad de unidades de tiempo que aparecen en el eje horizontal. La correlación para el primer rezago siempre es uno por lo que no deben tomarse en cuenta en las interpretaciones

La autocorrelación parcial identifica la relación entre los valores actuales y los valores anteriores de la serie cronológica original, después de quitar los efectos de las autocorrelaciones de orden inferior. El correlograma PACF de autocorrelaciones parciales puede utilizarse para determinar, dado que parece existir una relación entre las observaciones, el orden del modelo lineal que pudiera aplicarse. Una forma recursiva de calcularla es:

$$\phi(h,h) = \begin{cases} \rho(h) & \text{si } h = 1 \\ \dfrac{\rho(h) - \sum_{j=1}^{h-1} \rho(h-1, j)\rho(h-j)}{1 - \sum_{j=1}^{h-1} \phi(h-1, j)\rho(j)} & \text{si } h = 2,3,...k \end{cases}$$

$$\text{donde } \phi(h, j) = \phi(h-1, j) - \phi(h,h)\phi(h-1, h-j) \text{ con } j = 1,2,..., h-1$$

9.2 Criterios

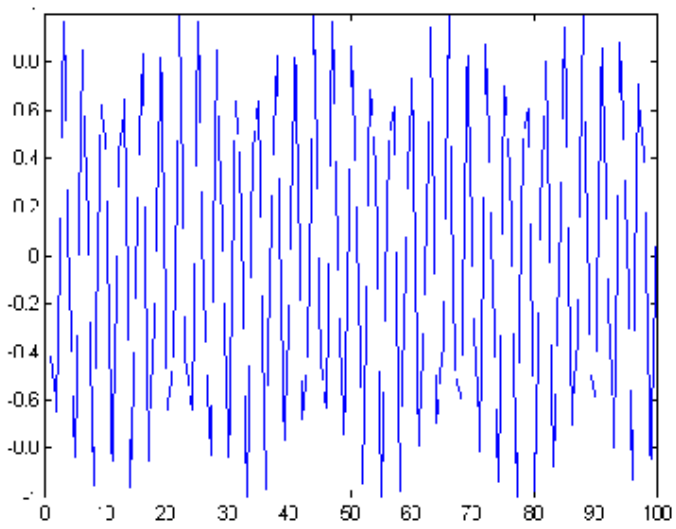
- Si ninguna de las autocorrelaciones es significativamente diferente de cero, la serie es esencialmente ruido blanco.
- Si las autocorrelaciones decrecen linealmente, pasando por el cero, o muestra un patrón cíclico, pasando por cero varias veces, la serie no es estacionaria. Se tendrá que diferenciarla una o más veces antes de modelarla.
- Si las autocorrelaciones muestran estacionalidad, o se tiene una alza cada periodo (cada 12 meses, por ejemplo), la serie no es estacionaria y hay que diferenciarla con un salto igual al periodo.
- Si las autocorrelaciones decrecen exponencialmente hacia cero y las autocorrelaciones parciales son significativamente no nulas sobre un pequeño número de rezagos, se puede usar un modelo autoregresivo
- Si las autocorrelaciones parciales decrecen exponencialmente hacia cero y las autocorrelaciones son significativamente no nulas sobre un pequeño número de rezagos, se puede usar un modelo de medias móviles
- Si las autocorrelaciones simples y parciales decrecen lentamente hacia cero, pero sin alcanzar el cero, se puede usar un modelo autoregresivo combinado con medias móviles

10 Ejemplos de series de tiempo

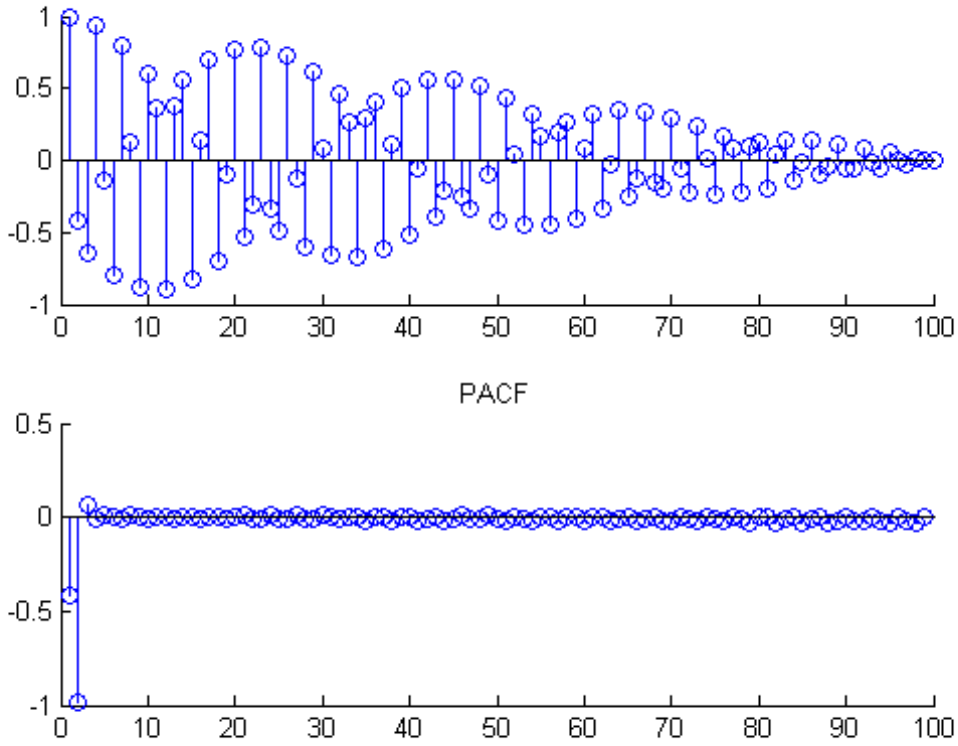
10.1 Función sinusoidal

Consideremos la función, generada en Matlab:

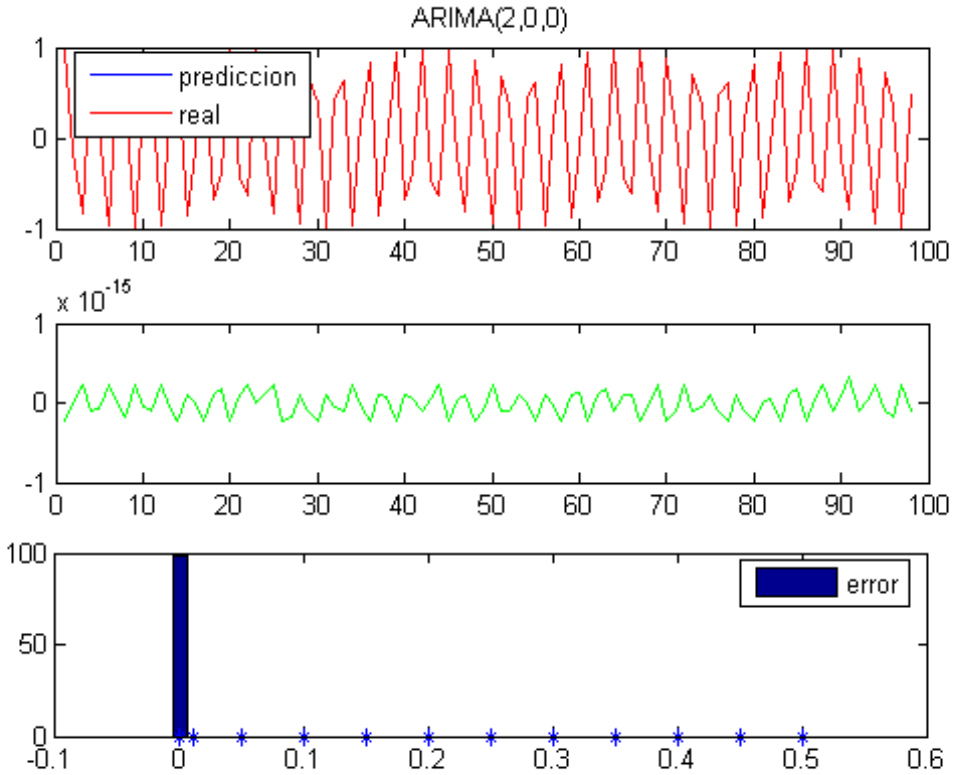
```
x=[1:1:100];
y=cos(2*x);
```



Si graficamos su correlograma obtenemos:



Observamos que sus autocorrelaciones totales se van a cero, y sus autocorrelaciones parciales son significativas en sus dos primeras componentes, lo que indica que el modelo a usar es ARIMA(2,0,0). Al calcular el modelo obtenemos el siguiente resultado:

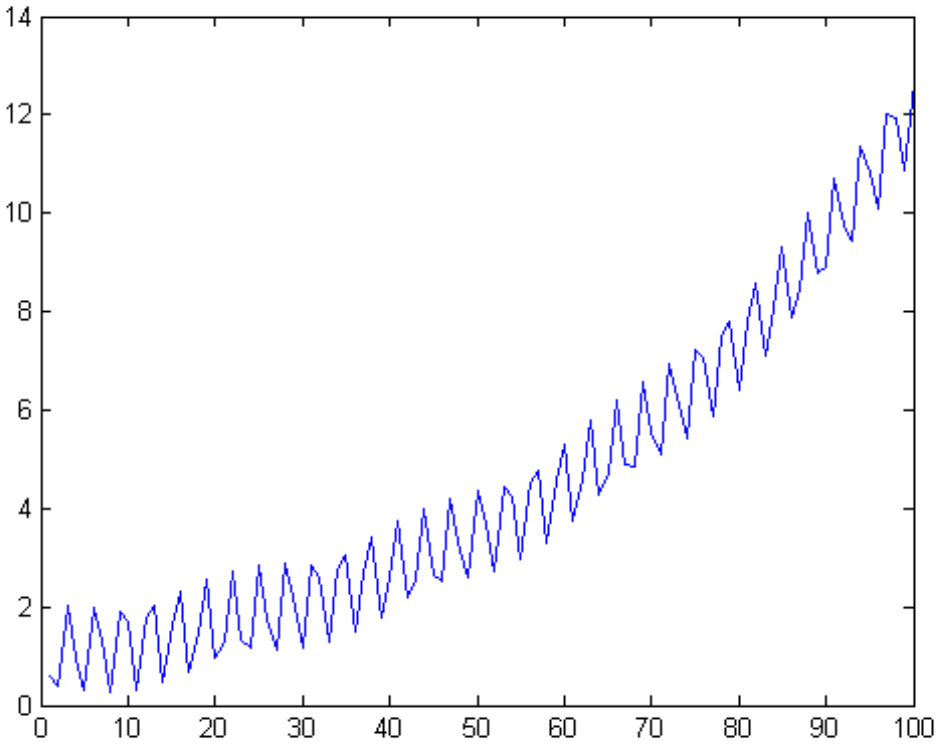


Podemos observar que el modelo es exacto, ya que el error es del orden de 10^{-15}, que se deben a efectos numéricos.

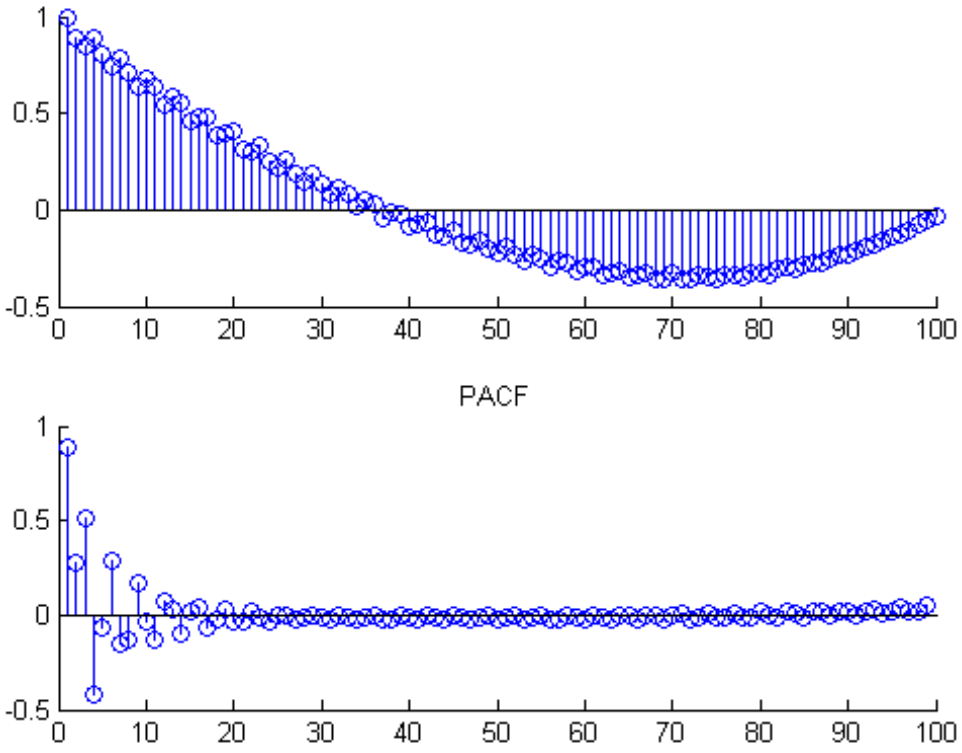
10.2 Función sinusoidal con tendencia

Consideremos la función, generada en Matlab:

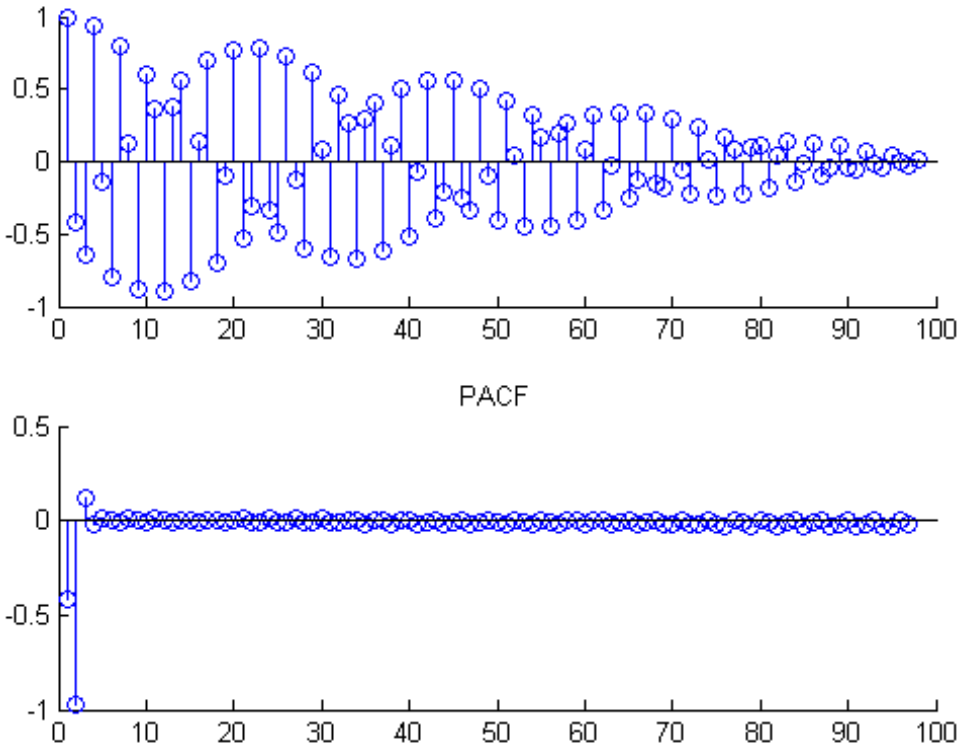
```
x=[1:1:100];
y=cos(2*x)+exp(x/40);
```



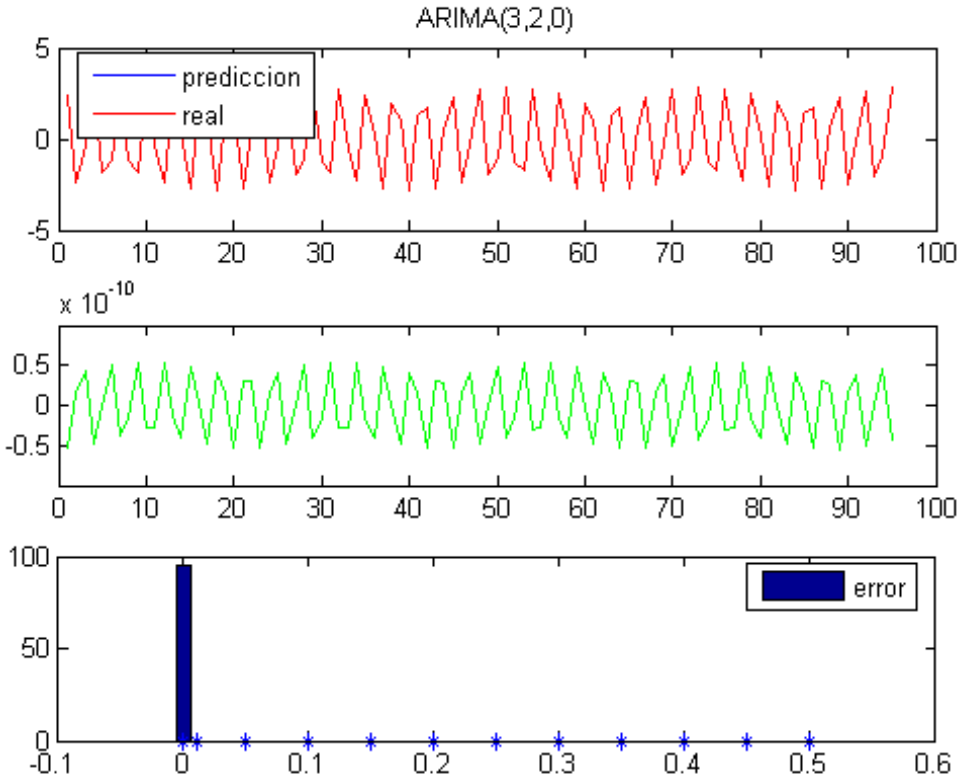
Al graficar su correlograma obtenemos:



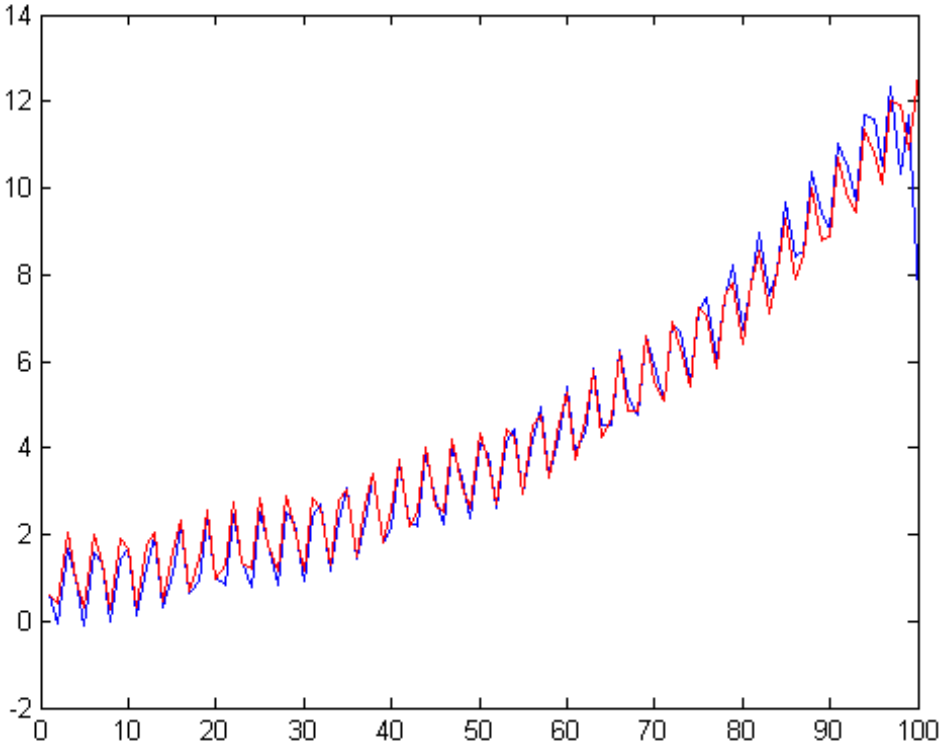
Como vemos, el correlograma indica que la serie no es estacionaria, por lo que se deberá diferenciar la serie. Luego de diferenciar dos veces la serie, su nuevo correlograma es:



Esto indica que se debe ocupar un modelo ARIMA(3,2,0), obteniendo el siguiente resultado:



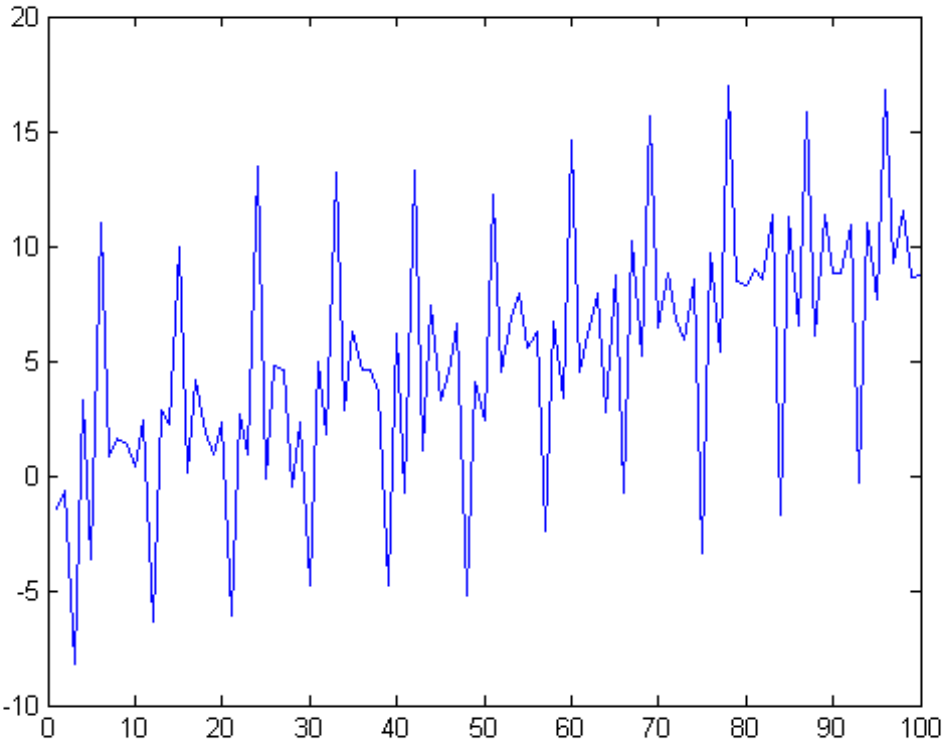
Nuevamente observamos que los resultados son exactos, ya que el error es del orden de 10^{-10}. Si reconstruimos la serie, integrando, obtenemos:



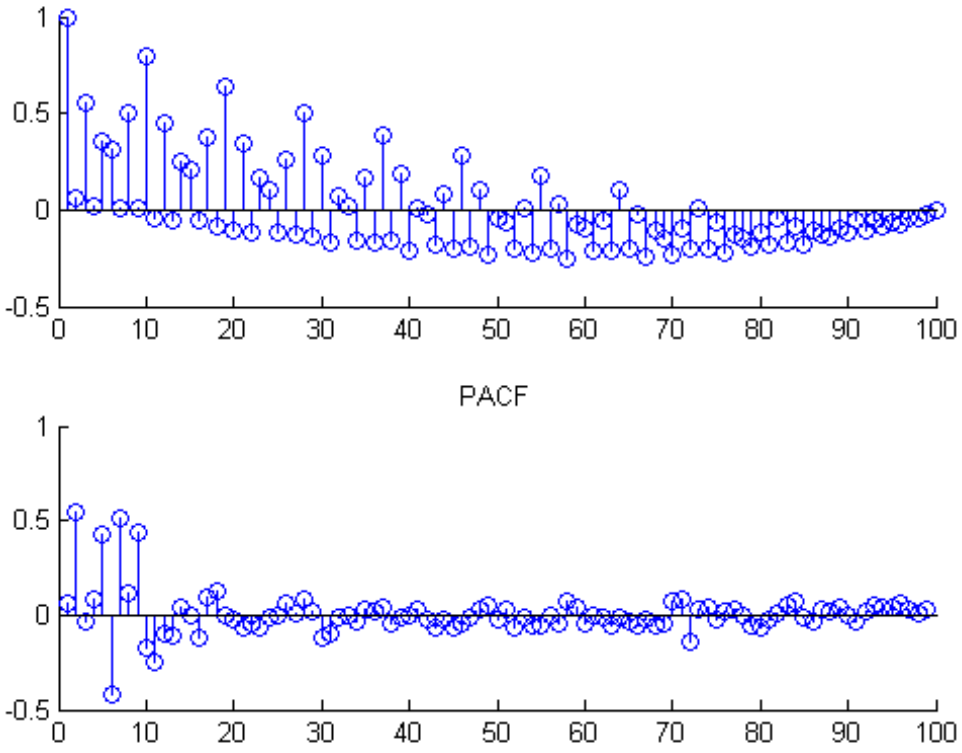
10.3 Función multisinusoidal con tendencia y componente aleatoria

Consideremos la función, generada en Matlab:

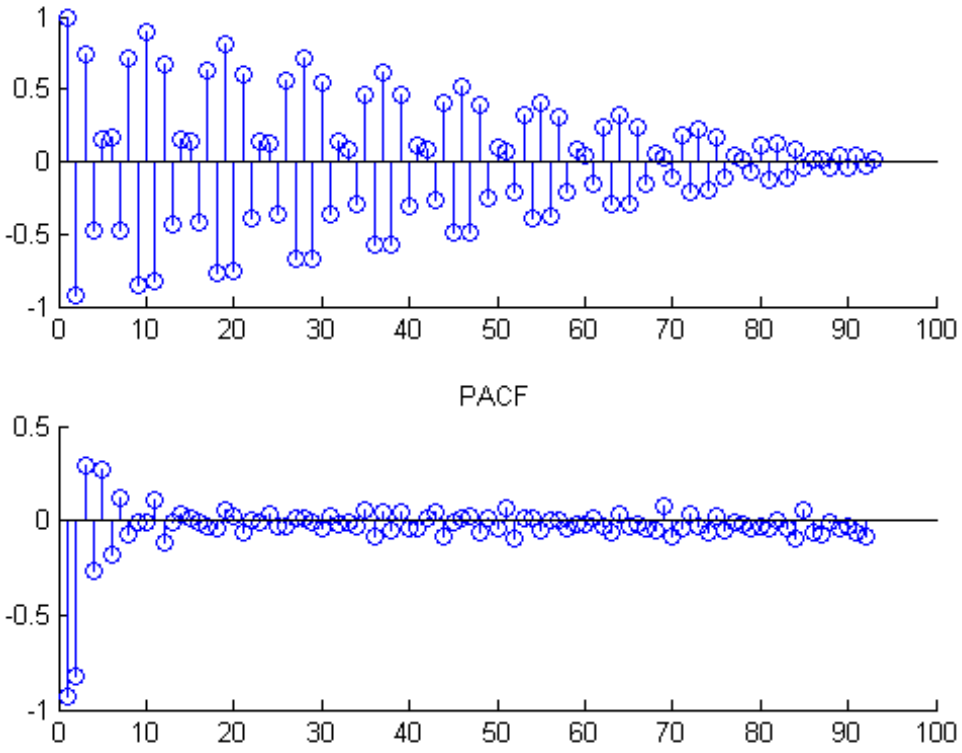
```
x=[1:1:100];
y=0.1*x+10*cos(x*pi/3).*sin(x*pi/9).*cos(x*pi/1.5)+1.5*randn(1,length(x));
```



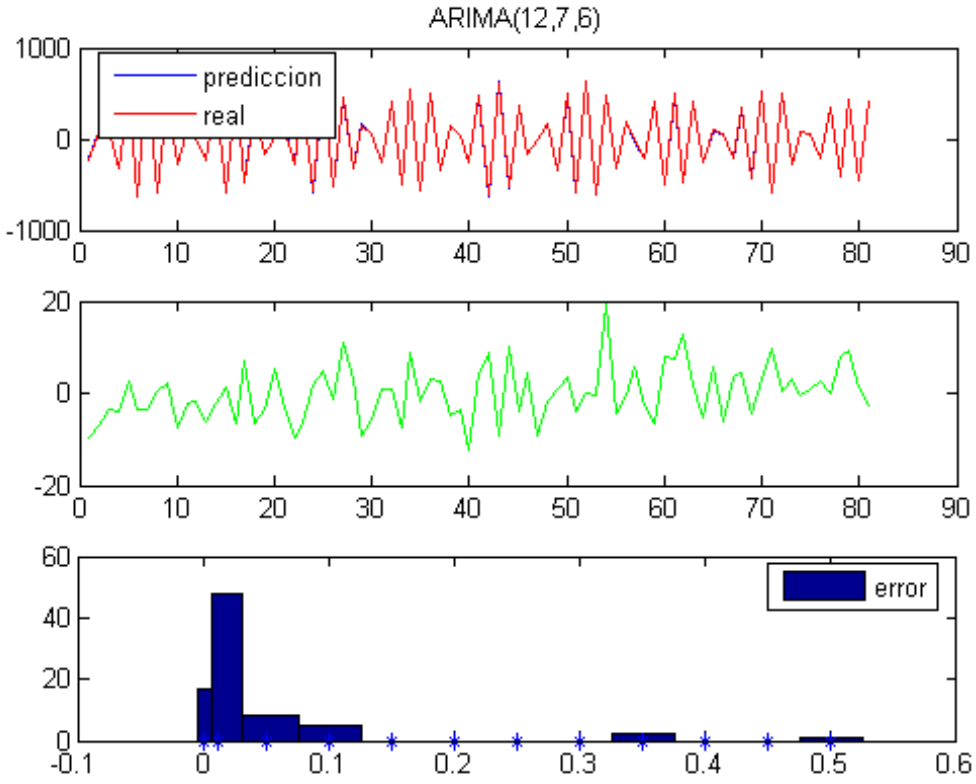
Al observar su correlograma, obtenemos:



Como vemos, el correlograma indica que la serie no es estacionaria, por lo que se deberá diferenciar la serie. Luego de diferenciar 8 veces la serie, su nuevo correlograma es:



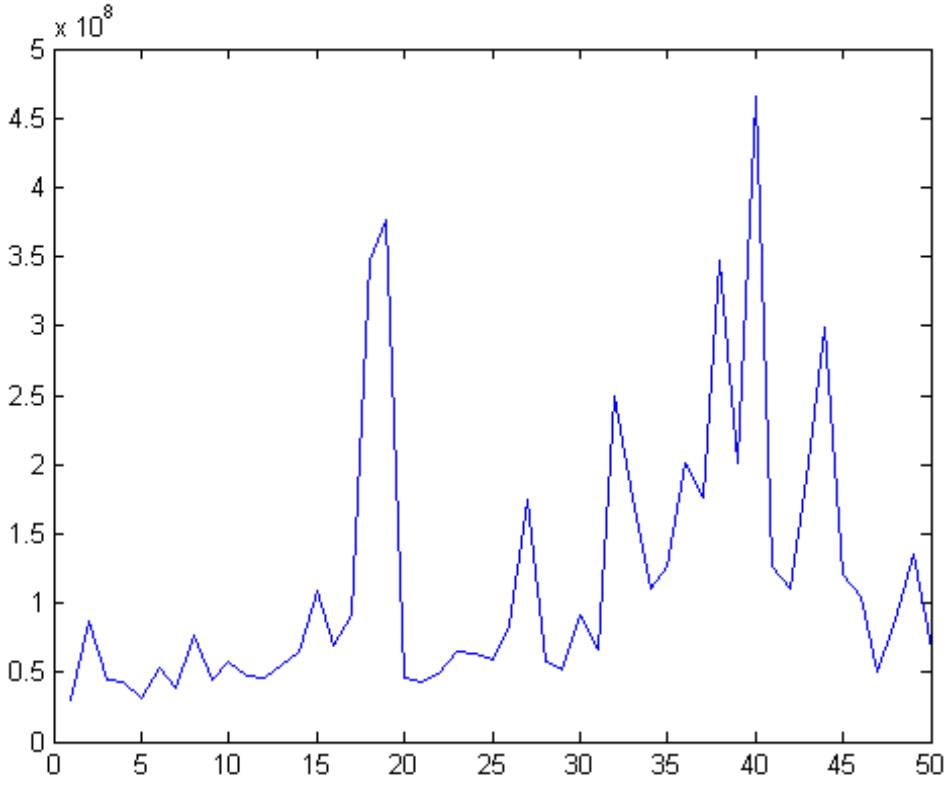
Podemos observar que su autocorrelación total se va a cero, pero su autocorrelación parcial es fuertemente significativa en las primeras 8 componentes, y luego se observa que son menos significativas, pero no cero. Esto indica que se debe usar un modelo ARIMA(12,7,q), donde q se obtiene de forma empírica. Luego de probar varios valores para q, obtenemos un modelo ARIMA(12,7,6):



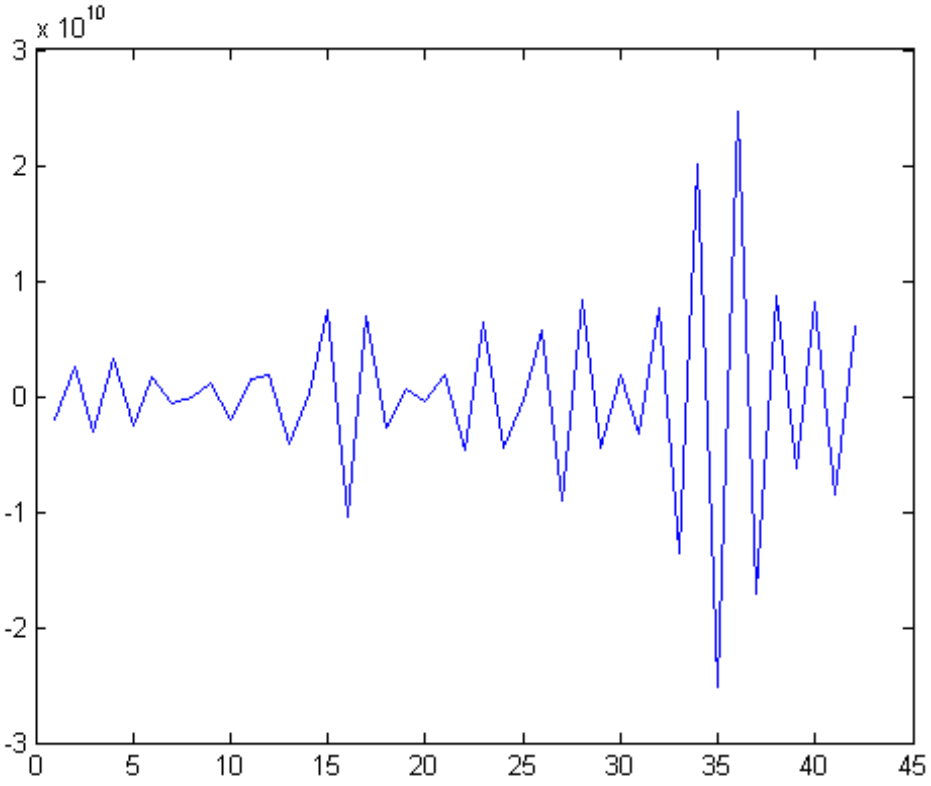
Podemos observar que el modelo es bastante preciso en general, ya que hay 3 datos con un error relativo mayor al 10%, 5 datos con error del 10%, 8 datos con error del 5%, 48 datos con un 1% de error y 17 datos con un error menor al 1%. Este error es sobre la serie diferenciada 7 veces, por lo que se debe reconstruir la serie original integrando la serie obtenida, pero el error que se obtiene sigue siendo relativamente pequeño.

10.4 Ventas mensuales de una empresa

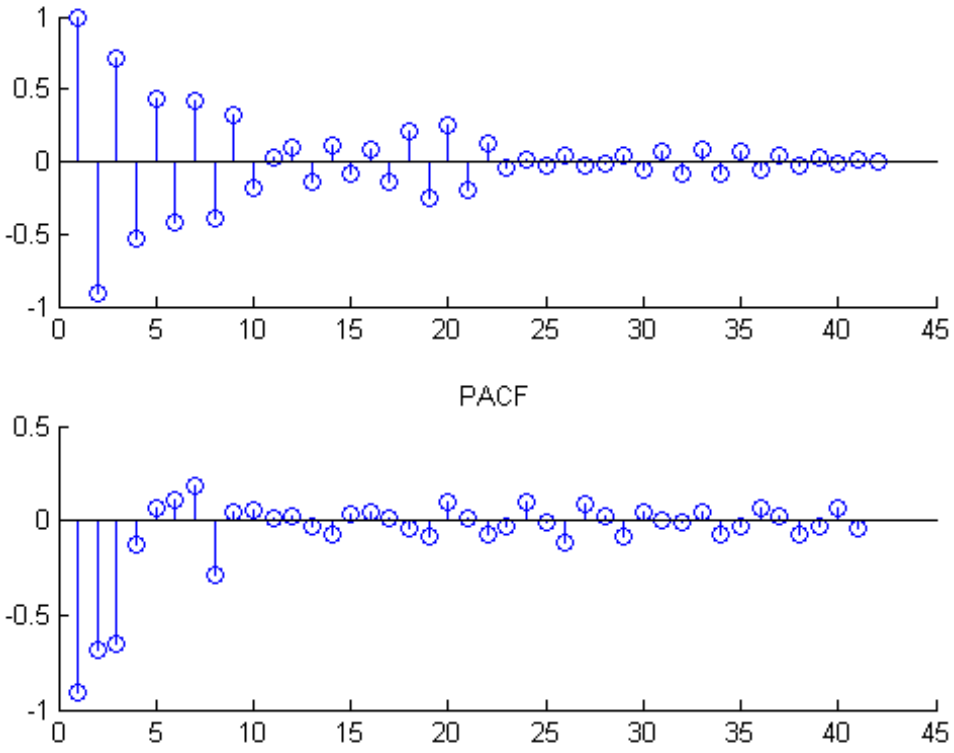
Se tienen las ventas mensuales de una empresa por un periodo de 50 meses. Se puede observar el siguiente gráfico:



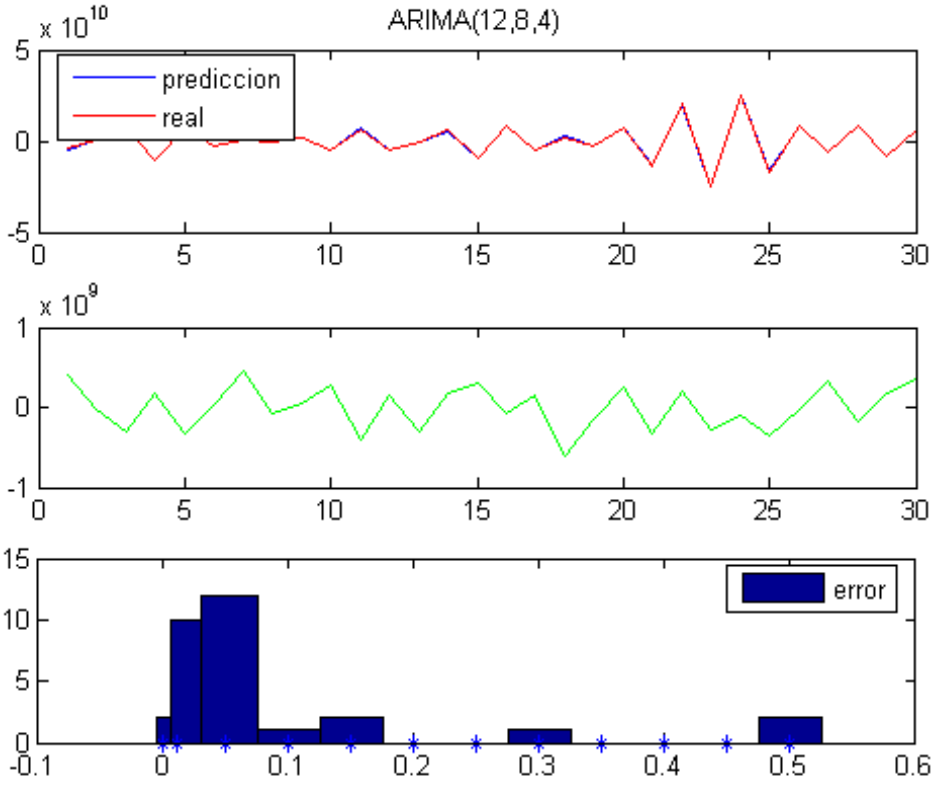
Después de hacer el análisis de las autocorrelaciones, se decide diferenciar la serie 8 veces, obteniendo el siguiente gráfico:



El correlograma muestra que se debe usar un modelo ARIMA(8,8,q)



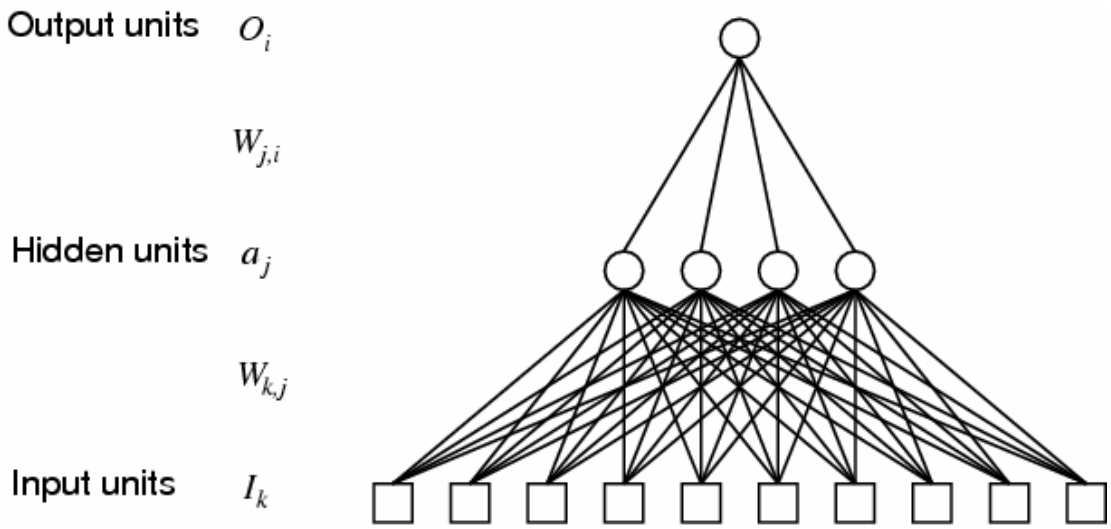
Luego de hacer las pruebas, nos quedamos con un modelo ARIMA(12,8,4), que los resultados se observan en el siguiente gráfico:



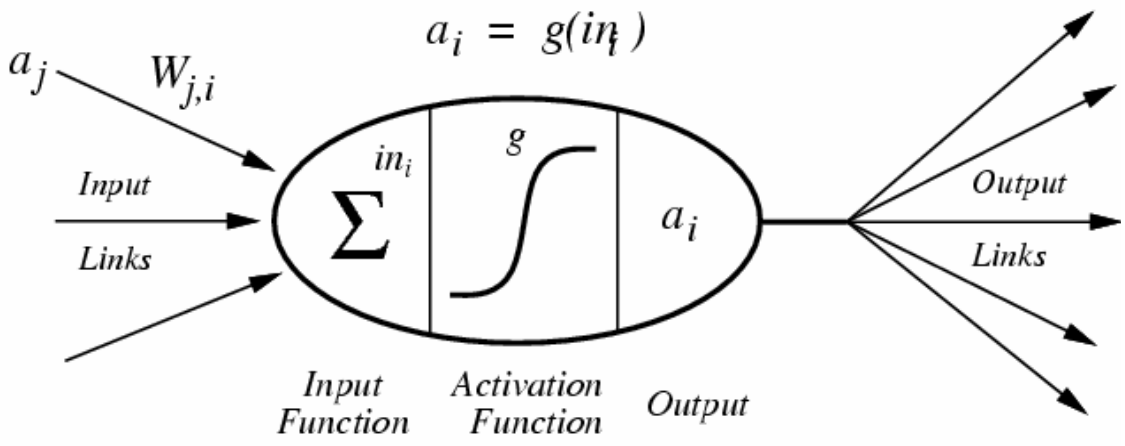
11 Técnicas de Inteligencia Computacional en Series de Tiempo

11.1 Redes Neuronales

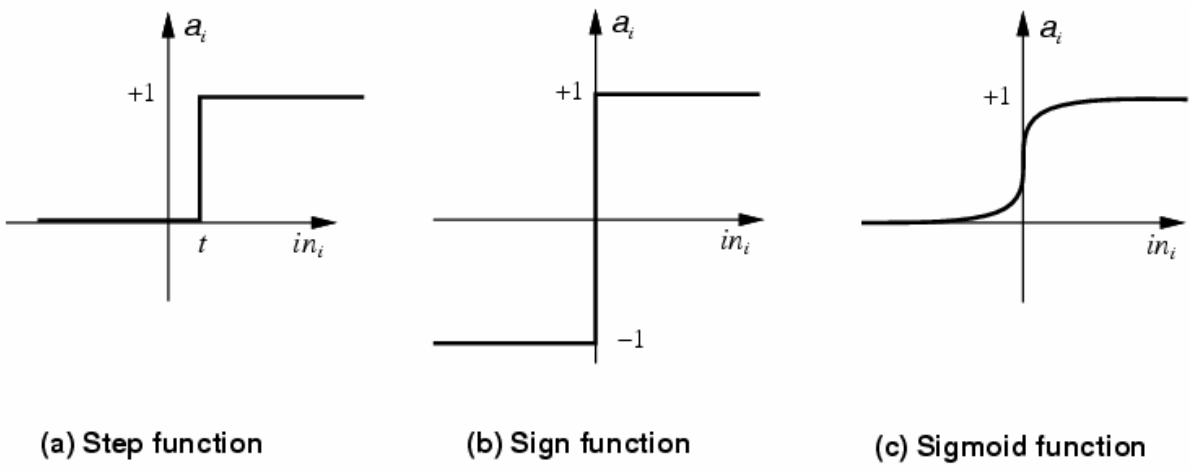
Una red neuronal puede ser descrita como un modelo de regresión no lineal cuya estructura se inspira en el funcionamiento del sistema nervioso. En términos generales, una red consiste en un gran número de unidades simples de proceso, denominas neuronas, que actúan en paralelo, estan agrupadas en capas y están conectadas mediante vínculos ponderados. Esto constituye la estructura de una red neuronal.



Cada neurona recibe inputs desde otras neuronas y genera un resultado que depende sólo de la información localmente disponible, ya sea almacenada internamente o plasmada en los ponderadores de las conexiones. El output generado por la neurona servirá de input para otras neuronas.



La llamada función de activación, es una función que emula el umbral presente en el sistema nervioso, que si la respuesta de una neurona no es lo suficiente mente grande, entonces esta no afecta en las siguientes neuronas. Las funciones más usuadas son la función escalón, signo, sigmoideal, gaussiana y lineal.



(a) Step function (b) Sign function (c) Sigmoid function

Mediante la adecuada modificación de los ponderadores de la red, en un proceso denominado aprendizaje, la red mejorará su desempeño en el desarrollo de la tarea para la cual fue construida. Este aprendizaje se basa en minimizar el error de la red neuronal. Los algoritmos clásicos que se usan para el aprendizaje de la red neuronal es el método del gradiente, gradiente conjugado y Levenberg-Marquardt, siendo este último el que presenta mejores resultados, ya que la convergencia es más estable y rápida.

Las redes neuronales tienen el potencial de implementar funciones complejas. Se puede demostrar que una red neuronal suficientemente grande, con una estructura y ponderadores adecuados, es capaz de aproximar cualquier función con el nivel de precisión que se desee.

El diseño de una red para resolver un problema con éxito puede ser una tarea muy compleja y larga dado la gran cantidad de decisiones de diseño que se deben tomar y la gran cantidad de parámetros que se deben definir. La enumeración completa de todas las alternativas no es práctica por requerir de un elevado número de evaluaciones, motivándose la aparición de variadas heurísticas y reglas basadas en la experiencia. Sin embargo, ninguna heurística ha mostrado la capacidad de entregar modelos con buen desempeño predictivo en cualquier conjunto de datos. Temas como la determinación automática del número de capas o neuronas ocultas están actualmente bajo investigación, lo que hace que, en la práctica, el método más común para el diseño de redes neuronales sea el de "prueba y error", cuya duración podría ser prolongada dado que no se debe descartar una red mientras ésta no haya completado su aprendizaje.

En el desarrollo de pronósticos de series de tiempo, existen antecedentes de buenos resultados mediante el empleo combinado de modelos de series de tiempo (ARIMA) y redes neuronales. [1]

11.1.1 Aplicación de redes neuronales en series de tiempo

El enfoque de las redes neuronales en series de tiempo, es que el valor de la serie en el tiempo T depende de forma no lineal de los valores de la serie en T-1,...,T-k, es decir:

$$y_t = f(y_{t-1}, ..., y_{t-k})$$

Usualmente, se preprocesan los datos, diferenciando la serie para obtener la estacionalidad, y una normalización para dejar la serie en el intervalo [-1,1], con la fórmula:

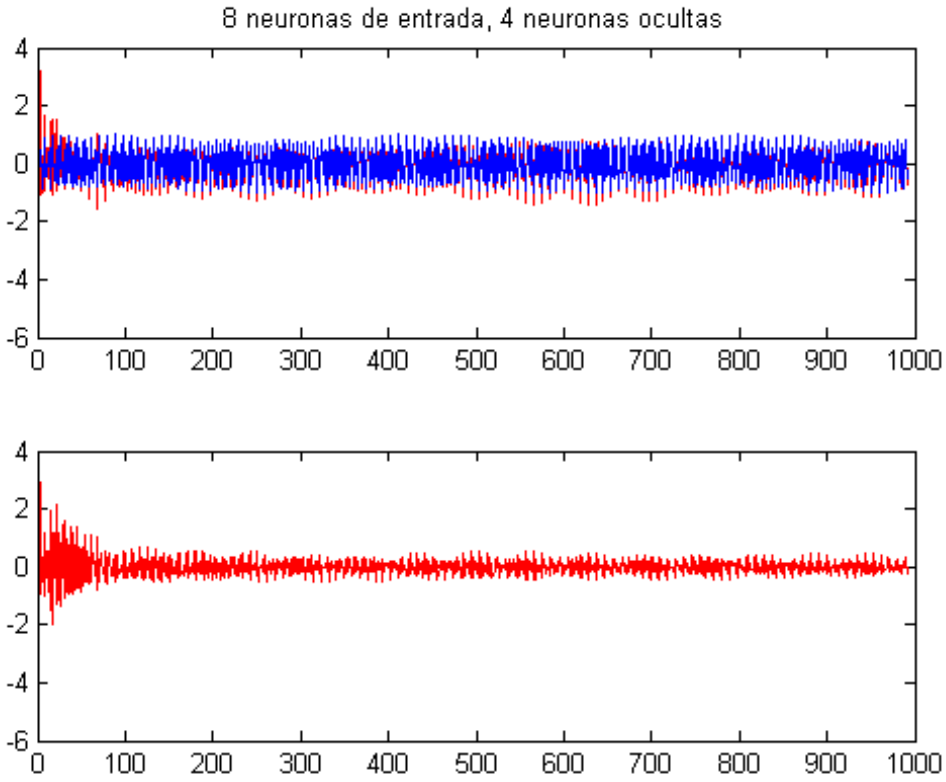
$$y\_aux = \frac{2 * y - (max(y) + min(y))}{max(y) - min(y)}$$

Una estructura típica es considerar k neuronas de entrada, que corresponde al vector $(y_{t-1}, ..., y_{t-k})$, una capa de n neuronas ocultas, con n menor que k, y una neurona de salida, que corresponde a y_t. La conección más usual es que cada una de las neuronas de entrada se conecta con todas las neuronas ocultas, y todas las neuronas ocultas

se conectan con la neurona de salida. Usualmente se usando la función sigmoide como función de activación entre las neuronas, y la función lineal como función de salida.

Una observación técnica es que se crean dos neuronas fantasmas, una en la capa de entrada y otra en la capa oculta, y siempre su respuesta es igual a 1. Esto se hace para considerar las respectivas constantes de umbral en la transferencia entre las capas.

En nuestro ejemplo, consideramos 8 neuronas de entrada y 4 neuronas oculatas, donde la función es y=3*sin(7*x)+5*cos(30*x) luego de diferenciarla una vez y normalizandola.



Podemos observar que hay casos el error de la red se mantiene a lo largo del tiempo. Pueden haber varias causas de este fenómeno:

1. El algoritmo de aprendizaje cayó en un mínimo local
2. La cantidad de neuronas es menor al óptimo
3. La estructura de la red no es la adecuada
4. Las funciones de activación no es la adecuada

11.1.2 Redes ARIMA

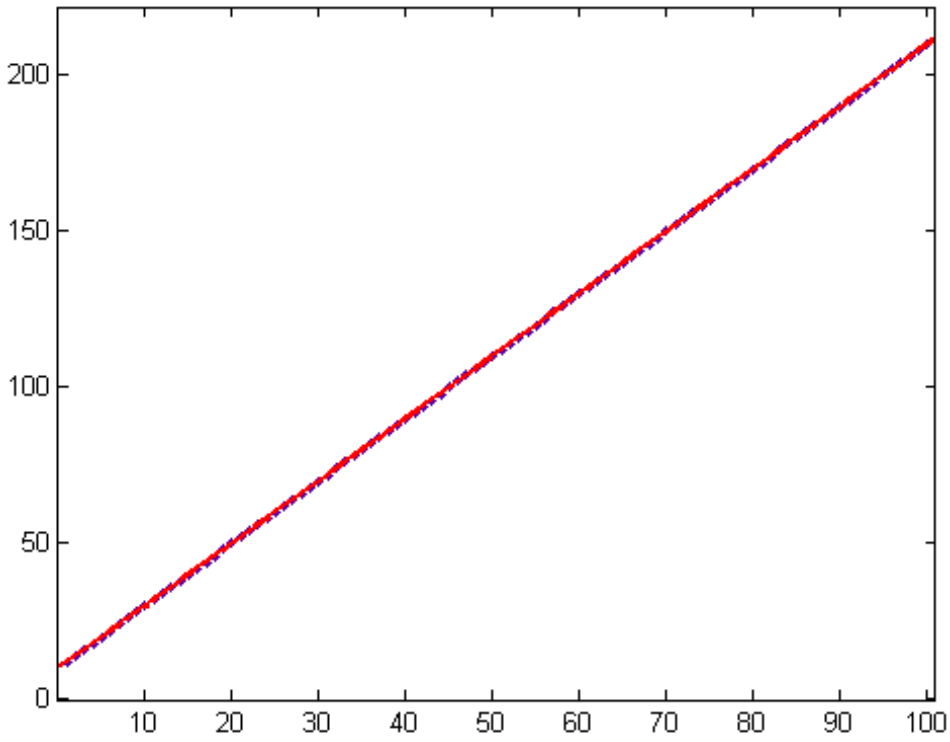
Un modelo bastante interesante de redes neuronales es simular un modelo ARIMA. Por ejemplo, si deseamos simular un modelo ARIMA(4,0,0), entonces consideramos una red neuronal de una capa de entrada con 4 neuronas y una neurona de salida, con la función lineal como activación.

Teóricamente, el modelo matemático es el mismo, pero en la práctica, este modelo tiene características diferentes al modelo ARIMA original. Esto se debe netamente a la forma de entrenar el modelo, ya que el modelo ARIMA se entrana através de regresiones lineales, mientras que las redes neuronales através de algoritmos no lineales.

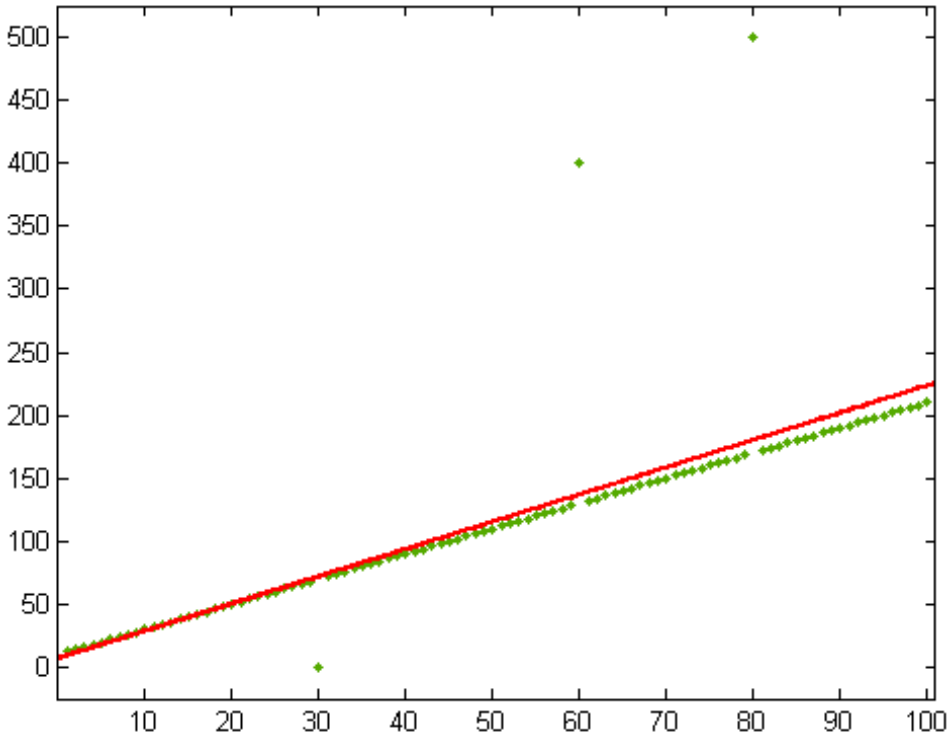
Esta diferencia en el entrenamiento implica dos diferencias importantes en los modelos obtenidos.

Outlayers En el caso del entrenamiento con regresión lineal, el modelo ARIMA es sensible a los outlayers, convergiendo a un modelo con ruido. En el caso de las redes neuronales, como el entrenamiento es punto a punto, si la cantidad de outlayers es menor, entonces los datos correctos arreglarán el error generado por los outlayers, convergiendo al modelo exacto.

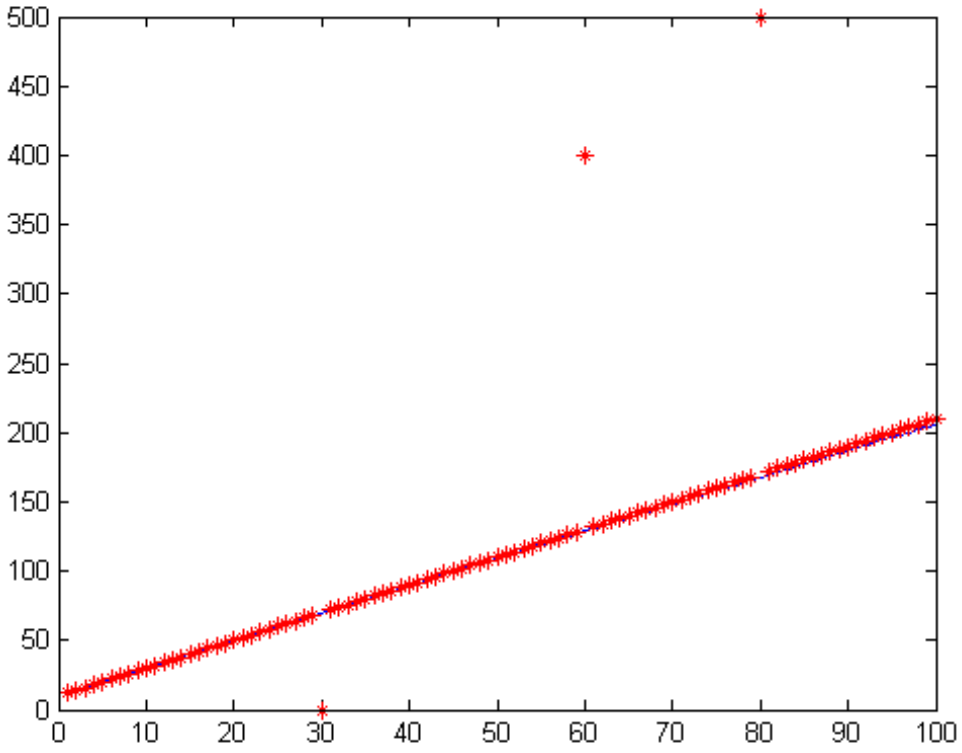
Por ejemplo, el siguiente gráfico muestra una regresión exacta, ya que todos los puntos son colineales: y=2*x+10



Ahora, si agregamos 3 outlayers, la regresión deja de ser exacta:



Si al mismo set de datos entrenamos la curva con redes neuronales, obtenemos el siguiente resultado:



Luego, esto indica que la forma de entrenar con redes neuronales ayudará a no tomar en cuenta los datos outlayers, siempre y cuando sean en una proporción pequeña.

Cantidad de parámetros En el caso del entrenamiento con regresión lineal, el modelo ARIMA es sensible a la cantidad de parámetros, por ejemplo, si se define la función:

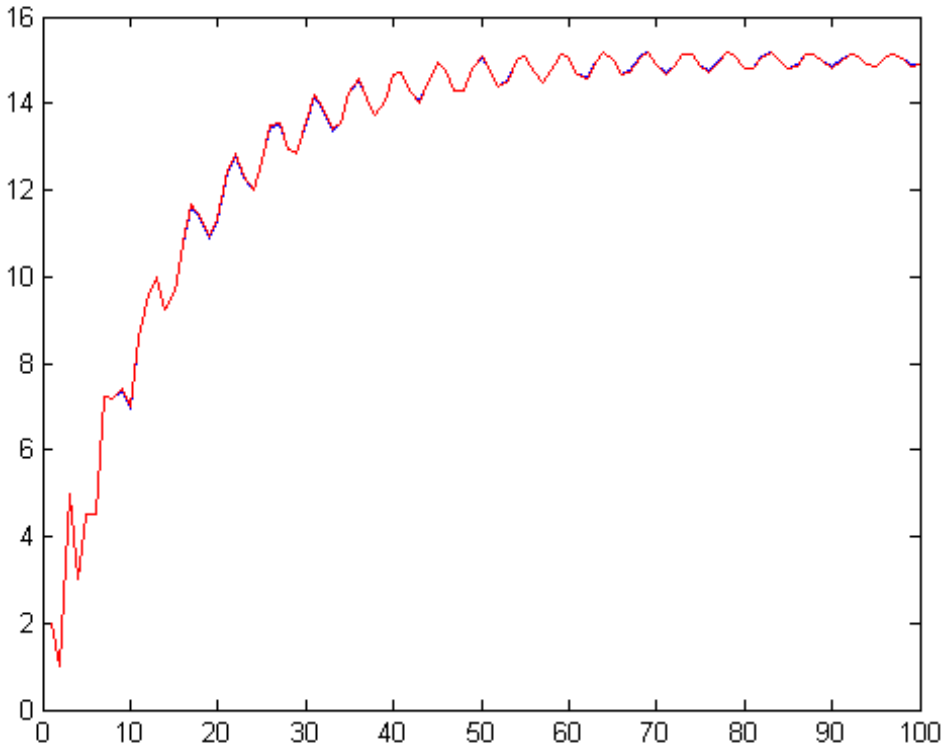
```
y(1)=2;
y(2)=1;
y(3)=5;
y(4)=3;
for k=5:1:length(x)
    y(k)=0.6*y(k-1)-0.3*y(k-2)-0.2*y(k-3)+0.7*y(k-4)+3;
end
```

Entonces el modelo ARIMA(4,0,0) va a converger de forma exacta al modelo, pero ARIMA(5,0,0) no converge a la solución. Es decir, es muy sensible a la precisión de la cantidad de parámetros. Por otro lado, al aplicar las redes arima al ejemplo anterior, simulando ARIMA(4,0,0), el método converge a la solución exacta, y al emular ARIMA(5,0,0), converge a otra solución de parámetros: [1.3382, -0.7403, 0.0213, 0.8467, -0.5174, 0.7734]. Si reconstruimos esta función

```
z(1)=2;
z(2)=1;
z(3)=5;
z(4)=3;
z(5)=4.5;
for k=6:1:length(x)
    z(k)=1.3382*z(k-1)-0.7403*z(k-2)+0.0213*z(k-3)+0.8467 *z(k-4)-0.5174*z(k-5)+0.7734;
```

end

y graficamos ambas funciones, observamos que es la misma, pero codificada de otra forma:



En otras palabras, con esta forma de entrenar no es necesario conocer la cantidad exacta de parámetros del modelo.

11.2 Support Vector Machines

SVM fueron creados por Boser, Guyon y Vapnik en 1992. La formulación original está motivada por la resolución de problemas de clasificación, donde la idea básica de SVM para abordar tal problema consiste en "mapear los datos desde el espacio original a un espacio de mayor dimensión a través de una transformación no lineal escogida a priori, para luego contruir el hiperplano de separación óptimo en el nuevo espacio". De esta manera, mediante la resolución de un problema lineal en el nuevo espacio, se tiene un modelo no lineal en espacio original.

En base a la misma filosofía, el método se extendió luego a problemas de regresión y de clustering. Desde su creación, SVM han acaparado gran atención teórica, siendo el método aplicado con éxito a problemas prácticos de predicción de series de tiempo de distinta naturaleza.

Dentro de las principales características de SVM se cuentan [6]:

- la resolución de un problema convexo y la imposibilidad de entrampamiento en óptimos locales
- la representación de la solución en base a una fracción del total de puntos disponibles (estos puntos son llamados Support Vectors)
- la capacidad de generalización a nuevos datos, debido a que el algoritmo SVM encarna el principio de minimización del riesgo estructural propuesto en la Teoría de Aprendizaje Estadístico de Vapnik

- la capacidad de modelas fenómenos no lineales mediante la ya citada transformación de los datos desde el espacio original a un espacio de mayor dimensión, espacio en el cual se obtiene un modelo lineal que equivale a un modelo lineal en el espacio original

11.2.1 Definiciones del modelo

Funciones de Kernel Un kernel se define como una función K, tal que $\forall x, z \in X$

$$K(x, z) =< \Phi(x), \Phi(z) >$$

donde X es el espacio de los datos de entrada (finito, generalmente $\mathbb{R}^n$), y Φ es una función de mapeo de los datos de entrada desde X a un espacio F de mayor dimensión, donde <•,•> es el producto interno de F.

Se puede probar que K(x,z) es una función de kernel si y sólo si la matriz M=$(\mathrm{K}(\mathrm{x}_i, x_j))_{i,j=1}^n$ es semidefinida positiva. Alguno de los kernel más comunes son:

- Lineal: K(x,x')=<x,x'>
- Polinomial: K(x,x')=(<x,x'>+1)^d
- RBF: K(x,x')=exp(-$\|x - x'\|^2 / 2\sigma^2$)

Estructura de la SVM El modelo de SVM se puede ver como capas de nodos, en donde:

- La primera capa consta de n nodos, que corresponden al vector de entrada
- La segunda capa consta de N nodos, que es la transformación no lineal a base de support vectors
- La tercerca capa contiene 1 solo nodo, que es la predicción
- Cada capa se conecta de forma completa con la siguiente
- Los nodos que llegan al nodo de output se ponderan por constantes, que son a determinar por el modelo, y luego se suman

Durante el proceso de aprendizaje, la primera capa selecciona las bases $\mathrm{K}(\mathrm{x}_i, X), i = 1, .., N$, dentro del conjunto de bases posibles, en tanto que la segunda capa construye una función lineal en el nuevo espacio, lo que es equivalente a encontrar un modelo no lineal en el espacio de entrada. Las N bases seleccionadas son aquellas inducidas por los puntos denominados Support Vectors.

Funciones de pérdida El modelo que se busca es de la forma y=f(x)+e, donde f(x) es una función no lineal y e el error. Luego, uno desea minimizar el valor de $\mathrm{y}_i - f(x_i)$=e, para cada i, y para esto se usa una funcion de pérdida. Las más comunes son

- Cuadrática: L(f(x),y)=(f(x)-y)^2
- $\varepsilon - sensible$: L(f(x),y,ε)=$\begin{cases} 0 & \text{si} \quad |\mathrm{f(x)\text{-}y}|<\varepsilon \\ |\mathrm{f(x)\text{-}y}|\text{-}\varepsilon & \text{si} \quad \text{no} \end{cases}$
- $Huber$: L(f(x),y,u)=$\begin{cases} \frac{1}{2}(\mathrm{f(x)\text{-}y})^2 & \text{si} \quad |\mathrm{f(x)\text{-}y}|<u \\ \mathrm{u}|\mathrm{f(x)\text{-}y}|\text{-}\frac{u^2}{2} & \text{si} \quad \text{no} \end{cases}$

11.2.2 Algoritmo de Regresión SVM

El problema de optimización que encuentra los pesos del modelo, usando función de pérdida $\varepsilon - sensible$ es

$$\begin{aligned} & \min \frac{1}{2}\|w\|^2 \\ s.a. y_i - & < w, \Phi(x_i) > -b \leq \varepsilon \\ -y_i + & < w, \Phi(x_i) > +b \geq \varepsilon \end{aligned}$$

El problema es que puede ser que no exista solución, por lo que se reformula como

$$\begin{aligned} & \min \frac{1}{2}\|w\|^2 + \mathrm{C}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ s.a. y_i - & < w, \Phi(x_i) > -b \leq \varepsilon + \xi_i \\ -y_i + & < w, \Phi(x_i) > +b \geq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0,\ i = 1, 2, ..., l \end{aligned}$$

donde C es un parámetro a fijar, que representa el trade-off entre la complejidad y la exactitud del modelo, y el parámetro ε representa el rango de tolerancia a los errores en el modelo. Este problema tiene solución, y además es convexo, por lo que los métodos de optimización convergen bien a la solución, y el planteamiento del dual es bastante más sencilla que el problema primal.

Una vez encontrados los pesos w, entonces nuestro modelos es:

$$y = \sum_{i=1}^{N} w_i K(X_i, x)$$

12 Modelo para un conjunto de series de tiempo

Todos los modelos clásicos de series de tiempo presentados parten del supuesto que uno conoce una historia prolongada de la serie, sigue algún comportamiento estacionario (u obtenerlo al diferenciar la serie), y se tiene una sola serie de tiempo. En muchos casos, estos supuestos no se cumplen, por lo que los modelos presentados anteriormente fallan. En esta sección presentaremos una metodología para abordar un tipo de problema de serie de tiempo bastante diferente.

12.1 Definición del problema

En muchos casos, se tiene una gran base de datos con una historia relativamente pequeña de series de tiempo, pero la cantidad de series de tiempos diferentes es enorme. Ejemplos de estos son los bancos, que tienen la historia de sus clientes, pero en muchos casos estos son solo un par de meses.

Además, estos datos pueden presentar una variabilidad enorme, por lo que el uso de estadísticos de variabilidad comunes como la desviación estándar y el coeficiente de variación no sean de gran utilidad.

A pesar de esta gran variabilidad, se desea hacer pronósticos a corto plazo de los valores de las distintas series de tiempo, pero esta predicción debe ser diferente para cada serie de tiempo en particular.

12.2 Algunos principios claves

12.2.1 Concepto de dato "normal"

El problema principal se reduce a saber si un dato está dentro de un intervalo esperado, en otras palabras, si el valor es "normal", ¿Pero que significa que un dato sea "normal"?. Debemos definir el concepto de "normal":

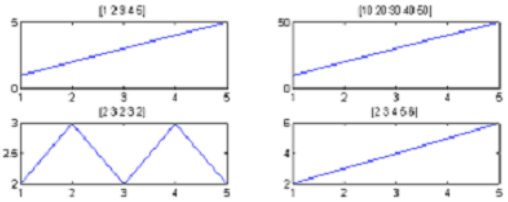
- Normal es el término por el que es conocido cualquier cosa que carece de diferencias significativas con su colectivo.
- Normal también es un término estadístico, que hace referencia al promedio aceptado.
- Pero para que algo sea "normal", debe existir un patrón a seguir, alguna regla.
- Ahora, para una misma característica, esta regla de "normalidad" puede ir cambiando dependiendo del entorno.
- Por ejemplo, algo tan cotidiano como el desayuno, el concepto de "normal puede variar drásticamente:
  - Un desayuno japonés normal, consiste en una sopa, arroz, y un vegetal.
  - En Chile se desayuna generalmente café o té, con o sin leche, acompañdo de tostadas con mantequilla, huevo revuelto, palta, queso o jamón.
  - En Argentina se desayuna mate cocido, té o café con leche, con algo dulce (medialunas, facturas, masitas dulces).

Pero para los japoneses, el desayuno japonés es "normal". Esto es porque todos ellos tienen tienen un comportamiento similar. Para poder ver si un dato es normal, debemos compararlo con sus vecinos, que son los individuos que más se parecen. Pero ahora aparece otro problema, ¿Qué clientes se parecen entre sí? Para ver si dos clientes se parecen, debemos definir una cierta "distancia" entre los clientes La vecindad de un cliente en particular serán aquellos clientes que estén más cerca. Definamos que significa el concepto de "distancia" entre los clientes.

12.2.2 Concepto de "distancia"

Para poder definir el concepto de distancia, hay que tomar una escala de comparación. Por ejemplo:

- ¿[1 2 3 4 5] se parece a [10 20 30 40 50]?
- ¿[1 2 3 4 5] se parece a [2 3 2 3 2]?
- ¿[1 2 3 4 5] se parece a [2 3 4 5 6]?
- Lo que importa para considerar la distancia no son sus valores en sí, sino la forma de estos:



12.3 Características fundamentales del modelo

12.3.1 Normalizando los datos

Para hacer los datos comparables, debemos normalizarlos. La normalización que conserva la forma de los datos, pero los lleva a una misma escala es

$$\tilde{X} = \frac{X - \mu}{\sigma}$$

Además, se cumple que $E(\tilde{X}) = 0$ y $Var(\tilde{X}) = 1$. Aplicando esta normalización a los datos anteriores obtenemos:

- $[1\ 2\ 3\ 4\ 5] \Longrightarrow \begin{bmatrix} -1.2649 & -0.6325 & 0 & 0.6325 & 1.2649 \end{bmatrix}$
- $[10\ 20\ 30\ 40\ 50] \Longrightarrow \begin{bmatrix} -1.2649 & -0.6325 & 0 & 0.6325 & 1.2649 \end{bmatrix}$

- $[2\ 3\ 2\ 3\ 2] \Longrightarrow \begin{bmatrix} -0.7303 & 1.0954 & -0.7303 & 1.0954 & -0.7303 \end{bmatrix}$
- $[2\ 3\ 4\ 5\ 6] \Longrightarrow \begin{bmatrix} -1.2649 & -0.6325 & 0 & 0.6325 & 1.2649 \end{bmatrix}$

Ahora, los números muestran lo mismo que los gráficos.

12.3.2 Función de distancia

Ahora bien, para generalizar la distancia entre cualquier conjunto de datos, tomamos la correlación entre los datos normalizados:

$$corr(X,Y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}$$

La correlación indica la fuerza y la dirección de una relación lineal entre los datos. Aplicandolo en el ejemplo anterior:

- $corr([-1.2649\ -0.6325\ 0\ 0.6325\ -1.2649], [-1.2649\ -0.6325\ 0\ 0.6325\ -1.2649]) = 1$
- $corr([-1.2649\ -0.6325\ 0\ 0.6325\ -1.2649], [-0.7303\ 1.0954\ -0.7303\ 1.0954\ -0.7303]) = 0$

Luego, nuestra vecindad serán aquellos clientes que tienen una correlación cercana a 1

12.3.3 Características de la vecindad

Dado que tenemos datos através del tiempo, estos son una serie temporal, que proviene de un proceso estocástico discreto. Este proceso estocástico tiene una ley de probabilidad conjunta. Al tomarnos una vecindad del cliente, los vecinos tendrán una ley de probabilidad conjunta muy similar a la original, por lo menos en el intervalo de comparación. Si promediamos todos los vecinos, obtendremos otro proceso estocástico, pero con una característica muy importante:

- $E(\frac{1}{N}\sum X) = E(X)$
- $Var(\frac{1}{N}\sum X) = \frac{Var(X)}{n}$

Luego, el proceso promedio tendrá el mismo valor esperado, pero una varianza mucho menos si n es grande. Como n es el número de vecinos, se debe tomar el máximo valor de vecinos, pero que tengan una ley de probabilidad conjunta similar a la original.

12.3.4 Independencia de los datos con el tiempo

Recordemos que lo más importante de nuestra función de distancia, es que capture la "forma" de los datos. Es decir, importa el orden de los datos, pero no importa la posición temporal de los datos:

- Si en el año 2007 tengo los datos [1 2 3 4 5 6 7 8 9 10 11], mi predicción para el mes de Diciembre será 12.
- Si en el año 2008 tengo los datos [1 2 3 4 5 6 7 8 9 10 11], mi predicción para el mes de Diciembre será 12.

Además, no es necesario conocer toda la historia de un cliente para predecir el valor del próximo mes, basta con una "ventana de tiempo". La memoria no es eterna, tiene una largo finito. Los hechos del pasado lejano influyeron sobre el pasado cercano, y estos influyen sobre el presente. Luego, al considerar una ventana de tiempo fija, y que la posición temporal de los datos no importa, podemos "viajar al pasado"

12.3.5 Principio fundamental del modelo

"Si en el pasado se conoce nuestro presente, entonces se conoce el futuro del pasado". Si tomamos una ventana de tiempo fija, con los ultimos datos del cliente, podemos comparar su comportamiento en el pasado con otros clientes. Al viajar al pasado, buscamos los clientes que más se parecen en esta ventana de tiempo. Luego, la ley de probabilidad conjunta de cada una de las ventanas de tiempo se van a parecer entre sí en una vecindad de la ventana. Como conosco los datos fuera de la ventana de los clientes, puedo estimar el futuro del cliente. Este estimador será el promedio de los vecinos, ya que de esta forma se mantiene el promedio, pero la varianza disminuye, dependiende del número de clientes seleccionados.

12.3.6 Ejemplo

- Se tiene un cliente con la serie [1 5 2 6 3 7 4 5 6 7 8 9]
- Consideramos la ventana de tiempo [4 5 6 7 8 9]
- Se tiene otro cliente con la serie [2 3 4 5 6 7 8 11 15 16 14 11]
- Si viajo al pasado 6 meses, la ventana de tiempo del último cliente será [2 3 4 5 6 7]
- Luego, la correlación entre [4 5 6 7 8 9] y [2 3 4 5 6 7] es 1, luego este dato es un vecino, y el valor estimado para el cliente original será 10
- Luego, este procedimiento se repite con todos los vecinos, y el valor esperado será el promedio de las estimaciones.

12.4 Explicación matemática del modelo

Conozco 20 meses de N clientes. Sea k el cliente que quiero predecir sus ventas en el mes 21. Si consideramos una ventana de tiempo de largo L=8, entonces tomo los meses [13 14 15 16 17 18 19 20] del cliente k. Denotemos como X_k este vector. Normalizamos este vector:

$$\tilde{X_k} = \frac{X_k - \bar{X}_k}{std(X_k)}$$

De los demás datos consideramos los meses [10 11 12 13 14 15 16 17], ya que debemos retroceder 3 meses. Denotemos X_i el vector asociado al cliente i$\neq k$. Ahora normalizamos cada uno de estos vectores:

$$\tilde{X_i} = \frac{X_i - \bar{X}_i}{std(X_i)}$$

Comparamos $\tilde{X_k}$ con cada uno de los $\tilde{X_i}$ y nos quedamos con los q vectores más parecidos. Estos son los q vectores con la correlación $corr(\tilde{X_k}, \tilde{X_i})$ más cercana a 1, o de forma equivalente, con el menor error cuadratico.

$$E_i = \sum_{j=1}^{L} (\tilde{X_i^j} - \tilde{X_k^j})^2$$

Considere q=10 vecinos. Tomamos el valor del mes 18 de cada uno de estos vecinos y los normalizamos según el intervalo anterior, es decir:

$$\tilde{Y_j} = \frac{Y_j - \bar{X}_j}{std(X_j)}$$

donde Y_j es el valor en el mes 18 (ya que 21-3=18) del vecino j, y X_j es el vector de los meses [10 11 12 13 14 15 16 17] de ese vecino. Luego, tomamos el promedio de cada uno de eso $\tilde{Y_j}$:

$$\tilde{Y} = \frac{\sum_{j=1}^{q} \tilde{Y_j}}{q}$$

Repito el mismo procedimiento, pero ahora considerando los meses [1 2 3 4 5 6 7 8], ya que

1	2	3	4	5	6	7	8	9	10	11	12
10	11	12	13	14	15	16	17	18	19	20	21
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9

Luego, la nueva predicción será a base del mes 9=21-12. Denotemos $\tilde{Z}_j$ a la normalizacion de cada una de las predicciones de los vecinos, y llamemos $\tilde{Z}$ al promedio de las predicciones. Ahora, promediamos ambas predicciones:

$$\tilde{W} = \frac{\tilde{Z} + \tilde{Y}}{2}$$

Finalmente, la predicción para el mes 21 del cliente k será:

$$P = \tilde{W} * std(X_k) + \bar{X}_k$$

donde X_k es el vector de los meses [13 14 15 16 17 18 19 20] del cliente k
El intervalo de confianza será

$$[I_{\inf} * std(X_k) + \bar{X}_k, I_{\sup} * std(X_k) + \bar{X}_k]$$

donde

$$\begin{aligned} I_{\inf} &= \tilde{W} - \sqrt{\frac{E}{L-1}} - z * \frac{\sqrt{\sum_{j=1}^{q}(\tilde{Y}_j - \tilde{W})^2 + \sum_{j=1}^{q}(\tilde{Z}_j - \tilde{W})^2}}{2*q-1} \\ I_{\sup} &= \tilde{W} + \sqrt{\frac{E}{L-1}} + z * \frac{\sqrt{\sum_{j=1}^{q}(\tilde{Y}_j - \tilde{W})^2 + \sum_{j=1}^{q}(\tilde{Z}_j - \tilde{W})^2}}{2*q-1} \end{aligned}$$

con z=2.861, y E=$\frac{1}{2*q}\sum_{i=1}^{2*q}\sum_{j=1}^{L}(\tilde{X}_i^j - \tilde{X}_k^j)^2$, donde $\tilde{X}_i^j$ es la componente j del vecino i normalizado, y $\tilde{X}_k^j$ es la componente j del cliente k normalizado. En otras palabras, E es el promedio de los errores cuadraticos medios.
El intervalo de rango será

$$[I_{\min} * std(X_k) + \bar{X}_k, I_{\max} * std(X_k) + \bar{X}_k]$$

donde

$$\begin{aligned} I_{\min} &= \min(\min_j \tilde{Y}_j, \min_j \tilde{Z}_j) \\ I_{\max} &= \max(\max_j \tilde{Y}_j, \max_j \tilde{Z}_j) \end{aligned}$$

Obs: Recuerde que $\tilde{Y}_j$ denota al j-esimo vecino de la primera ventana de tiempo, y $\tilde{Z}_j$ denota al j-esimo vecino de la segunda ventana de tiempo.
Para considerar las transacciones, se reemplaza la tabla de ventas por la de transacciones, y lo demás es igual.

12.5 Resultados