CSE 6740 Lecture 15

How Do I Treat Temporal Data? (Time Series Analysis)

Alexander Gray

agray@cc.gatech.edu

Georgia Institute of Technology

Quiz Answers

1. Association rule mining, in general, returns a large result. T.
2. Mixture of Gaussians clustering is a procedural method. F.

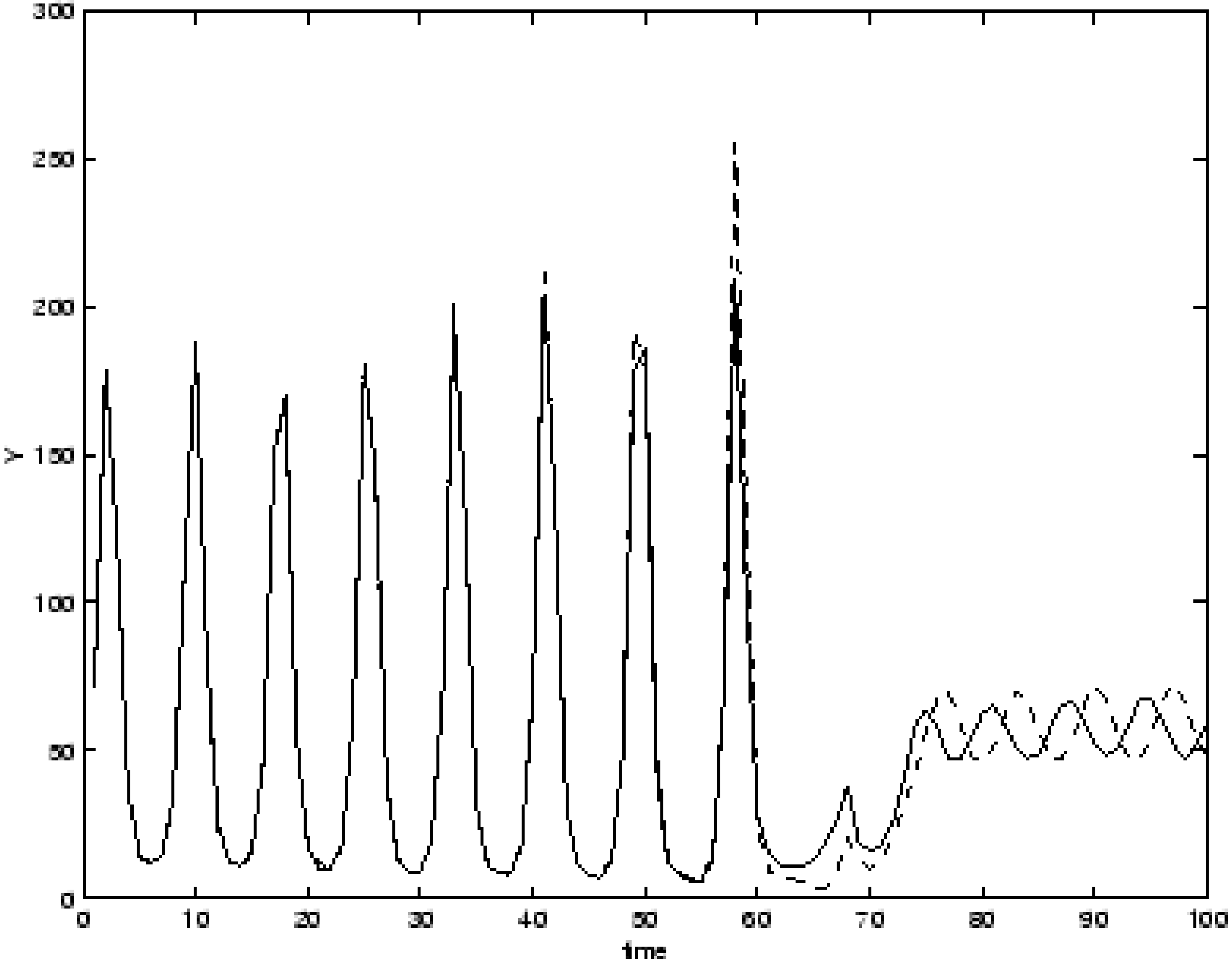
Today

1. Time Series
2. Univariate Linear Methods
3. Extensions

Time Series

General concepts in time series analysis.

Time Series



Time series are not IID. Each data point is somehow dependent on previous ones.

Parts of a Time Series

Some sources of variation:

- *Seasonality.* For example, unemployment is typically high in the winter and lower in summer.
- *Trend.* A long-term change in the mean level.

Example of a trend:

$$X_t = m_t + \epsilon_t \tag{1}$$

where

$$m_t = \beta_0 + \beta t. \tag{2}$$

and ϵ_t is zero-mean noise.

Parts of a Time Series

Seasonality can be modeled in different ways, including:

$$X_t + m_t + S_t + \epsilon_t \quad \text{additive} \tag{3}$$

$$X_t + m_t S_t + \epsilon_t \quad \text{multiplicative} \tag{4}$$

$$X_t m_t S_t \epsilon_t \quad \text{multiplicative} \tag{5}$$

Note that the third model can be made into the first one with a logarithmic transformation.

Stationarity

A time series is said to be *stationary*, roughly speaking, if there is no systematic change in mean (no trend) or variance and if strictly periodic variations have been removed.

In other words, the properties of one section of the data look much like that of any other section.

(Actually, only models are stationary, not data.)

Differencing

Differencing is a pre-processing operation which is effective for removing a trend. It is performed on the time series $\{x_1, \ldots, x_N\}$ to obtain a new time series $\{y_2, \ldots, y_N\}$ by

$$y_t = x_t - x_{t-1} \equiv \nabla x_t. \tag{6}$$

Occasionally second-order differencing is required:

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1} = x_t - 2x_{t-1} + x_{t-2}. \tag{7}$$

A seasonal effect can be removed with seasonal differencing, e.g.:

$$\nabla_{12} x_t = \nabla x_t - \nabla x_{t-12}. \tag{8}$$

Stochastic Process

A *stochastic process* or *random process* is a sequence of random variables $X_1, \ldots, X_t, \ldots, X_N$ ordered in time. The mean function is defined by

$$\mu(t) = \mathbb{E}(X_t). \tag{9}$$

The variance function is defined by

$$\sigma^2(t) = \mathbb{V}(X_t). \tag{10}$$

The *autocovariance function* is defined by

$$\gamma(t_1, t_2) = \mathbb{E}\left\{(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})\right\}, \tag{11}$$

which generalizes the variance function when $t_1 \neq t_2$.

Stationary Process

A time series is said to be *strictly stationary* if the joint distribution of $X_1, \ldots, X_N$ is the same as the joint distribution of $X_{1+\tau}, \ldots, X_{N+\tau}$ for all $t_1, \ldots, t_N$ and τ.

In other words, shifting the time origin by an amount τ has no effect on the joint distributions, which thus must depend only on the intervals between $t_1, \ldots, t_N$.

Stationary Process

Strict stationarity, using the case $N = 1$, also implies that the distribution of X_t is the same for all t, or $\mu(t) = \mu$ and $\sigma^2(t) = \sigma^2$. Using the case $N = 2$, it also implies that $\gamma(t_1, t_2)$ depends only on the *lag* τ, i.e. $\gamma(t_1, t_2) = \gamma(\tau)$.

If $\mathbb{E}(X_t) = \mu$ and Cov$(X_t, X_{t+\tau}) = \gamma(\tau)$ we say the process is *second-order stationary* or *weakly stationary*.

Univariate Linear Methods

Some linear models, for a single time series.

Purely Random Process

A process is called a *purely random process* or *white noise* if it consists of a sequence of random variables $\{Z_t\}$ which are IID.

Suppose they are normally distributed with mean zero and variance σ_Z^2. Then Cov$(X_t, X_{t+\tau}) = \gamma(\tau) = \sigma_Z^2$ for $\tau = 0$, otherwise 0.

Random Walk

Suppose that $\{Z_t\}$ is a process with mean μ and variance σ_Z^2. A process $\{X_t\}$ is called a *random walk* if

$$X_t = X_{t-1} + Z_t. \tag{12}$$

Starting the process with $X_1 = Z_1$, we have $\mathbb{E}(X_t) = t\mu$, $V(X_t) = t\sigma_Z^2$. Since the mean and variance change with t, the process is non-stationary.

However, note that first-order differencing

$$\nabla X_t = X_t - X_{t-1} = Z_t \tag{13}$$

forms a purely random process, making it stationary.

Moving Average Process

Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2. A process $\{X_t\}$ is called a *moving average process* of order q (MA(q) process) if

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}. \quad (14)$$

This process is weakly stationary.

Autoregressive Process

Suppose that $\{Z_t\}$ is a purely random process with mean zero and variance σ_Z^2. A process $\{X_t\}$ is called an *autoregressive process* of order p (AR(p) process) if

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t. \tag{15}$$

The “auto” refers to the fact that the prediction is based on past values of X_t.

AR $\leftrightarrow$ MA

Let's look at the AR(1) process, or *Markov process*:

$$X_t = \alpha X_{t-1} + Z_t. \tag{16}$$

By successive substitution we have

$$\begin{aligned} X_t &= \alpha(\alpha X_{t-2} + Z_{t-1}) + Z_t && (17)\\ &= \alpha^2(\alpha X_{t-3} + Z_{t-2}) + \alpha Z_{t-1} + Z_t && (18)\\ &= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \dots && (19) \end{aligned}$$

provided $-1 < \alpha < 1$ so that the sum converges.

In general there is a duality between AR and MA processes, i.e. each process can be written as the other.

ARMA Process

A mixed autoregressive/moving-average process containing p AR terms and q MA terms is said to be an ARMA process of order (p, q):

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}. \quad (20)$$

It turns out that a stationary time series can often be adequately modelled by an ARMA model with fewer parameters than a pure AR or MA process alone.

ARIMA Process

The *integrated ARMA process* accounts for non-stationarity by differencing first. Replacing X_t by $W_t = \nabla^d X_t$, we obtain an ARIMA(p, d, q) model:

$$X_t = \alpha_1 W_{t-1} + \ldots + \alpha_p W_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q}. \quad (21)$$

The *seasonal ARIMA* (SARIMA) process includes seasonal differencing.

Continuous Time

It is interesting to consider the continuous-time version of these models, which turns out to be much less tractable mathematically. Consider an AR(1) process $X_t = \alpha X_{t-1} + Z_t$. Then

$$(1-\alpha)X_t + \alpha \nabla X_t = Z_t. \tag{22}$$

This has the continuous analog

$$aX(t) + \frac{dX(t)}{dt} = Z(t) \tag{23}$$

where $Z(t)$ is continuous white noise, which leads to the Langevin equation. One interpretation of this analogy is that we are effectively learning differential equations.

State-Space Model

We assume the data are obtained from a linear combination of some hidden state variables s_t, plus noise:

$$X_t = m_t^T s_t + \epsilon_t \tag{24}$$

where m_t is considered to be known and $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$. We also assume we know how s_t changes in time:

$$s_t = A_t s_{t-1} + c_t \tag{25}$$

where the matrix A_t is considered known and c_t is multivariate normal with zero mean vector and known covariance matrix C_t. This Markov model is called a *state-space model*.

State-Space Model

It turns out many models can be represented in the state-space model, including all ARIMA models. As an example, consider the AR(2) model $X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t$.

By defining $s_t = (X_t, \alpha_2 X_{t-1})$, we have $X_t = (1, 0)s_t$ with $m_t^T = (1, 0)$ and $\sigma_\epsilon^2 = 0$, and

$$s_t = \begin{pmatrix} \alpha_1 & 1 \\ \alpha_2 & 0 \end{pmatrix} s_{t-1} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} Z_t \qquad (26)$$

since $s_{t-1}^T = (X_{t-1}, X_{t-2})$.

Kalman Filter

In the usual formulation, s_t is the unknown quantity to be estimated. The *Kalman filter*, or *linear dynamical system* is a way to estimate s_t when a new observation in the time series becomes available, i.e. in an *online* fashion, consisting of two stages – “prediction” and “correction”. Because it is online, it can follow the movement of a system where the underlying model is evolving through time.

Suppose we’ve observed a univariate time series up to time $t-1$, and $\widehat{s}_{t-1}$ is the best (minimum MSE) estimator for s_{t-1} up to this time.

Kalman Filter

The prediction stage is concerned with forecasting s_t from data up to time $t-1$, to obtain a prediction

$$\widehat{s}_{t|t-1} = A_t \widehat{s}_{t-1} \tag{27}$$

with covariance matrix

$$P_{t|t-1} = A_t P_{t-1} A_t^T + C_t. \tag{28}$$

When the new observation at time t, X_t, has been observed, the estimator for s_t can then be corrected to account for this additional information.

Kalman Filter

At time $t-1$, the best forecast of X_t is given by $m_t^T \widehat{s}_{t|t-1}$ so that the prediction error is

$$e_t = X_t - m_t^T \widehat{s}_{t|t-1}. \tag{29}$$

Using this error feedback, we can update the various quantities by

$$\begin{aligned} \widehat{s}_t &= \widehat{s}_{t|t-1} + K_t e_t && (30) \\ P_t &= P_{t|t-1} - K_t m_t^T P_{t|t-1} && (31) \\ K_t &= P_{t|t-1} m_t / (m_t^T P_{t|t-1} m_t + \sigma_\epsilon^2). && (32) \end{aligned}$$

K_t is called the *Kalman gain* matrix.

Extensions

Multivariate and nonlinear.

Linear Multivariate Formulations

The univariate state-space model is easily generalized to the case where X_t is a vector by making m_t a matrix and ϵ_t a vector.

By simply using lagged variables (past values), we can construct the usual data matrix for IID regression and use, say, standard linear regression. Note that the machinery we have discussing is more general – it can handle linear coefficients which change in time, for example.

Linear Multivariate Formulations

An AR model can easily use past values from other time series to predict values of the time series of interest. We can also account for feedback between different time series by using simultaneous equations – the natural extension to this setting is called *vector ARMA* (VARMA).

Inspired by the idea that the relationship between two or more stocks might remain constant, *co-integration* models focus on finding stationary linear combinations of time series variables.

Nonlinear Models

A locally linear approximation to a nonlinear system can be obtained by a simple Taylor expansion derivation, resulting in a nonlinear model called the *extended Kalman filter*.

Instead of propagating a single state mean and covariance, we can propagate a set of possible states, represented as points. One variant using a few determistically-chosen points is called *unscented Kalman filtering*.

An approach using many points to obtain a nonparametric-ish approximation of the state density is called *particle filtering*. It has been useful for cases where there are multiple modes, for example due to spurious measurements.

Nonlinear Models

A framework which allows parametric forms beyond linear models and Gaussians is called the *generalized state-space model*.

Another viewpoint is that of graphical models, which allow easy specification of many of the models we've discussed at some level. Graphical models which are repeated over time, such as Kalman filters and hidden Markov models, are called *dynamic Bayesian networks*.

Piecewise linear modelling, corresponding to different regimes of the time series, is another way of achieving nonlinearity. Mixtures of Kalman filters where regimes are defined by a discrete hidden Markov model-like process, or *switching state-space linear dynamical systems*, comprise one way to do this. *Threshold autoregressive models* are another way.

Nonlinear Models

We can always generalize the linear AR model with some class of nonlinear functions from past values to the current value.

Various ways of making linear parameters change over time constitute overall nonlinear models.

In econometrics, models which account for time-varying variance called *generalized conditionally heteroscedastic models* (GARCH) have been explored heavily.

Other Topics

- Temporal cross-validation. h-*block cross-validation*: leave a space of h points between each point to be predicted and the rest of the data used for training.
- Fourier space. Good for periodicities.
- Chaos. Gave us phase plots.

Main Things You Should Know

- What stationarity is
- What differencing does
- What an autoregressive process is
- What a state-space model and Kalman filter are

Quiz

1. (T/F) A stationary process is IID.
2. (T/F) A purely random process is IID.
3. (T/F) A Markov model is a special case of an autoregressive process.