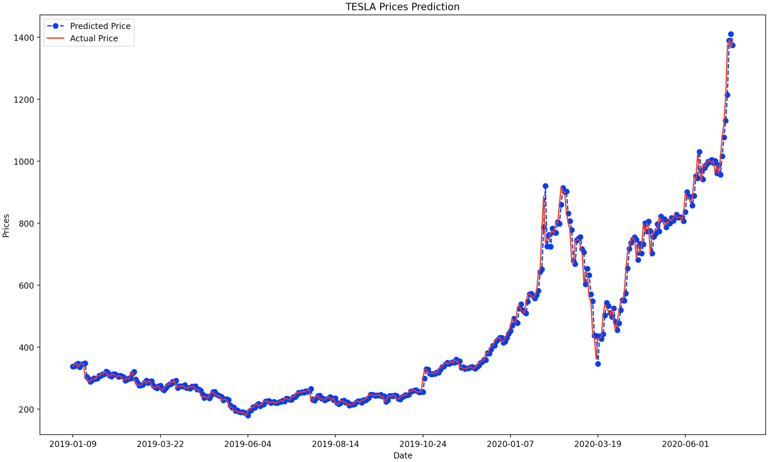
ARIMA: AutoRegressive Integrated Moving Average

Team Fourier

UC Berkeley, Fall 2020, CS 189

December 2020

Time Series


TESLA Prices Prediction
Predicted Price
Actual Price
Prices
1400
1200
1000
800
600
400
200
2019-01-09
2019-03-22
2019-06-04
2019-08-14
2019-10-24
2020-01-07
2020-03-19
2020-06-01
Date

Time Series

Time series are functions, $f(t)$, that require a special type of analysis versus normal functions. They are used to model quantities over time, such as population or price.

Recall System Identification

Recall the idea of system identification in control systems. We have this model:

$$X_{t+1} = AX_t + Bu_t$$

X_t = Observation
A = State transition rule
Bu_t = User input and effect on system

Timeseries: System Identification

The goal of timeseries modeling and forecasting is finding the proper model, in a slightly altered format to the previous system identification problem, with the correct parameter A. There is no user input, however, and instead there are residual terms, which you can think of as perturbations from the model due to changes in environment.

Timeseries: Example Model, Auto-regressive

Auto-regressive Model:

$$AR(p) = y_t = c + (\phi_1 L + \phi_2 L^2 + \phi_3 L^3 ... + \phi_p L^p) y_t + Z_t$$

L = lag operator, $Ly_t = y_{t-1}$, etc.
This is very reminiscent of the system identification model, in which current observations are a linear combination of previous observations. The ϕ's can be thought of as the A from the system identification model.

Time Series Formulation with Trend

Time series data with a trend can be formulated as follows:

$$X_t = m_t + Z_t$$

X_t = Observation at Time t
m_t = Trend
Z_t = Residual
This formulation is based on the idea that the time series data can be fully described by some function of t, m_t, and these residuals, or deviations from the function, at each time step. In time series analysis, we seek to find m_t as well as develop a model to predict Z_t or X_t directly.

Modeling and Forecasting Time Series

There are two major methods in which to model a trend time series:

- Exponential or Moving average smoothing
- ARIMA

Exponential and Simple Moving Average Smoothing

Simple Moving Average

$$\hat{m}_t = \frac{1}{2p+1} \sum_{i=-p}^{p} m_{t+i} \tag{1}$$

Simple Moving Average smoothing can be used to estimate the underlying trend m_t by averaging the values over $2p+1$ observations. Intuitively, by taking more samples, the effect of each residual diminishes.

Exponential and Simple Moving Average Smoothing

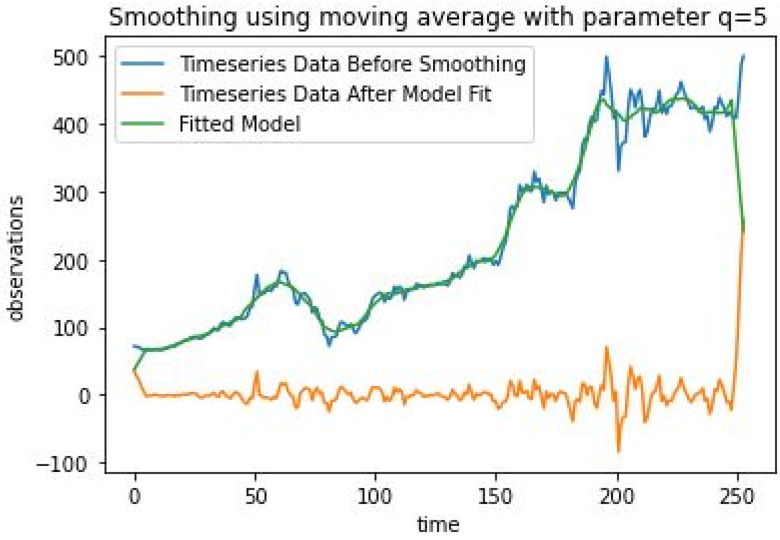
Exponential Moving Average

$$\hat{m}_t = \frac{1-\phi}{\phi}\sum_{i=1}^{\infty}\phi^i X_{t-i} \tag{2}$$

Exponential Moving Average smoothing can be used to estimate the underlying trend m_t by taking a weighted average of i samples, weighing each sample exponentially less the farther back in time from the current time. This modeling is based on the intuition that each observation is more likely influenced by the immediately previous observation than an observation some time in the past.

Exponential and Simple Moving Average Smoothing

Exponential and Simple Moving Average Smoothing



ARIMA: Time Series Analysis

ARIMA seeks to fit a model to timeseries data as well as forecast future values based on the assumption that current observations have some sort of dependence, or correlation, on previous values. However, in order to model this correlation on previous values, the data must satisfy a few properties:

- Trendless
- Stationarity

Trendless Time Series

In our original formulation, we had a timeseries formulation:

$$X_t = m_t + Z_t$$

In order to properly use ARIMA to model and forecast observations, we need to remove the trend, m_t. We already saw the use of using exponential or simple moving average smoothing to approximate the trend.

Differencing to Remove Trend

Another effective technique is differencing, in which we generate a new timeseries:

$$Y_t = X_t - X_{t-1}$$

Differencing to Remove Trend

With a linear trend, $m_t = at + b$, we immediately see a result:

$$\begin{aligned} Y_t &= X_t - X_{t-1} \\ Y_t &= at + b + Z_t - (a(t-1) + b + Z_{t-1}) \\ Y_t &= a + Z_t - Z_{t-1} \end{aligned}$$

Our result is a timeseries with no trend, with bias term, a, which can be easily removed by zero-meaning the timeseries.

Differencing to Remove Trend

With a quadratic trend, $m_t = at^2 + bt + c$, we find we have to perform differencing twice:

$$
\begin{aligned}
Y_t &= X_t - X_{t-1} \\
Y_t &= at^2 + bt + c + Z_t - (a(t-1)^2 + bt + c + Z_{t-1}) \\
Y_t &= 2at - a + Z_t - Z_{t-1} \\
Y_{t-1} &= 2a(t-1) - a + Z_{t-1} - Z_{t-2} \\
Y_t^2 &= Y_t - Y_{t-1} \\
Y_t^2 &= 2a + Z_t - 2Z_{t-1} + Z_{t-2}
\end{aligned}
$$

Our result is a timeseries with no trend, with bias term, $2a$, which can once again be easily removed by zero-meaning the timeseries.

Stationarity

Stationarity by definition is the property that any t consecutive samples taken from the time series have the same joint distribution. Intuitively, relating back to the system identification, problem, this is the idea that the A matrix defining the state transition is the same regardless of where we are in the time series.

Stationarity

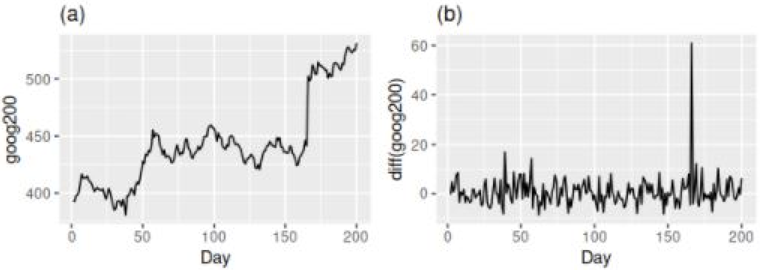


Figure a depicts a non-stationary time-series as the observations demonstrate a shifting mean. Figure b depicts a stationary time-series as the observations are very horizontal in nature.

Autocorrelation and Autocavariance Function

$$
\begin{aligned}
K_{XX}(t,s) &= \mathrm{Cov}[X_t, X_s] \\
&= \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)] \\
&= \mathbb{E}[X_t, X_s] - \mu_1 \mu_2
\end{aligned}
$$

$$
\rho(t,s) = \frac{K_{XX}(t,s)}{\sqrt{\mathrm{Var}(X_t), \mathrm{Var}(X_s)}}
$$

AR model

$$AR(p) = y_t = c + (\phi_1 L + \phi_2 L^2 + \phi_3 L^3 ... + \phi_p L^p) y_t + Z_t$$

The current observation is dependent on previous observations from the time series.

AR model: Causality

An Autoregressive model is causal if the current observation is only dependent on itself and past observations, and not future observations in the time series. This is very desirable for modeling and forecasting.

MA model

$$MA(q) = X_t = \mu + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}$$

The current observation is a combination of previous residuals.
These residuals are latent, un-observable variables.

MA model: Invertibility

$$
\begin{aligned}
Z_t &= \pi(B)X_t \\
&= \sum_{j=0}^{\infty} \pi_j X_{t-j}
\end{aligned}
\tag{3}
$$

This condition is simply an enforcement on the MA model as a bijection, meaning there is a unique solution for the parameters.

Solving for Parameters: AR model

$$\underbrace{\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{t-1} \\ X_t \end{bmatrix}} = \mu + \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ X_1 & 0 & \dots & 0 & 0 \\ & & \vdots & & \\ X_{t-2} & X_{t-3} & \dots & X_{t-p+2} & X_{t-p+1} \\ X_{t-1} & X_{t-2} & \dots & X_{t-p+1} & X_{t-p} \end{bmatrix}} \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{p-1} \\ \phi_p \end{bmatrix}}$$

We can solve for the phi's using least squares.

Solving for Parameters: MA model

Unlike AR models, the MA model is built on latent variables, residuals, that aren't observed but rather a result of the model itself. Therefore, solving for the parameters is more difficult. There are several options:

- Yule-Walker
- Iterative Least Squares

Solving for Parameters: MA model, Yule Walker

- Use sample autocovariance function as an estimation of true autocovariance
- Solve for θ's and ϕ's algebraically

Solving for Parameters: MA model, Iterative Least Squares

- Initialize parameters, θ's and ϕ's
- Recursively solve for residuals by plugging into model
- Solve for θ^*'s and ϕ^*'s given the current residuals
- Repeat with new θ's and ϕ's until model converges and the parameters no longer change

ARIMA Model: Putting it together

$$ARIMA(p,d,q) = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 ... - \phi_p L^p)(1-L)^d y_t = \\ c + (1 + \theta_1 L + \theta_2 L^2 ... + \theta_q L^q)\epsilon_t$$

An AR component, with the ϕ's, the MA component, with the θ's, and the I component, with the $(1-L)^d$. In practice, the integrated component, differencing, is run on the the time series data prior to solving for the ARMA parameters.

SARIMA Model: Putting it together

$$SARIMA(p,d,q)(P,D,Q)_m = (1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 ... - \phi_p L^p)$$
$$(1 - \Phi_1 L^m - \Phi_2 L^{2m} - \Phi_3 L^{3m} ... - \Phi_P L^{Pm})(1-L)^d (1-L)^{Dm} y_t =$$
$$c + (1 + \theta_1 L + \theta_2 L^2 ... + \theta_q L^q)$$
$$(1 + \Theta_1 L^m + \Theta_2 L^{2m} ... + \Theta_Q L^{Qm})\epsilon_t$$

Simply adds seasonal terms, multiplied with the non-seasonal terms. Allows for time series data with seasonality, or cyclical behavior.

Evaluating an ARIMA model: AIC

$$AIC = -2log(L) + 2k$$
$$k = p + q + 2$$

Akaike Information Criterion
Tradeoff between likelihood of the data, L, and complexity of the model.

Evaluating an ARIMA model: BIC

$$BIC = -2log(L) + klog(n)$$
$$k = p + q + 2$$

Bayesian Information Criterion
Tradeoff between likelihood of the data, L, and complexity of the model. Stronger penalty than AIC, incentivizes a smaller model than AIC.