



[Click to Take the FREE Time Series Crash-Course](#)

Search...



How to Create an ARIMA Model for Time Series Forecasting in Python

by Jason Brownlee on January 9, 2017 in [Time Series](#)

Tweet

Share

Share

Last Updated on August 19, 2020

A popular and widely used statistical method for time series forecasting is the ARIMA model.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a class of model that captures a suite of different standard temporal structures in time series data.

In this tutorial, you will discover how to **develop an ARIMA model for time series** forecasting in Python.

After completing this tutorial, you will know:

- About the ARIMA model the parameters used and assumptions made by the model.
- How to fit an ARIMA model to data and use it to make forecasts.
- How to configure the ARIMA model on your time series problem.

Kick-start your project with my new book [Time Series Forecasting With Python](#), including *step-by-step tutorials* and the *Python source code* files for all examples.

Let's get started.

- **Updated Apr/2019:** Updated the link to dataset.
- **Updated Sept/2019:** Updated examples to use latest API.

Autoregressive Integrated Moving Average Model

An [ARIMA](#) model is a class of statistical models for analyzing and forecasting time series data.

[Start Machine Learning](#)

It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

ARIMA is an acronym that stands for AutoRegressive Integrated Moving Average. It is a generalization of the simpler AutoRegressive Moving Average and adds the notion of integration.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

- **AR:** *Autoregression*. A model that uses the dependent relationship between an observation and some number of lagged observations.
- **I:** *Integrated*. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
- **MA:** *Moving Average*. A model that uses the dependency of an observation on a moving average model applied to lagged observations.

Each of these components are explicitly specified in the ARIMA model via parameters. For example, we can use ARIMA(p, d, q) where the parameters are substituted for the three components of the ARIMA model. The specific ARIMA model being used.

The parameters of the ARIMA model are defined as follows:

- **p:** The number of lag observations included in the model, also called the order of autoregression.
- **d:** The number of times that the raw observations are differenced.
- **q:** The size of the moving average window, also called the order of moving average.

A linear regression model is constructed including the specified number and type of terms, and the data is prepared by a degree of differencing in order to make it stationary, i.e. to remove trend and seasonal structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates to not use that element of the model. This way, the ARIMA model can be configured to perform the function of an ARMA model, and even a simple AR, I, or MA model.

Adopting an ARIMA model for a time series assumes that the underlying process that generated the observations is an ARIMA process. This may seem obvious, but helps to motivate the need to confirm the assumptions of the model in the raw observations and in the residual errors of forecasts from the model.

Next, let's take a look at how we can use the ARIMA model in Python. We will start with loading a simple univariate time series.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

[START MY EMAIL COURSE](#)

Stop learning Time Series Forecasting the slow way!

Take my free 7-day email course and discover how to stop learning Time Series Forecasting the slow way!

[Start Machine Learning](#)

Click to sign-up and also get a free PDF Ebook version of the course.

[Start Your FREE Mini-Course Now!](#)

Shampoo Sales Dataset

This dataset describes the monthly number of sales of shampoo over a 3 year period.

The units are a sales count and there are 36 observations. The original dataset is credited to Makridakis, Wheelwright, and Hyndman (1998).

- Download the dataset.

Download the dataset and place it in your current working directory as "shampoo-sales.csv".

Below is an example of loading the Shampoo Sales dataset into memory. Note the use of the parse_dates argument to convert the date-time field. The dataset is baselined in an arbitrary year of 1901.

```

1 from pandas import read_csv
2 from pandas import datetime
3 from matplotlib import pyplot
4
5 def parser(x):
6     return datetime.strptime('1901'+x, '%Y-%m')
7
8 series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser)
9 print(series.head())
10 series.plot()
11 pyplot.show()
```

Running the example prints the first 5 rows of the dataset.

1	Month
2	1901-01-01 266.0
3	1901-02-01 145.9
4	1901-03-01 183.1
5	1901-04-01 119.3
6	1901-05-01 180.3
7	Name: Sales, dtype: float64

The data is also plotted as a time series with the month along the x-axis and sales figures on the y-axis.

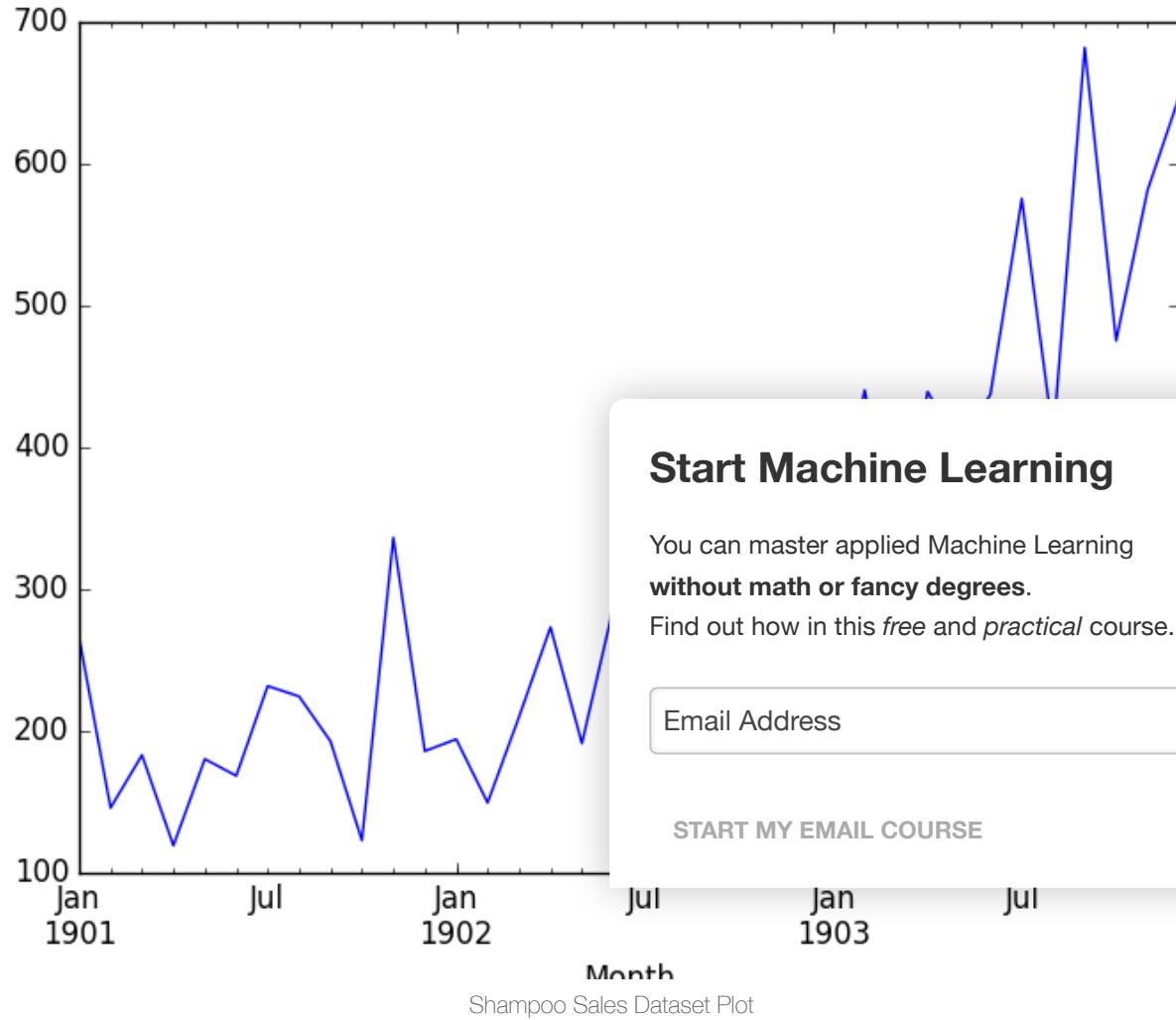
Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



We can see that the Shampoo Sales dataset has a clear trend.

This suggests that the time series is not stationary and will require differencing to make it stationary, at least a difference order of 1.

Let's also take a quick look at an autocorrelation plot of the time series. This is also built-in to Pandas. The example below plots the autocorrelation for a large number of lags in the time series.

```

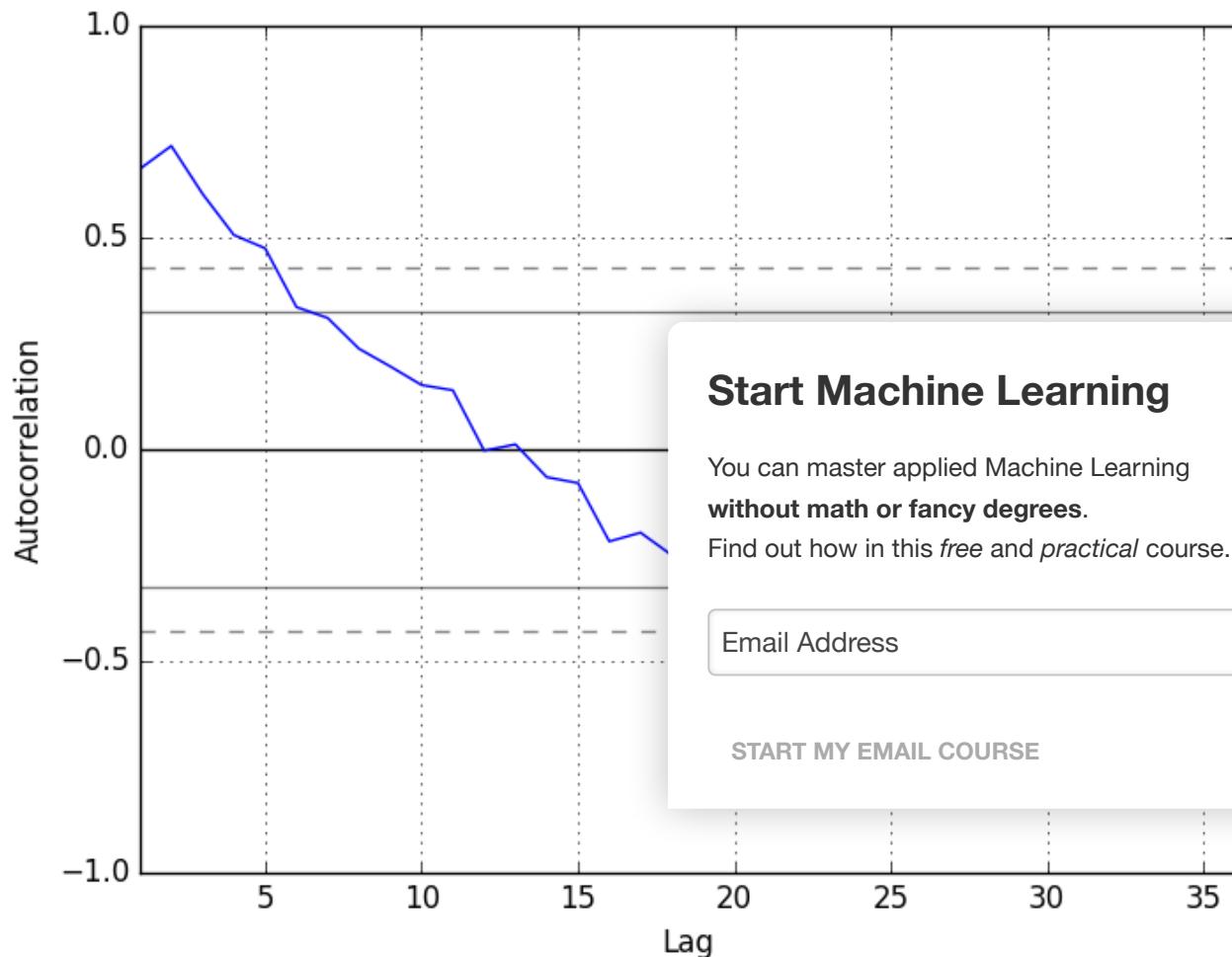
1 from pandas import read_csv
2 from pandas import datetime
3 from matplotlib import pyplot
4 from pandas.plotting import autocorrelation_plot
5
6 def parser(x):
7     return datetime.strptime('190'+x, '%Y-%m')
8
9 series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser)
10 autocorrelation_plot(series)
11 pyplot.show()

```

Running the example, we can see that there is a positive correlation with the first 10-to-12 lags that is perhaps significant for the first 5 lags.

[Start Machine Learning](#)

A good starting point for the AR parameter of the model may be 5.



Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

ARIMA with Python

The statsmodels library provides the capability to fit an ARIMA model.

An ARIMA model can be created using the statsmodels library as follows:

1. Define the model by calling `ARIMA()` and passing in the p , d , and q parameters.
2. The model is prepared on the training data by calling the `fit()` function.
3. Predictions can be made by calling the `predict()` function and specifying the index of the time or times to be predicted.

Let's start off with something simple. We will fit an ARIMA model to the entire Shampoo Sales dataset and review the residual errors.

First, we fit an ARIMA(5,1,0) model. This sets the lag value to 5 for autoregression, uses a difference order of 1 to make the time series stationary, and uses a mo

Start Machine Learning

When fitting the model, a lot of debug information is provided about the fit of the linear regression model. We can turn this off by setting the `disp` argument to 0.

```

1 from pandas import read_csv
2 from pandas import datetime
3 from pandas import DataFrame
4 from statsmodels.tsa.arima_model import ARIMA
5 from matplotlib import pyplot
6
7 def parser(x):
8     return datetime.strptime('190'+x, '%Y-%m')
9
10 series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, do
11 # fit model
12 model = ARIMA(series, order=(5,1,0))
13 model_fit = model.fit(disp=0)
14 print(model_fit.summary())
15 # plot residual errors
16 residuals = DataFrame(model_fit.resid)
17 residuals.plot()
18 pyplot.show()
19 residuals.plot(kind='kde')
20 pyplot.show()
21 print(residuals.describe())

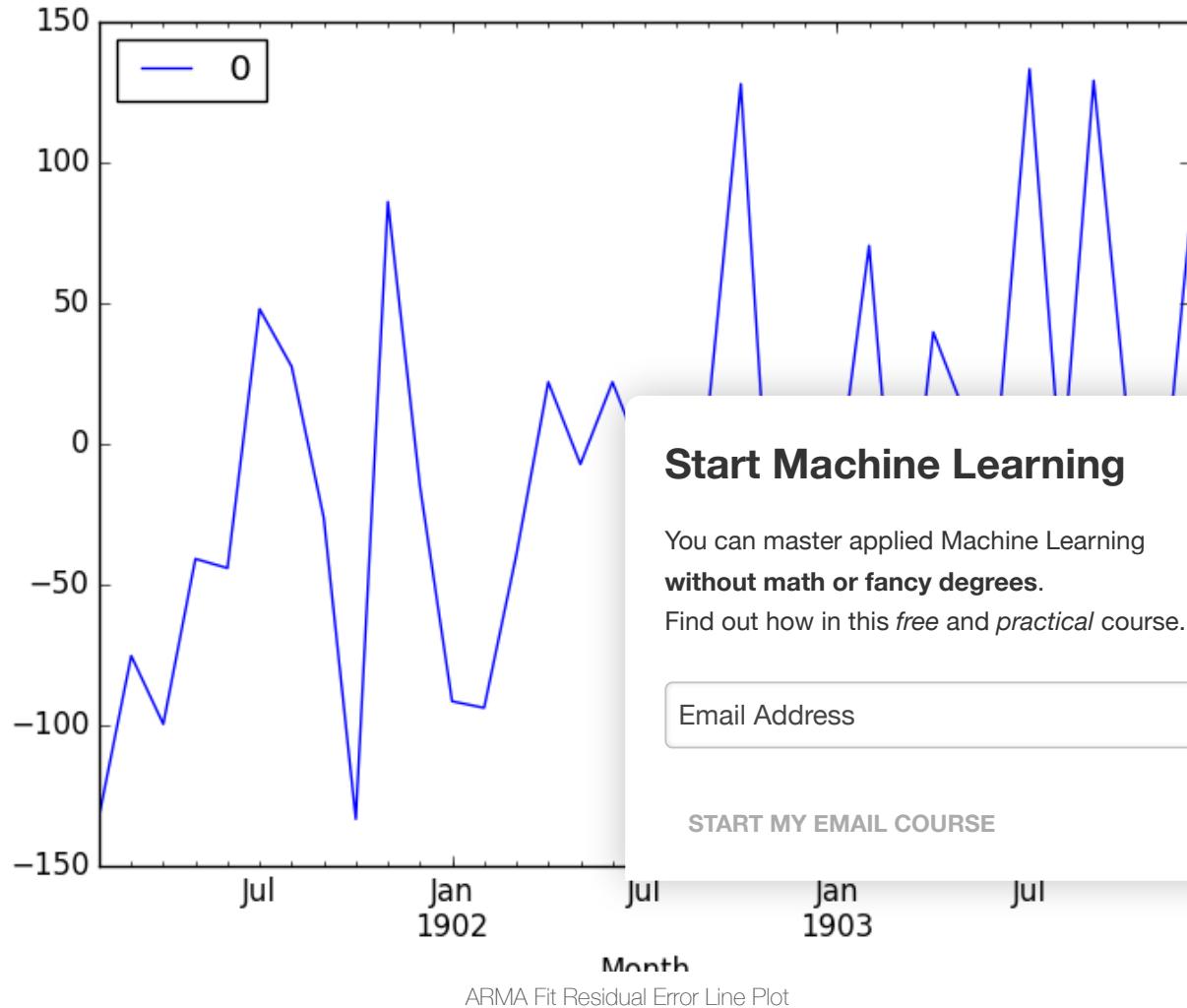
```

Running the example prints a summary of the fit model, as well as the skill of the fit on the in-sample observations.

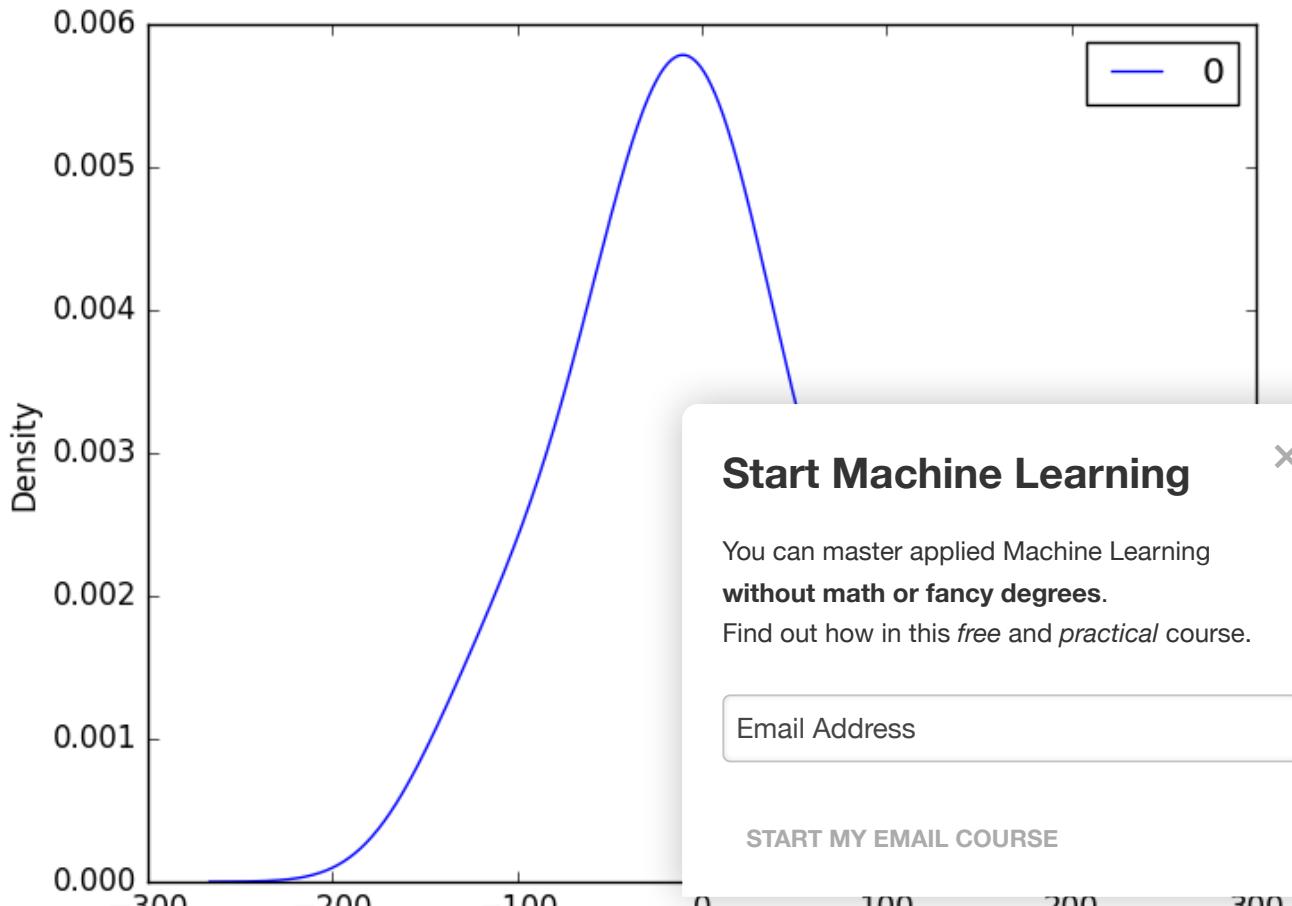
ARIMA Model Results						
1	Dep. Variable:	D.Sales	No. Obs:			
2	Model:	ARIMA(5, 1, 0)	Log Likelihood:	-196.170		
3	Method:	css-mle	S.D. of innovations:	64.241		
4	Date:	Mon, 12 Dec 2016	AIC:	406.340		
5	Time:	11:09:13	BIC:	417.227		
6	Sample:	02-01-1901	HQIC:	410.098		
7		- 12-01-1903				
8						
9						
10						
11	coef	std err	z	P> z	[95.0% Conf. Int.]	
12						
13	const	12.0649	3.652	3.304	0.003	4.908 19.222
14	ar.L1.D.Sales	-1.1082	0.183	-6.063	0.000	-1.466 -0.750
15	ar.L2.D.Sales	-0.6203	0.282	-2.203	0.036	-1.172 -0.068
16	ar.L3.D.Sales	-0.3606	0.295	-1.222	0.231	-0.939 0.218
17	ar.L4.D.Sales	-0.1252	0.280	-0.447	0.658	-0.674 0.424
18	ar.L5.D.Sales	0.1289	0.191	0.673	0.506	-0.246 0.504
19						
20						
21	Real	Imaginary		Modulus	Frequency	
22						
23	AR.1	-1.0617	-0.5064j	1.1763	-0.4292	
24	AR.2	-1.0617	+0.5064j	1.1763	0.4292	
25	AR.3	0.0816	-1.3804j	1.3828	-0.2406	
26	AR.4	0.0816	+1.3804j	1.3828	0.2406	
27	AR.5	2.9315	-0.0000j	2.9315	-0.0000	
28						

First, we get a line plot of the residual errors, suggesting that there may still be some trend information not captured by the model.

[Start Machine Learning](#)



Next, we get a density plot of the residual error values, suggesting the errors are Gaussian, but may not be centered on zero.



ARMA Fit Residual Error Density Plot

The distribution of the residual errors is displayed. The results show that indeed there is a bias in the prediction (a non-zero mean in the residuals).

1	count	35.000000
2	mean	-5.495213
3	std	68.132882
4	min	-133.296597
5	25%	-42.477935
6	50%	-7.186584
7	75%	24.748357
8	max	133.237980

Note, that although above we used the entire dataset for time series analysis, ideally we would perform this analysis on just the training dataset when developing a predictive model.

Next, let's look at how we can use the ARIMA model to make forecasts.

Rolling Forecast ARIMA Model

The ARIMA model can be used to forecast future time

[Start Machine Learning](#)

We can use the `predict()` function on the `ARIMAResults` object to make predictions. It accepts the index of the time steps to make predictions as arguments. These indexes are relative to the start of the training dataset used to make predictions.

If we used 100 observations in the training dataset to fit the model, then the index of the next time step for making a prediction would be specified to the prediction function as `start=101, end=101`. This would return an array with one element containing the prediction.

We also would prefer the forecasted values to be in the original scale, in case we performed any differencing ($d > 0$ when configuring the model). This can be specified by setting the `typ` argument to the value '`levels`': `typ='levels'`.

Alternately, we can avoid all of these specifications by step forecast using the model.

We can split the training dataset into train and test set prediction for each element on the test set.

A rolling forecast is required given the dependence on the AR model. A crude way to perform this rolling forecast is to receive each observation is received.

We manually keep track of all observations in a list called `history`, to which new observations are appended each iteration.

Putting this all together, below is an example of a rolling forecast with the ARIMA model in Python.

```

1 from pandas import read_csv
2 from pandas import datetime
3 from matplotlib import pyplot
4 from statsmodels.tsa.arima_model import ARIMA
5 from sklearn.metrics import mean_squared_error
6
7 def parser(x):
8     return datetime.strptime('190'+x, '%Y-%m')
9
10 series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True, date_parser=parser)
11 X = series.values
12 size = int(len(X) * 0.66)
13 train, test = X[0:size], X[size:len(X)]
14 history = [x for x in train]
15 predictions = list()
16 for t in range(len(test)):
17     model = ARIMA(history, order=(5,1,0))
18     model_fit = model.fit(disp=0)
19     output = model_fit.forecast()
20     yhat = output[0]
21     predictions.append(yhat)
22     obs = test[t]
23     history.append(obs)
24     print('predicted=%f, expected=%f' % (yhat, obs))
25 error = mean_squared_error(test, predictions)
26 print('Test MSE: %.3f' % error)
27 # plot

```

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

```
28 pyplot.plot(test)
29 pyplot.plot(predictions, color='red')
30 pyplot.show()
```

Running the example prints the prediction and expected value each iteration.

We can also calculate a final mean squared error score (MSE) for the predictions, providing a point of comparison for other ARIMA configurations.

```
1 predicted=349.117688, expected=342.300000
2 predicted=306.512968, expected=339.700000
3 predicted=387.376422, expected=440.400000
4 predicted=348.154111, expected=315.900000
5 predicted=386.308808, expected=439.300000
6 predicted=356.081996, expected=401.300000
7 predicted=446.379501, expected=437.400000
8 predicted=394.737286, expected=575.500000
9 predicted=434.915566, expected=407.600000
10 predicted=507.923407, expected=682.000000
11 predicted=435.483082, expected=475.300000
12 predicted=652.743772, expected=581.300000
13 predicted=546.343485, expected=646.900000
14 Test MSE: 6958.325
```

A line plot is created showing the expected values (blue) and the predictions (red). We can see the values show some trend and are in the same direction.

Start Machine Learning

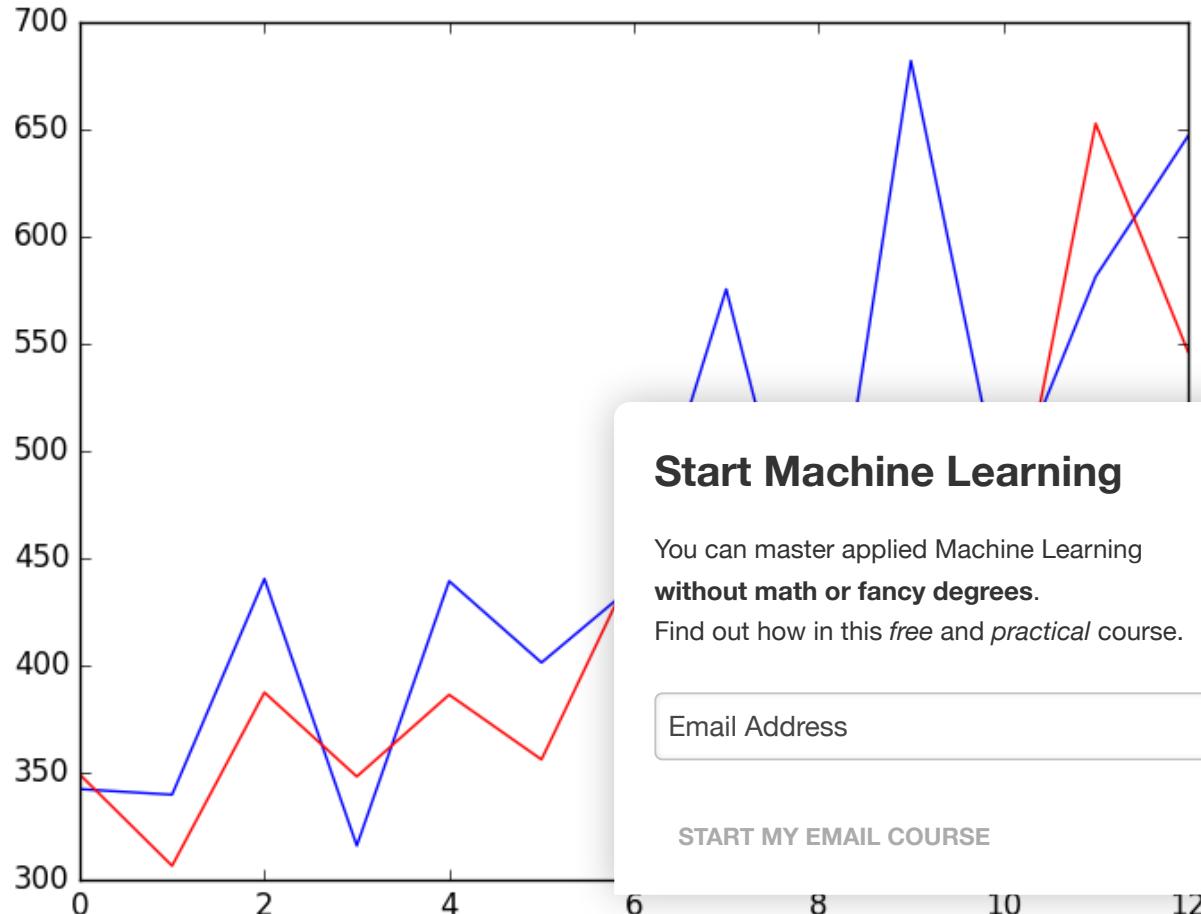
You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Start Machine Learning



Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

ARIMA Rolling Forecast Line Plot

The model could use further tuning of the p, d, and maybe even the q parameters.

Configuring an ARIMA Model

The classical approach for fitting an ARIMA model is to follow the [Box-Jenkins Methodology](#).

This is a process that uses time series analysis and diagnostics to discover good parameters for the ARIMA model.

In summary, the steps of this process are as follows:

- 1. Model Identification.** Use plots and summary statistics to identify trends, seasonality, and autoregression elements to get an idea of the amount of differencing and the size of the lag that will be required.
- 2. Parameter Estimation.** Use a fitting procedure to find the coefficients of the regression model.
- 3. Model Checking.** Use plots and statistical tests of the residual errors to determine the amount and type of temporal structure not captured by the model

Start Machine Learning

The process is repeated until either a desirable level of fit is achieved on the in-sample or out-of-sample observations (e.g. training or test datasets).

The process was described in the classic 1970 textbook on the topic titled [Time Series Analysis: Forecasting and Control](#) by George Box and Gwilym Jenkins. An updated 5th edition is now available if you are interested in going deeper into this type of model and methodology.

Given that the model can be fit efficiently on modest-sized time series datasets, grid searching parameters of the model can be a valuable approach.

For an example of how to grid search the hyperparameters of the ARIMA model, see the tutorial:

- [How to Grid Search ARIMA Model Hyperparameters](#)

Summary

In this tutorial, you discovered how to develop an ARIMA model for time series forecasting in Python.

Specifically, you learned:

- About the ARIMA model, how it can be configured, and how it can be used.
- How to perform a quick time series analysis using the ARIMA model.
- How to use an ARIMA model to forecast out of sample.

Do you have any questions about ARIMA, or about this tutorial?

Ask your questions in the comments below and I will do my best to answer.

Start Machine Learning

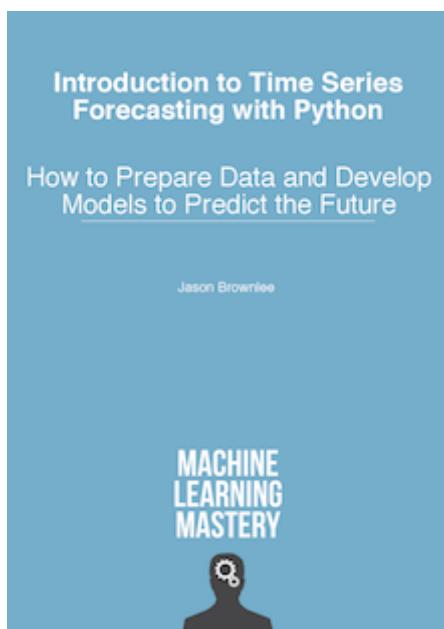
You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Want to Develop Time Series Forecasts with Python?



Develop Your Own Forecasts in Minutes

...with just a few lines of python code

Discover how in my new Ebook:

[Introduction to Time Series Forecasting With Python](#)

It covers **self-study tutorials** and **end-to-end projects** on topics like: *Loading data, visualization, modeling, algorithm tuning, and much more...*

Finally Bring Time Series Forecasting to Your Own Projects

Skip the Academics. Just Results.

[SEE WHAT'S INSIDE](#)

[Start Machine Learning](#)

[Tweet](#)[Share](#)[Share](#)

About Jason Brownlee

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on tutorials.

[View all posts by Jason Brownlee →](#)

◀ How to Visualize Time Series Residual Forecast Errors with Python

How to Model

656 Responses to *How to Create an ARIMA Model for Time Series Forecasting in Python*

SalemAmeen January 9, 2017 at 7:22 am #

Many thank

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

[START MY EMAIL COURSE](#)

Jason Brownlee January 9, 2017 at 7:53 am #

[REPLY ↗](#)

You're welcome.

Hugo Santillan April 10, 2019 at 1:52 pm #

[REPLY ↗](#)

Hi Jason! Great tutorial.

Just a reeal quick question ..how can I fit and run the last code for multiple varialbles?..the dataset that looks like this:

Date,CO,NO2,O3,PM10,SO2,Temperature
2016-01-01 00:00:00,0.615,0.01966,0.00761,49.92,0.00055,18.1

Jason Brownlee April 11, 2019 at 6:26 am #

[REPLY ↗](#)

You can model the target variable ^----^

[Start Machine Learning](#)

Alternately you can provide the other variables as exog variables, such as SARIMAX.

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

Finally, you could use a neural network:

https://machinelearningmastery.com/start-here/#deep_learning_time_series

Andre February 2, 2020 at 10:38 am #

Hey,

Nice article, it helped me a lot.

I have a question as to how to make predictions for new data that was not included in the training set. For each item in the test set, after a prediction is made, the prediction, along with the true value, is added to the history.

How can I make predictions when I do not have training data from?

Jason Brownlee February 3, 2020 at 10:30 pm #

Good question, see this tutorial:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

joseph November 5, 2019 at 8:22 pm #

REPLY ↗

I have a question am doing a project concerning data analytics insights for retail company sales case study certain supermarket in my area and am proposing to use ARIMA can it be appropriate and how can i apply it

Jason Brownlee November 6, 2019 at 6:32 am #

REPLY ↗

Perhaps start by modeling one product?

Jagan January 22, 2020 at 8:18 pm #

REPLY ↗

Hi Jason! Great Tutorial!!

I have a usecase of timeseries forecasting where I have to predict sales of different products out of the electronics store. There are around 300 types.

[Start Machine Learning](#)

sales on the next day for each of the product based on previous one year data. But not every product is being sold each day.

My guess is I have to create a tsa for each product. but the data quality for each product is low as not each product is being sold each day. And my use case is that I have to predict sales of each product.

Any way I can use time series on whole data without using tsa on each product individually?

Jason Brownlee January 23, 2020 at 6:29 am #

REPLY ↗

Good question, I have some suggestions:
<https://machinelearningmastery.com/faq-series-site-training-data/>
multiple-sites

Chris May 18, 2019 at 12:20 pm #

Hi Jason! Great tutorial.

I got a question that needs your kind help.

For some reason, I need to calculate residuals of a GARCH model. The calculated residuals are different of those directly from the R package such rugarch. I put the estimated parameters back to the model and use the training data to back out the residuals. How to get the starting residuals at $t=0$, $t=-1$ etc. Should I treat the fitted ARMA-GARCH just as an fitted ARMA model? In that case why we need to fit an ARMA-GARCH model to the training data.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee May 19, 2019 at 7:59 am #

REPLY ↗

Sorry, I'm not familiar with the "rugarch" package or how it functions.

A. Sharma August 1, 2019 at 6:51 am #

REPLY ↗

Hi Jason,

Could you do a GaussianProcess example with the same data. And compare the two- those two methods seem to be applicable to similar problems- I would love to see your insights.

Jason Brownlee August 1, 2019 at 6:57 am #

REPLY ↗

Thanks for the great suggestion. I hope

Start Machine Learning

a Sharma August 1, 2019 at 10:47 am #

REPLY ↩

Thanks. If you also did a comparative study of the two, that would be great- I realize that might be out of the regular, thought I'd still ask. Also can I sign up for email notification?

Jason Brownlee August 1, 2019 at 2:12 pm #

Thanks.

You can sign-up for notification about
<https://machinelearningmastery.com/>

Ahmed hesham March 16, 2020 at 7:52 pm #

Hello,

I have climate change data for the past 8 years and I want to forecast so I need at least climate data for 30 years which means I need data for the previous 22 years climate change using ARIMA based on the last 8 years. Thank you

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee March 17, 2020 at 8:13 am #

REPLY ↩

No, that would be way too much data. ARIMA is for small datasets – or at least the python implementation cannot handle much data.

Perhaps explore using a linear regression or other ML methods as a first step.

Blessing Ojeme January 9, 2017 at 1:20 pm #

REPLY ↩

Much appreciated, Jason. Keep them coming, please.

Jason Brownlee January 10, 2017 at 8:55 am #

REPLY ↩

Sure thing! I'm glad you're finding them useful.

What else would you like to see?

Start Machine Learning

Utkarsh July 22, 2017 at 10:31 pm #

REPLY ↗

Hi Jason ,can you suggest how one can solve time series problem if the target variable is categorical having around 500 categories.

Thanks

Jason Brownlee July 23, 2017 at 6:24 am #

REPLY ↗

That is a lot of categories.

Perhaps moving to a neural network type model would be appropriate given the vast amount of data to learn this problem.

Yash July 25, 2018 at 8:01 pm #

What if there are multiple columns in the data? For example, if I have a shampoo, there could be a column with item number, a column with price, a column with number of stores and finally a column with res

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee July 26, 2018 at 7:40 am #

REPLY ↗

If you have parallel input time series, you can use the other variables as exogenous variables. If you want to predict all variables, you can use VAR.

If you want to support multiple series generally as input, you can use ML methods, this will help as a start:

<https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/>

Aflal May 13, 2020 at 7:35 am #

REPLY ↗

OMG. Searching for weeks, never found an article like this one. Thank a lot.
I need your advice please,
I need to predict Retail sales data with variables like weather, sales Discount', holiday etc.

Which is the best model is to use? And why?

How can decide the best fit model?

(Can I use SARIMAX for this?)

Love from Sri Lanka

Start Machine Learning

Sorry for bad English

Jason Brownlee May 13, 2020 at 7:45 am #

REPLY ↩

You're welcome.

Perhaps test a few different models and discover what works best for your dataset.

Aflal May 16, 2020 at 6:10 am #

But I'm your suggestion tour
forecasting.

What is your suggestion??

Any link I could follow to find a solution
Thanks again

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee May 16, 2020

Perhaps try some of the techniques listed here.

https://machinelearningmastery.com/start-here/#deep_learning_time_series

Somayeh November 27, 2017 at 2:43 am #

REPLY ↩

Hi Jason,

Recently I am working on time series prediction, but my research is a little bit complicated for me to understand how to fix a time series models to predict future values of multi targets.

Recently I read your post in multi-step and multivariate time series prediction with LSTM. But my problem have a series input values for every time (for each second we have recorded more than 500 samples). We have 22 inputs and 3 targets. All the data has been collected during 600 seconds and then predict 3 targets for 600 next seconds. Please help me how can solve this problem?

It is noticed we have trend and seasonality pulses for targets during the time.

morteza February 19, 2018 at 2:58 am #

REPLY ↩

do you find a solution to this problem?

Start Machine Learning

Chow Xixi January 9, 2017 at 6:00 pm #

REPLY ↗

good, Has been paid close attention to your blog.

Jason Brownlee January 10, 2017 at 8:55 am #

REPLY ↗

Thanks!

Kevin January 17, 2017 at 12:58 am #

Gives me loads of errors:

Traceback (most recent call last):

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/date_parser.py", line 10, in date_parser
    date_parser(*date_cols), errors='ignore')
```

```
File "/Users/kevinoost/PycharmProjects/ARIMA/main.py", line 6, in parser
    return datetime.strptime('190' + x, '%Y-%m')
```

TypeError: strftime() argument 1 must be str, not numpy.datetime64

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 2285, in converter
    dayfirst=dayfirst),
```

```
File "pandas/src/inference.pyx", line 841, in pandas.lib.try_parse_dates (pandas/lib.c:57884)
```

```
File "pandas/src/inference.pyx", line 838, in pandas.lib.try_parse_dates (pandas/lib.c:57802)
```

```
File "/Users/kevinoost/PycharmProjects/ARIMA/main.py", line 6, in parser
    return datetime.strptime('190' + x, '%Y-%m')
```

```
File "/Users/kevinoost/anaconda/lib/python3.5/_strptime.py", line 510, in _strptime_datetime
    tt, fraction = _strptime(data_string, format)
```

```
File "/Users/kevinoost/anaconda/lib/python3.5/_strptime.py", line 343, in _strptime
    (data_string, format))
```

ValueError: time data '190Sales of shampoo over a three year period' does not match format '%Y-%m'

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "/Users/kevinoost/PycharmProjects/ARIMA/main.py", line 8, in parser
    series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True,
                      date_parser=date_parser)
```

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 562, in parser_
    return _read(filepath_or_buffer, kwds)
```

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 325, in _read
    return parser.read()
```

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 2285, in converter
    dayfirst=dayfirst),
```

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

```

ret = self._engine.read(nrows)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 1387, in read
index, names = self._make_index(data, alldata, names)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 1030, in
_make_index
index = self._agg_index(index)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 1111, in
_agg_index
arr = self._date_conv(arr)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 2288, in converter
return generic_parser(date_parser, *date_cols)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/date_converters.py", line 38, in
generic_parser
results[i] = parse_func(*args)
File "/Users/kevinoost/PycharmProjects/ARIMA/main.j
return datetime.strptime('190'+x, '%Y-%m')
File "/Users/kevinoost/anaconda/lib/python3.5/_strptime
tt, fraction = _strptime(data_string, format)
File "/Users/kevinoost/anaconda/lib/python3.5/_strptime
(data_string, format))
ValueError: time data '190Sales of shampoo over a thr
Process finished with exit code 1
Help would be much appreciated.

```

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Jason Brownlee January 17, 2017 at 7:39 am #

REPLY ↩

It looks like there might be an issue with your data file.

Open the csv in a text editor and confirm the header line looks sensible.

Also confirm that you have no extra data at the end of the file. Sometimes the datamarket files download with footer data that you need to delete.

Joseph Brown March 7, 2018 at 8:35 am #

REPLY ↩

Hi Jason,

I'm getting this same error. I checked the data and looks fine. I not sure what else to do, still learning. Please help.

Data

"Month";"Sales of shampoo over a three year period"
 "1-01";266.0

Start Machine Learning

```
"1-02";145.9
"1-03";183.1
"1-04";119.3
"1-05";180.3
"1-06";168.5
"1-07";231.8
"1-08";224.5
"1-09";192.8
"1-10";122.9
"1-11";336.5
"1-12";185.9
"2-01";194.3
"2-02";149.5
"2-03";210.1
"2-04";273.3
"2-05";191.4
"2-06";287.0
"2-07";226.0
"2-08";303.6
"2-09";289.9
"2-10";421.6
"2-11";264.5
"2-12";342.3
"3-01";339.7
"3-02";440.4
"3-03";315.9
"3-04";439.3
"3-05";401.3
"3-06";437.4
"3-07";575.5
"3-08";407.6
"3-09";682.0
"3-10";475.3
"3-11";581.3
"3-12";646.9
```

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Jason Brownlee March 7, 2018 at 3:02 pm #

REPLY ↗

The data you have pasted is separated by semicolons, not commas as expected.

AJ January 21, 2018 at 8:56 pm #

REPLY ↗

Start Machine Learning

Hi Kevin,

the last line of the data set, at least in the current version that you can download, is the text line “Sales of shampoo over a three year period”. The parser barfs on this because it is not in the specified format for the data lines. Try using the “nrows” parameter in read_csv.

```
series = read_csv('~/Downloads/shampoo-sales.csv', header=0, nrows=36, parse_dates=[0],  
index_col=0, squeeze=True, date_parser=parser)
```

worked for me.

Jason Brownlee January 22, 2018 at 4:43 am #

REPLY ↗

Great tip!

Serail April 7, 2018 at 4:13 pm #

Thanks for your excellent tip

Alex November 25, 2018 at 8:56 pm #

Thanks, had the same problem, worked!

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

NGUYEN Quang Anh January 19, 2017 at 6:28 pm #

REPLY ↗

Let say I have a time series data with many attribute. For example a row will have (speed, fuel, tire_pressure), how could we made a model out of this ? the value of each column may affect each other, so we cannot do forecasting on solely 1 column. I google a lot but all the example I've found so far only work on time series of 1 attribute.

Jason Brownlee January 20, 2017 at 10:19 am #

REPLY ↗

This is called multivariate time series forecasting. Linear models like ARIMA were not designed for this type of problem.

generally, you can use the lag-based representation of each feature and then apply a standard machine learning algorithm.

I hope to have some tutorials on this soon.

Start Machine Learning

rchesak May 30, 2017 at 12:37 pm #

REPLY ↗

Wanted to check in on this, do you have any tutorials on multivariate time series forecasting?

Also, when you say standard machine learning algorithm, would a random forest model work?

Thanks!

rchesak May 30, 2017 at 12:52 pm #

REPLY ↗

Update: the `statsmodels.tsa`.
says it takes the optional parameter `exog`,
optional array of exogenous variables'. This
agree?

I am trying to predict number of cases of a
data. So I believe the ARIMA model should

Thank you!

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee June 2, 2017 at 12:32 pm #

REPLY ↗

I have not experimented with this argument.

Jason Brownlee June 2, 2017 at 12:32 pm #

REPLY ↗

No multivariate examples at this stage.

Yes, any supervised learning method.

Muyi Ibidun February 7, 2017 at 9:36 am #

REPLY ↗

Hello Ng,

Your problem fits what VAR (Vector Autoregression) models is designed for. See the following links for more information. I hope this helps your work.

https://en.wikipedia.org/wiki/Vector_autoregression

http://statsmodels.sourceforge.net/devel/vector_ar.html

Start Machine Learning

Kelvid January 20, 2017 at 11:55 am #

REPLY ↗

Hi, would you have a example for the seasonal ARIMA post? I have installed latest statsmodels module, but there is an error of import the SARIMAX. Do help if you manage to figure it out. Thanks.

Jason Brownlee January 21, 2017 at 10:23 am #

REPLY ↗

Hi Kelvid, I don't have one at the moment. I'll prepare an example of SARIMAX and post it soon.

Muhammad Arsalan January 29, 2017 at 10:13 pm #

It is so informative..thankyou

Jason Brownlee February 1, 2017 at 10:16 am #

I'm glad to hear that Muhammad.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Sebastian January 31, 2017 at 3:33 am #

REPLY ↗

Great post Jason!

I have a couple of questions:

- Just to be sure. `model_fit.forecast()` is single step ahead forecasts and `model_fit.predict()` is for multiple step ahead forecasts?
- I am working with a series that seems at least quite similar to the shampoo series (by inspection). When I use predict on the training data, I get this zig-zag pattern in the prediction as well. But for the test data, the prediction is much smoother and seems to saturate at some level. Would you expect this? If not, what could be wrong?

Jason Brownlee February 1, 2017 at 10:28 am #

REPLY ↗

Hi Sebastian,

Yes, `forecast()` is for one step forecasts. You can do one step forecasts with `predict()` also, but it is more work.

I would not expect prediction beyond a few time s

Start Machine Learning

Sebastian February 3, 2017 at 9:25 am #

REPLY ↗

Thanks for the reply!

Concerning the second question. Yes, you are right the prediction is not very accurate. But moreover, the predicted time series has a totally different frequency content. As I said, it is smooth and not zig-zaggy as the original data. Is this normal or am I doing something wrong. I also tried the multiple step prediction (model_fit.predict()) on the training data and then the forecast seem to have more or less the same frequency content (more zig-zaggy) as the data I am trying to predict.

Jason Brownlee February 3, 2017 at 10:00 am #

Hi Sebastian, I see.

In the case of predicting on the training data, for example, if you predict the next 5 obs some time steps later, the model will predict t+5 rather than prediction(t+4).

In the case of predicting beyond the end of the training data, for example, if you make predictions (unless you provide them), it only makes sense to predict the next few time steps. The result is the errors compound over time.

Does that make sense/help?

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Sebastian February 3, 2017 at 6:34 pm #

That helped!

Thanks!

Jason Brownlee February 4, 2017 at 10:00 am #

Glad to hear it Sebastian.

satya May 22, 2017 at 9:19 pm #

Hi Jason,

I suppose my training set is 1949 to 1961. Can I use the data from 1970 until now for prediction? Predict function

Start Machine Learning

Thanks

Satya

Jason Brownlee May 23, 2017 at 7:51 am #

Yes, you would have to predict 10 years worth of data though. The predictions after 10 years would likely have a lot of error.

Ani July 6, 2018 at 1:22 am #

Hi Jason,

Continuing on this note, how far ahead can we predict with AR+GARCH or GARCH in Python? I'm guessing more than 10 years, but I'm not sure what the forecasting mechanism?

To give you a sense of my data, given the time series is non-stationary in terms of number of predictions can we expect jagged predictions in-sample, but essentially no semblance of that behavior and instead something smooth? That's what I think. Thanks!

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee July 6, 2018 at 6:43 am #

The skill of AR+GARH (or either) really depends on the choice of model parameters and on the specifics of the problem.

Perhaps you can try grid searching different parameters?

Perhaps you can review ACF/PACF plots for your data that may suggest better parameters?

Perhaps you can try non-linear methods?

Perhaps your problem is truly challenging/not predictable?

I hope that helps as a start.

Elliot January 31, 2017 at 10:07 am #

REPLY ↲

So this is building a model and then checking it off of the given data right?

-How can I predict what would come next after the last data point? Am I misunderstanding the code?

Start Machine Learning

Jason Brownlee February 1, 2017 at 10:34 am #

REPLY ↗

Hi Elliot,

You can predict the next data point at the end of the data by training on all of the available data then calling `model.forecast()`.

I have a post on how to make predictions here:

<http://machinelearningmastery.com/make-predictions-time-series-forecasting-python/>

Does that help?

Elliot February 1, 2017 at 11:44 am #

I tried the `model.forecast` at the end of the data.

"`AttributeError: 'ARIMA' object has no attribute 'forecast'`"

Also on your article: <http://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>

In step 3, when it says "Prediction: 46.755211" it uses the model to predict what would happen

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee February 2, 2017 at 1:58 pm #

REPLY ↗

Hi Elliot, the `forecast()` function is on the `ARIMAResults` object. You can learn more about it here:

http://statsmodels.sourceforge.net/stable/generated/statsmodels.tsa.arima_model.ARIMAResults.forecast.html

Muyi Ibidun February 7, 2017 at 9:38 am #

REPLY ↗

Thanks Jason for this post!

It was really useful. And your blogs are becoming a must read for me because of the applicable and piecemeal nature of your tutorials.

Keep up the good work!

Jason Brownlee February 7, 2017 at 10:25 am #

REPLY ↗

You're welcome, I'm glad to hear that.

Start Machine Learning

Kalin Stoyanov February 8, 2017 at 9:30 pm #

REPLY ↩

Hi,

This is not the first post on ARIMA, but it is the best so far. Thank you.

Jason Brownlee February 9, 2017 at 7:24 am #

REPLY ↩

I'm glad to hear you say that Kalin.

James Zhang February 10, 2017 at 7:42 pm #

Hey Jason,

thank you very much for the post, very good written! I model, but when I try to forecast the data that are out change history.append(obs) to history.append(yhat), any reason? and how do you actually do the out-of-sample dataset? Thank you very much!

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee February 11, 2017 at 5:00 am #

REPLY ↩

Hi james,

Each loop in the rolling forecast shows you how to make a one-step out of sample forecast.

Train your ARIMA on all available data and call forecast().

If you want to perform a multi-step forecast, indeed, you will need to treat prior forecasts as "observations" and use them for subsequent forecasts. You can do this automatically using the predict() function. Depending on the problem, this approach is often not skillful (e.g. a flat forecast).

Does that help?

James February 16, 2017 at 2:03 am #

REPLY ↩

Hi Jason,

thank you for your reply! so what could be the reason a flat forecast occurs and how to avoid it?

Start Machine Learning

Jason Brownlee February 16, 2017 at 11:09 am #

REPLY ↗

Hi James,

The model may not have enough information to make a good forecast.

Consider exploring alternate methods that can perform multi-step forecasts in one step – like neural nets or recurrent neural nets.

James February 16, 2017 at 7:41 pm #

Hi Jason,

thanks a lot for your information! still n

Jason Brownlee February 17,

I'm here to help James!

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Supriya February 16, 2017 at 1:27 am #

REPLY ↗

when i calculate train and test error , train rmse is greater than test rmse.. why is it so?

Jason Brownlee February 16, 2017 at 11:08 am #

REPLY ↗

I see this happen sometimes Supriya.

It suggests the model may not be well suited for the data.

Matias T February 18, 2017 at 12:04 am #

REPLY ↗

Hello Jason, thanks for this amazing post.

I was wondering how does the “size” work here. For example lets say i want to forecast only 30 days ahead. I keep getting problems with the degrees of freedom.

Could you please explain this to me.

Thanks

Start Machine Learning

REPLY ↩

Jason Brownlee February 18, 2017 at 8:40 am #

Hi Matias, the “size” in the example is used to split the data into train/test sets for model evaluation using walk forward validation.

You can set this any way you like or evaluate your model different ways.

To forecast 30 days ahead, you are going to need a robust model and enough historic data to evaluate this model effectively.

Matias R February 21, 2017 at 6:39 am #

I get it. Thanks Jason.

I was thinking, in this particular example, ¿will

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)
Jason Brownlee February 21, 2017

Great question Matias.

The amount of history is one variable to te

Design experiments to test if having more or less history improves performance.

ubald kuijpers February 24, 2017 at 10:05 pm #

REPLY ↩

Dear Jason,

Thank you for explaining the ARIMA model in such clear detail.

It helped me to make my own model to get numerical forecasts and store it in a database.

So nice that we live in an era where knowledge is de-mystified .

Jason Brownlee February 25, 2017 at 5:55 am #

REPLY ↩

I'm glad to here it!

Jacques Sauve February 25, 2017 at 6:41 am #

REPLY ↩

Hi Jason. Very good work!

It would be great to see how forecasting models can b

[Start Machine Learning](#)

Jason Brownlee February 26, 2017 at 5:26 am #

REPLY ↗

Great suggestion, thanks Jacques.

Mehran March 1, 2017 at 12:56 am #

REPLY ↗

Hi there. Many thanks. I think you need to change the way you parse the datetime to:

`datetime.strptime('19'+x, '%Y-%b')`

Many thanks

Jason Brownlee March 1, 2017 at 8:41 am #

Are you sure?

See this list of abbreviations:

<https://docs.python.org/3/library/datetime.html#strftime-strptime-behavior>

The "%m" refers to "Month as a zero-padded decimal number".

See a sample of the raw data file:

```
1 "Month","Sales"
2 "1-01",266.0
3 "1-02",145.9
4 "1-03",183.1
5 "1-04",119.3
6 "1-05",180.3
```

The "%b" refers to "Month as locale's abbreviated name." which we do not have here.

Niirkshith March 6, 2017 at 4:49 pm #

REPLY ↗

Hi Jason,

Lucky i found this at the begining of my project.. Its a great start point and enriching.
Keep it coming :).

This can also be used for non linear time series as well?

Thanks,
niri

Jason Brownlee March 7, 2017 at 9:31 am #

Start Machine Learning

Glad to hear it Niirkshith.

Try and see.

Anthony of Sydney March 8, 2017 at 9:00 am #

REPLY ↩

Dear Dr Jason,

In the above example of the rolling forecast, you used the rmse of the predicted and the actual value.

Another way of getting the residuals of the model is to get the std devs of the residuals of the fitted model

```
1 model = modelling.ARIMA(data, (ar,diff,ma));
2 model_fit = model.fit()
3 residuals = pd.DataFrame(model_fit.resid)
4 #calculate the std dev of the residuals, we
5 print("the std dev of the residuals = %f" %
6 residuals.plot()
```

Question, is the std dev of the residuals the same as the rmse?

Thank you

Anthony of Sydney NSW

what is the difference between measuring the std deviation of the residuals and the rmse of the rolling forecast will

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Jason Brownlee March 8, 2017 at 9:50 am #

REPLY ↩

No, they are not the same.

See this post on performance measures:

<http://machinelearningmastery.com/time-series-forecasting-performance-measures-with-python/>

The RMSE is like the average residual error, but not quite because of the square and square root that makes the result positive.

Niirkshith March 10, 2017 at 1:28 pm #

REPLY ↩

Hi Jason,

Great writeup, had a query, when u have a seasonal data and do seasonal differencing. i.e for ex $y(t) = y(t) - y(t-12)$ for yearly data. What will be the value of d in ARIMA(p,d,q).

Niirkshith March 10, 2017 at 1:29 pm #

REPLY ↩

typo, ex $y(t) = y(t) - y(t-12)$ for monthly data

Start Machine Learning

Jason Brownlee March 11, 2017 at 7:56 am #

REPLY ↩

Great question Niirkshith.

ARIMA will not do seasonal differencing (there is a version that will called SARIMA). The d value on ARIMA will be unrelated to the seasonal differencing and will assume the input data is already seasonally adjusted.

Niirkshith March 13, 2017 at 1:09 pm #

REPLY ↩

Thanks for getting back.

ivan March 19, 2017 at 5:17 am #

Hi, Jason

thanks for this example. My question how is chosen the best Ivan

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Jason Brownlee March 19, 2017 at 9:11 am #

REPLY ↩

You can use ACF and PACF plots to help choose the values for p and q.

See this post:

<http://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Narbukra March 30, 2017 at 4:21 am #

REPLY ↩

Hi Jason, I am wondering if you did a similar tutorial on multi-variate time series forecasting?

Jason Brownlee March 30, 2017 at 8:57 am #

REPLY ↩

Not yet, I am working on some.

Nirikshith May 12, 2017 at 1:02 pm #

Start Machine Learning

Hi Jason,
any updates on the same

Shruti June 8, 2018 at 6:54 pm #

REPLY ↩

Hi Jason,
Nice post.

Can you please suggest how should I resolve this error: LinAlgError: SVD did not converge
I have a univariate time series.



Jason Brownlee June 9, 2018 at 6:

Sounds like the data is not a good fit for ARIMA. It may have another quirk.

David March 30, 2017 at 8:53 am #

Hi Jason,

Thanks for the great post! It was very helpful. I'm currently trying to forecast with the ARIMA model using order (4, 1, 5) and I'm getting an error message "The computed initial MA coefficients are not invertible. You should induce invertibility, choose a different model order, or you can pass your own start_params." The model works when fitting, but seems to error out when I move to model_fit = model.fit(disp=0). The forecast works well when using your parameters of (0, 1, 5) and I used ACF and PACF plots to find my initial p and q parameters. Any ideas on the cause/fix for the error? Any tips would be much appreciated.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

mostafa kotb October 17, 2017 at 4:52 am #

REPLY ↩

i have the same problem as yours, i use ARIMA with order (5,1,2) and i have been searching for a solution, but still couldn't find it.

Vit January 30, 2019 at 9:13 pm #

REPLY ↩

Hi, I have exactly the same problem. Have you already found any solution to that?
Thank you for any information,
Vit

Start Machine Learning



Jason Brownlee January 31, 2019 at 5:32 am #

REPLY ↗

Perhaps try a different model configuration?

Long August 8, 2020 at 12:06 pm #

Sorry, it is difficult for (3,1,3) as well.

It worked for prediction for the first step of the test data, but gave out the error on the second prediction step.

My code is as follow:

```

1 train_size = 0.66
2 series = shampoo_data["Sales"]
3 train, test = series[:int(series.size * train_size)]
4 predictions = []
5 true_values = []
6 history = [x for x in train]
7 for t in range(test.size):
8     model = ARIMA(history, order=(3,1,3))
9     model_fit = model.fit(disp=0)
10    forecast_results = model_fit.forecast()
11    pred = forecast_results[0]
12    std_error = forecast_results[1]
13    ci = forecast_results[2]
14    obs = test[t]
15    import pdb; pdb.set_trace()
16    print("Predicted: %.3f +- %.3f (CI: %.3f, %.3f), Expected: %.3f" % (pred,
17        std_error, ci[0], ci[1], obs))
18    predictions.append(pred)
19    true_values.append(obs)
20    history.append(obs)
21 rmse = mean_squared_error(true_values, predictions, squared=False)
22 mse = mean_squared_error(true_values, predictions)
23 print("RMSE: %.3f" % rmse)
24 print("MSE: %.3f" % mse)
25 fig = plt.figure(figsize=(20,10))
26 months = shampoo_data["Month"].values[int(series.size * train_size):]
27 plt.plot(months, test, label="true_values")
28 plt.plot(months, predictions, color="red", label="predictions")
29 plt.legend()
30 plt.show()
```

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

tom reilly April 27, 2017 at 6:39 am #

REPLY ↗

It's a great blog that you have, but the PACF determines the AR order not the ACF.



Jason Brownlee April 27, 2017 at 8:49 am #

REPLY ↗

Thanks Tom.

Start Machine Learning

I believe ACF and PACF both inform values for q and p:

<http://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Evgeniy May 2, 2017 at 1:22 am #

REPLY ↗

Good afternoon!

Is there an analog to the function auto.arima in the package for python from the package of the language R.
For automatic selection of ARIMA parameters?

Thank you!



Jason Brownlee May 2, 2017 at 6:00 am #

Yes, you can grid search yourself, see how

<http://machinelearningmastery.com/grid-search-arima-models/>

timer May 18, 2017 at 7:23 pm #

Hi. Great one. Suppose I have multiple airlines on daily basis. Now I want to predict for each airline number of possible passengers over next few months. How can I fit these time series models. Separate model for each airline or one single model?

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee May 19, 2017 at 8:16 am #

REPLY ↗

Try both approaches and double down on what works best.

Kashif May 26, 2017 at 2:06 am #

REPLY ↗

Hi Jason, if in my dataset, my first column is date (YYYYMMDD) and second column is time (hhmmss) and third column is value at given date and time. So could I use ARIMA model for forecasting such type of time series ?



Jason Brownlee June 2, 2017 at 11:47 am #

REPLY ↗

Yes, use a custom parse function to combine the date and time into one index column.

Start Machine Learning

Ashwini March 26, 2020 at 5:39 pm #

REPLY ↗

I have very similar data set. So how to train arima/sarima single model with above kind of data, i.e.. multiple data points at each timestep?



Jason Brownlee March 27, 2020 at 6:04 am #

REPLY ↗

I'm not sure these models can support data of that type.

Perhaps start here:

https://machinelearningmastery.com/start-here/#deep_learning_time_series

Kashif May 25, 2017 at 6:30 pm #

Hi Sir, Do you have tutorial about vector auto forecasting?)



Jason Brownlee June 2, 2017 at 11:42 am #

Not at the moment.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Ebrahim Aly May 30, 2017 at 5:03 am #

REPLY ↗

Thanks a lot, Dr. Jason. This tutorial explained a lot. But I tried to run it on an oil prices data set from Bp and I get the following error:

SVD did not converge

I used $(p,d,q) = (5, 1, 0)$

Would you please help me on solving or at least understanding this error?



Jason Brownlee June 2, 2017 at 12:29 pm #

REPLY ↗

Perhaps consider rescaling your input data and explore other configurations?

Alex June 9, 2017 at 8:01 am #

Start Machine Learning

Hi Jason,

I have a general question about ARIMA model in the case of multiple Time Series:
suppose you have not only one time series but many (i.e. the power generated per hour at 1000 different wind farms). So you have a dataset of 1000 time series of N points each and you want to predict the next N+M points for each of the time series.

Analyzing each time series separately with the ARIMA could be a waste. Maybe there are similarities in the time evolution of these 1000 different patterns which could help my predictions. What approach would you suggest in this case?



Jason Brownlee June 10, 2017 at 8:11 am #

REPLY ↗

You could not use ARIMA.

For linear models, you could use vector autoregression.

For nonlinear methods, I'd recommend a neural network.

I hope that helps as a start.

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee June 14, 2017 at 8:45 am #

REPLY ↗

Do you mean multiple series?

See VAR:

http://www.statsmodels.org/dev/vector_ar.html

TaeWoo Kim June 23, 2017 at 3:22 am #

REPLY ↗

"First, we get a line plot of the residual errors, suggesting that there may still be some trend information not captured by the model."

So are you looking for a smooth flat line in the curve?



Jason Brownlee June 23, 2017 at 6:47 am #

REPLY ↗

No, the upward trend that appears to exist

Start Machine Learning

Ukesh June 24, 2017 at 12:37 am #

REPLY ↗

At the end of the code, when I tried to print the predictions, it printed as the array, how do I convert it to the data points???

```
print(predictions)
```

```
[array([ 309.59070719]), array([ 388.64159699]), array([ 348.77807261]), array([ 383.60202178]), array([ 360.99214813]), array([ 449.34210105]), array([ 395.44928401]), array([ 434.86484106]), array([ 512.30201612]), array([ 428.59722583]), array([ 625.99359188]), array([ 543.53887362])]
```

Ukesh June 24, 2017 at 12:53 am #

Never mind.. I figured it out...

```
forecasts = numpy.array(predictions)
```

```
[[ 309.59070719]
 [ 388.64159699]
 [ 348.77807261]
 [ 383.60202178]
 [ 360.99214813]
 [ 449.34210105]
 [ 395.44928401]
 [ 434.86484106]
 [ 512.30201612]
 [ 428.59722583]
 [ 625.99359188]
 [ 543.53887362]]
```

Keep up the good work Jason.. Your blogs are extremely helpful and easy to follow.. Loads of appreciation..

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee June 24, 2017 at 8:03 am #

REPLY ↗

Glad to hear it.

Vincent June 29, 2017 at 6:53 pm #

REPLY ↗

Hi Jason and thank you for this post, its really helpful!

I have one question regarding ARIMA computation time.

Start Machine Learning

I'm working on a dataset of 10K samples, and I've tried rolling and "non rolling" (where coefficients are only estimated once or at least not every new sample) forecasting with ARIMA :

- rolling forecast produces good results but takes a big amount of time (I'm working with an old computer, around 3/6h depending on the ARMA model);
- "non rolling" doesn't forecast well at all.

Re-estimating the coefficients for each new sample is the only possibility for proper ARIMA forecasting?

Thanks for your help!



Jason Brownlee June 30, 2017 at 8:11 am #

REPLY ↗

I would focus on the approach that gives get caught up on "proper".

Kashif July 12, 2017 at 11:29 pm #

Dear Respected Sir, I have tried to use ARIMA/ are following,

YYYYMMDD hhmmss Duration

20100916 130748 18

20100916 131131 99

20100916 131324 214

20100916 131735 72

20100916 135342 37

20100916 144059 250

20100916 150148 87

20100916 150339 0

20100916 150401 180

20100916 154652 248

20100916 183403 0

20100916 210148 0

20100917 71222 179

20100917 73320 0

20100917 81718 25

20100917 93715 15

But when I used ARIMA model for such type of dataset, the prediction was very bad and test MSE was very high as well, My dataset has irregular pattern and autocorrelation is also very low. so could ARIMA model be used for such type of dataset ? or I have to do some modification in my dataset for using ARIMA model? Looking forward.

Thanks

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 13, 2017 at 9:56 am #

REPLY ↗

Perhaps try data transforms?

Perhaps try other algorithms?

Perhaps try gathering more data.

Here are more ideas:

<http://machinelearningmastery.com/machine-learning-performance-improvement-cheat-sheet/>

Vaibhav Agarwal July 14, 2017 at 6:53 am #

REPLY ↗

Hi Jason,

```
def parser(x):
    return datetime.strptime('190'+x, '%Y-%m')
series = read_csv('/home/administrator/Downloads/sh
index_col=0, squeeze=True, date_parser=parser)
print(series.head())
```

for these lines of code, I'm getting the following error

ValueError: time data '190Sales of shampoo over a thr

Please help.

Thanks

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 14, 2017 at 8:37 am #

REPLY ↗

Check that you have deleted the footer in the raw data file.

Kushal July 14, 2017 at 6:53 pm #

REPLY ↗

Hi Jason

Does ARIMA have any limitations for size of the sample. I have a dataset with 18k rows of data, ARIMA just doesn't complete.

Thanks

Kushal

Jason Brownlee July 15, 2017 at 9:41 am #

Start Machine Learning



Yes, it does not work well with lots of data (linalg methods under the covers blow up) and it can take forever as you see.

You could fit the model using gradient descent, but not with statsmodels, you may need to code it yourself.

Olivia July 18, 2017 at 4:51 am #

REPLY ↗

Love this. The code is very straightforward and the explanations are nice.

I would like to see a HMM model on here. I have been struggling with a few different packages (pomegranate and hmmlearn) for some time now. would like to see what you can do with it! (particularly a stock market example)



Jason Brownlee July 18, 2017 at 8:48 am #

Thanks Olivia, I hope to cover HMMs in th

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Ben July 19, 2017 at 11:27 am #

Good evening,

In what I am doing, I have a training set and a test set. In the training set, I am fitting an ARIMA model, let's say ARIMA(0,1,1) to the training set. What I want to do is use this model and apply it to the test set to get the residuals.

So far I have:

```
model = ARIMA(data,order = (0,1,1))
model_fit = model.fit(disp=0)
res = model_fit.resid
```

This gives me the residuals for the training set. So I want to apply the ARIMA model in 'model' to the test data.

Is there a function to do this?

Thank you



Jason Brownlee July 19, 2017 at 4:09 pm #

REPLY ↗

Hi Ben,

You could use your fit model to make a prediction for the test dataset then compare the predictions vs the real values to calculate the residual errors.

Start Machine Learning

Ben July 19, 2017 at 11:04 pm #

REPLY ↗

Could you give me an example of the syntax? I understand that idea, but when I would try the results were very poor.



Jason Brownlee July 20, 2017 at 6:20 am #

REPLY ↗

I provide a suite of examples, please search the blog for ARIMA or start here:
<http://machinelearningmastery.com/start-here/#timeseries>

Shaun July 27, 2017 at 9:29 am #

Hi Jason,

In your example, you append the real data set to the history before prediction?

history.append(obs), where obs is test[t].

In a real example, you don't have access to the real "future" data, so what do you do if dates beyond the data given in the csv, the results are

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 28, 2017 at 8:25 am #

REPLY ↗

We are doing walk-forward validation.

In this case, we are assuming that the real ob is made available after the prediction is made and before the next prediction is required.

Jai July 31, 2017 at 3:59 pm #

REPLY ↗

Hi,

How i do fix following error ?

```
-----
ImportError Traceback (most recent call last)
in ()
6 #fix deprecated - end
7 from pandas import DataFrame
--> 8 from statsmodels.tsa.arima_model import ARIM
```

Start Machine Learning

9

```
10 def parser(x):
```

ImportError: No module named 'statsmodels'

i have already install the statsmodels module.

```
(py_env) E:\WinPython-64bit-3.5.3.1Qt5_2\virtual_env\scikit-learn>pip3 install --upgrade "E:\WinPython\packages\statsmodels-0.8.0-cp35-cp35m-win_amd64.whl"
Processing e:\winpython\packages\statsmodels-0.8.0-cp35-cp35m-win_amd64.whl
Installing collected packages: statsmodels
Successfully installed statsmodels-0.8.0
```

<http://www.lfd.uci.edu/~gohlke/pythonlibs/>

Start Machine Learning



Jai July 31, 2017 at 5:25 pm #

problem fixed,

```
from statsmodels.tsa.arima_model import ARIMA
#this must come after statsmodels.tsa.arima_mod
from matplotlib import pyplot
```

You can master applied Machine Learning

without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

REPLY ↩



Jason Brownlee August 1, 2017 at 7:52 am #

Glad to hear it.

REPLY ↩



Jason Brownlee August 1, 2017 at 7:50 am #

It looks like statsmodels was not installed correctly or is not available in your current environment.

You installed using pip3, are you running a python3 env to run the code?

REPLY ↩

Jai August 1, 2017 at 4:18 pm #

interestingly, under your Rolling Forecast ARIMA Model explanation, matplotlib was above statsmodels.

```
from matplotlib import pyplot
from statsmodels.tsa.arima_model import ARIMA
```

i am using jupyter notebook from WinPython-64bit-3.5.3.1Qt5_2\virtual_env\scikit-learn>pip3 install --upgrade "E:\WinPython\packages\statsmodels-0.8.0-cp35-cp35m-win_amd64.whl"

Start Machine Learning

explanation

```
from matplotlib import pyplot
from pandas import DataFrame
from statsmodels.tsa.arima_model import ARIMA
```

Jai August 1, 2017 at 4:21 pm #

REPLY ↗

i think it could be i need to restart the virtual environment to let the environment recognize it, today i re-test the following declarations it is ok.

```
from matplotlib import pyplot
from pandas import DataFrame
from statsmodels.tsa.arima_model import ARIMA
```

thanks for the replies. case close



Jason Brownlee August 2, 2017 at 7:46 am #

X

Glad to hear it.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 2, 2017 at 7:46 am #

REPLY ↗

You will need to install statsmodels.

Fathi July 31, 2017 at 5:44 pm #

REPLY ↗

Great explanation

can anyone help me to write code in R about forecasting such as (50,52,50,55,57) i need to forecasting the next 3 hour, kindly help me to write code using R with ARIMA and SARIMA Model
thanks in advance



Jason Brownlee August 1, 2017 at 7:55 am #

REPLY ↗

I have a good list of books to help you with ARIMA in R here:

<http://machinelearningmastery.com/books-on-time-series-forecasting-with-r/>

Start Machine Learning

Fathi August 9, 2017 at 10:49 pm #

REPLY ↗

Dear :sir

i hope all of you fine

could any help me to analysis my data I will pay for him
if u can help me plz contact me fathi_nias@yahoo.com
thanks



Jason Brownlee August 10, 2017 at 6:57 am #

REPLY ↗

Consider hiring someone on upwork.com

Quentin August 11, 2017 at 10:37 pm #

Can the ACF be shown using bars so you can see the MA model? Or have you done a tutorial on interpreting

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 12, 2017 at 6:50 am #

REPLY ↗

Yes, consider using the blog search. Here it is:

<http://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Amritanshu August 18, 2017 at 8:20 pm #

REPLY ↗

Hi Jason

I am getting the error when trying to run the code:

```
from matplotlib import pyplot
from pandas import DataFrame
from pandas.core import datetools
from pandas import read_csv
from statsmodels.tsa.arima_model import ARIMA

series = read_csv('sales-of-shampoo-over-a-three-year.csv', header=0, parse_dates=[0], index_col=0)
# fit model
model = ARIMA(series, order=(0, 0, 0))
model_fit = model.fit(disp=0)
print(model_fit.summary())
# plot residual errors
```

Start Machine Learning

```
residuals = DataFrame(model_fit.resid)
residuals.plot()
pyplot.show()
residuals.plot(kind='kde')
pyplot.show()
print(residuals.describe())
```

Error Mesg on Console :

```
C:\Python36\python.exe C:/Users/aamrit/Desktop/untitled1/am.py
C:/Users/aamrit/Desktop/untitled1/am.py:3: FutureWarning: The pandas.core.datetools module is
deprecated and will be removed in a future version. Please use the pandas.tseries module instead.
from pandas.core import datetools
```

Traceback (most recent call last):

```
File "C:\Python36\lib\site-packages\pandas\core\tools\numerics.py", line 11, in _convert_listlike
    values, tz = tslib.datetime_to_datetime64(arg)
File "pandas\_libs\tslib.pyx", line 1810, in pandas._libs.tslib.datetime_to_datetime64
(pandas\_libs\tslib.c:33275)
TypeError: Unrecognized value type:
```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "C:\Python36\lib\site-packages\statsmodels\tsa\base.tsmodel.py", line 11, in _convert_listlike
    dates = to_datetime(dates)
File "C:\Python36\lib\site-packages\pandas\core\tools\numerics.py", line 11, in _convert_listlike
    result = _convert_listlike(arg, box, format, name=arg.name)
File "C:\Python36\lib\site-packages\pandas\core\tools\datetimes.py", line 447, in _convert_listlike
    raise e
```

```
File "C:\Python36\lib\site-packages\pandas\core\tools\datetimes.py", line 435, in _convert_listlike
    require_iso8601=require_iso8601
```

```
File "pandas\_libs\tslib.pyx", line 2355, in pandas._libs.tslib.array_to_datetime
(pandas\_libs\tslib.c:46617)
File "pandas\_libs\tslib.pyx", line 2538, in pandas._libs.tslib.array_to_datetime
(pandas\_libs\tslib.c:45511)
File "pandas\_libs\tslib.pyx", line 2506, in pandas._libs.tslib.array_to_datetime
(pandas\_libs\tslib.c:44978)
File "pandas\_libs\tslib.pyx", line 2500, in pandas._libs.tslib.array_to_datetime
(pandas\_libs\tslib.c:44859)
File "pandas\_libs\tslib.pyx", line 1517, in pandas._libs.tslib.convert_to_tsobject
(pandas\_libs\tslib.c:28598)
File "pandas\_libs\tslib.pyx", line 1774, in pandas._libs.tslib._check_dts_bounds
(pandas\_libs\tslib.c:32752)
```

```
pandas._libs.tslib.OutOfBoundsDatetime: Out of bounds nanosecond timestamp: 1-01-01 00:00:00
```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "C:/Users/aamrit/Desktop/untitled1/am.py", line 9, in _fit
    model = ARIMA(series, order=(0, 0, 0))
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 997, in __new__
    return ARMA(endog, (p, q), exog, dates, freq, missing)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 452, in __init__
    super(ARMA, self).__init__(endog, exog, dates, freq, missing)
```

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

```
File "C:\Python36\lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 44, in __init__
    self._init_dates(dates, freq)
File "C:\Python36\lib\site-packages\statsmodels\tsa\base\tsa_model.py", line 58, in _init_dates
    raise ValueError("Given a pandas object and the index does ")
ValueError: Given a pandas object and the index does not contain dates
```

Process finished with exit code 1



Jason Brownlee August 19, 2017 at 6:17 am #

REPLY ↗

Ensure you have removed the footer data from the CSV data file.

Amritanshu August 18, 2017 at 11:44 pm #

Hi Jason

Please help me to resolve the error

I am getting error :

Traceback (most recent call last):

```
File "C:/Users/aamrit/Desktop/untitled1/am.py", line 10, in <module>
    model_fit = model.fit(disp=0)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 1151, in fit
    callback, start_ar_lags, **kwargs)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 956, in fit
    start_ar_lags)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 578, in _fit_start_params
    start_params = self._fit_start_params_hr(order, start_ar_lags)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 508, in _fit_start_params_hr
    endog == np.dot(exog, ols_params).squeeze()
TypeError: Cannot cast ufunc subtract output from dtype('float64') to dtype('int64') with casting rule 'same_kind'
```

Code :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from datetime import datetime
from statsmodels.tsa.arima_model import ARIMA

data = pd.read_csv('AirPassengers.csv', header=0, parse_dates=[0], index_col=0)
model = ARIMA(data, order=(1,1,0), exog=None, dates=None, freq=None, missing='none')
model_fit = model.fit(disp=0)
print(model_fit.summary())
```

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 19, 2017 at 6:21 am #

REPLY ↗

Sorry, I have not seen this error before, consider posting to stack overflow.

kyci November 27, 2017 at 6:16 pm #

REPLY ↗

It is a bug in statsmodels. You should convert the integer values in 'data' to float first (e.g., by using np.float()).



Jason Brownlee November 28, 2017 at 8:30 am #

Great tip.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Vicente Queiroz March 30, 2018 at 10:30 pm #

@kyci is correct as you can check
<https://github.com/statsmodels/statsmodels>

I was following this tutorial for my dataset, and what fixed my problem was just converting to float, like this:

```
X = series.values
X = X.astype('float32')
```

Anup May 18, 2018 at 11:04 pm #

REPLY ↗

How can I add multiple EXOG variables in the model?

Amritanshu August 29, 2017 at 8:00 pm #

REPLY ↗

Jason, I am able to implement the model but the results are very vague for the predicted....

how to find the exact values for p,d and q ?



Jason Brownlee August 30, 2017 at 6:14 am #

REPLY ↗

Start Machine Learning

My best advice is to use a grid search for those parameters:

<https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/>

Amritanshu August 30, 2017 at 8:13 pm #

REPLY ↗

Thanks a lot Jason.... if value of d=0 then we should not bother about using differncing methods ?



Jason Brownlee August 31, 2017 at 10:30 am #

DEEPIKA ↗

It depends.

The d only does a 1-step difference. You n

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Amritanshu August 31, 2017 at 5:21 pm #

Jason, Can I get a link to understand it in a better way

REPLY ↗

Jason Brownlee September 1, 2017 at 6:43 am #

You can get started with time series here:

<https://machinelearningmastery.com/start-here/#timeseries>

Amritanshu September 5, 2017 at 11:22 pm #

REPLY ↗

Hi Jason

I am trying to predict values for the future. I am facing issue.

My data is till 31st July and I want to have prediction of 20 days.....

My Date format in excel file for the model is 4/22/17 –MM-DD-YY

output = model_fit.predict(start='2017-01-08',end='2017-20-08')

Error :

Traceback (most recent call last):

File "C:/untitled1/prediction_new.py", line 31, in

output = model_fit.predict(start='2017-01-08',end='2017-20-08')

File "C:\Python36\lib\site-packages\statsmodels\base

Start Machine Learning

```

obj = data.wrap_output(func(results, *args, **kwargs), how)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 1492, in predict
    return self.model.predict(self.params, start, end, exog, dynamic)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 733, in predict
    start = self._get_predict_start(start, dynamic)
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima_model.py", line 668, in _get_predict_start
    method)
File "C:\Python36\lib\site-packages\statsmodels\tsa\base\datetools.py", line 52, in _index_date
    date = dates.get_loc(date)
AttributeError: 'NoneType' object has no attribute 'get_loc'

```

Can you please help ?



Jason Brownlee September 7, 2017 at 12:45

Sorry, I'm not sure about the cause of this there?

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees.** Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Amritanshu September 20, 2017 at 10:20 pm #

Not working ... can you please help ?

Kashif September 6, 2017 at 8:11 pm #

REPLY ↗

Hi Sir

Please help me to resolve this error

```

from pandas import read_csv
from pandas import datetime
from matplotlib import pyplot

def parser(x):
    return datetime.strptime('190'+x, '%Y-%m')

series = read_csv('E:/data/csv/shampoo-sales.csv', header=0, parse_dates=[0], index_col=0,
squeeze=True, date_parser=parser)
print(series.head())
series.plot()
pyplot.show()

```

ERROR is

runfile('C:/Users/kashi/Desktop/prog/Date_time.py', w

Start Machine Learning

Traceback (most recent call last):

File "", line 1, in

```
runfile('C:/Users/kashi/Desktop/prog/Date_time.py', wdir='C:/Users/kashi/Desktop/prog')
```

File "C:\Users\kashi\Anaconda3\lib\site-packages\spyder\utils\site\sitecustomize.py", line 866, in runfile
execfile(filename, namespace)

File "C:\Users\kashi\Anaconda3\lib\site-packages\spyder\utils\site\sitecustomize.py", line 102, in execfile
exec(compile(f.read(), filename, 'exec'), namespace)

File "C:/Users/kashi/Desktop/prog/Date_time.py", line 10, in

```
series = read_csv('E:/data/csv/shampoo-sales.csv', header=0, parse_dates=[0], index_col=0,  
squeeze=True, date_parser=parser)
```

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 691, in
read
return _read(filepath_or_buffer, kwds)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 680, in
read
return parser.read()

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 670, in
read
ret = self._engine.read(nrows)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 659, in
read
index, names = self._make_index(data, alldata, names)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 649, in
read
index = self._agg_index(index)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1111, in _agg_index
arr = self._date_conv(arr)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 2288, in converter
return generic_parser(date_parser, *date_cols)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\date_converters.py", line 38, in generic_parser
results[i] = parse_func(*args)

File "C:/Users/kashi/Desktop/prog/Date_time.py", line 8, in parser
return datetime.strptime('190'+x, '%Y-%m')

File "C:\Users\kashi\Anaconda3\lib_strptime.py", line 510, in _strptime_datetime
tt, fraction = _strptime(data_string, format)

File "C:\Users\kashi\Anaconda3\lib_strptime.py", line 343, in _strptime
(data_string, format))

ValueError: time data '1901-Jan' does not match format '%Y-%m'

I have already removed the footer note from the dataset and I also open dataset in text editor. But I couldn't remove this error. But when I comment "date_parser=parser" my code runs but doesn't show years,
How to resolve it?

Thanks

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee September 7, 2017 at 12:53 pm #

REPLY ↗

Perhaps %m should be %b?

Alec September 21, 2017 at 6:41 pm #

REPLY ↗

Getting this problem:

```
File "/shampoo.py", line 6, in parser
    return datetime.strptime('190'+x, '%Y-%m')
TypeError: ufunc 'add' did not contain a loop with sign
dtype('<U32')
```

I've tried '%Y-%b' but that only gives me the "does no

Any ideas?

/ Thanks



Jason Brownlee September 22, 2017 at 5:35

Hi Alex, sorry to hear that.

Confirm that you downloaded the CSV version of the dataset and that you have deleted the footer information from the file.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Alec September 22, 2017 at 5:41 pm #

REPLY ↗

Hey,

I got it to work right after I wrote the post...

The header in the .csv was written as "Month,""Sales" and that caused the error, so I just changed it to "month", "sales" and it worked.

Thanks for putting in the effort to follow up on posts!



Jason Brownlee September 23, 2017 at 5:36 am #

REPLY ↗

Glad to hear that Alec!

Start Machine Learning

Teja October 6, 2017 at 8:15 am #

REPLY ↗

Hey,

I've two years monthly data of different products and their sales at different stores. How can I perform Time series forecasting on each product at each location?

Thanks in advance.



Jason Brownlee October 6, 2017 at 11:04 am #

REPLY ↗

You could explore modeling products separately and then combine the data. See what works best.

Shud October 23, 2017 at 7:47 pm #

Hey Jason,

You mentioned that since the residuals doesn't have mean zero. The standard deviation of the residuals is in the order of 10^5. So i would like to know your thoughts please?

Start Machine Learning

×

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Shud October 23, 2017 at 8:20 pm #

REPLY ↗

Btw my mean is ~400

zhifeng November 4, 2017 at 1:17 am #

REPLY ↗

For those who came with an error of ValueError: time data '1901-Jan' does not match format '%Y-%m'

please replace the month column with following:

Month

1-1

1-2

1-3

1-4

1-5

1-6

1-7

1-8

Start Machine Learning

1-9
1-10
1-11
1-12
2-1
2-2
2-3
2-4
2-5
2-6
2-7
2-8
2-9
2-10
2-11
2-12
3-1
3-2
3-3
3-4
3-5
3-6
3-7
3-8
3-9
3-10
3-11
3-12

Start Machine Learning X

You can master applied Machine Learning **without math or fancy degrees.** Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

cuongquyet November 10, 2017 at 9:59 pm #

REPLY ↗

Dear Jason,

Firstly, I would like to thanks about your sharing

Secondly, I have a small question about ARIMA with Python. I have about 700 variables need to be forecasted with ARIMA model. How Python supports this issuse Jason

For example, I have data of total orders in a country, and it will be contributte to each districts
So I need to forecast for each districts (about 700 districts)

Thanks you so much



Jason Brownlee November 11, 2017 at 9:22 am

REPLY ↗

Start Machine Learning

Generally, ARIMA only supports univariate time series, you may need to use another method.

That is a lot of variables, perhaps you could explore a multilayer perceptron model?

volity November 13, 2017 at 10:11 pm #

REPLY ↗

The result of `model_fit.forecast()` is like `(array([242.03176448]), array([91.37721802]), array([[62.93570815, 421.12782081]]))`. The first number is `yhat`, can you explain what the other numbers mean in the result? thank you!



Jason Brownlee November 14, 2017 at 10:13 am #

It may be the confidence interval:

<https://machinelearningmastery.com/time-series-forecasting-with-python/>

Chetan November 14, 2017 at 10:32 am #

Great blogpost Jason!

Had a follow up question on the same topic.

Is it possible to do the forecast with the ARIMA model at a higher frequency than the training dataset?

For instance, let's say the training dataset is sampled at 15min interval and after building the model, can I forecast at 1second level intervals?

If not directly as is, any ideas on what approaches can be taken? One approach I am entertaining is creating a Kernel Density Estimator and sampling it to create higher frequency samples on top of the forecasts.

Thanks, much appreciate your help!

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee November 15, 2017 at 9:44 am #

REPLY ↗

Hmm, it might not be the best tool. You might need something like a neural net so that you can design a one-to-many mapping function for data points over time.

Monsoon November 18, 2017 at 3:23 am #

REPLY ↗

Hi Jason,

Your tutorial was really helpful to understand the concept. But I have small doubt regarding the steps you followed at t

Start Machine Learning

```
X = series.values
size = int(len(X) * 0.66)
train, test = X[0:size], X[size:len(X)]
history = [x for x in train]
predictions = list()
for t in range(len(test)):
    model = ARIMA(history, order=(5,1,0))
    model_fit = model.fit(disp=0)
    output = model_fit.forecast()
    yhat = output[0]
    predictions.append(yhat)
    obs = test[t]
    history.append(obs)
print('predicted=%f, expected=%f' % (yhat, obs))
error = mean_squared_error(test, predictions)
```

Note: 1) here in the above for each iteration you're adding value because in real forecasting we don't have future trying to explain something and I'm not getting it.

2) Second doubt, aren't you suppose to perform "reverse differencing" in the model?

Kindly, please clear my doubt

Note: I have also went through one of your other tutorials on temperature in Australia.

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

here the steps you followed were convincing, also you have performed "inverse difference" step to scale the prediction to original scale.

I have followed the steps from the one above but I'm unable to forecast correctly.



Jason Brownlee November 18, 2017 at 10:23 am #

REPLY ↗

In this case, we are assuming the real observation is available after prediction. This is often the case, but perhaps over days, weeks, months, etc.

The differencing and reverse differencing were performed by the ARIMA model itself.

Somayeh November 28, 2017 at 12:39 am #

REPLY ↗

Hi Jason,

Recently I am working on time series prediction, but my research is a little bit complicated for me to understand how to fix a time series models to predict

[Start Machine Learning](#)

Recently I read your post in multi-step and multivariate time series prediction with LSTM. But my problem have a series input values for every time (for each second we have recorded more than 500 samples). We have 22 inputs and 3 targets. All the data has been collected during 600 seconds and then predict 3 targets for 600 next seconds. Please help me how can solve this problem?
It is noticed we have trend and seasonality pulses for targets during the time.



Jason Brownlee November 28, 2017 at 8:37 am #

REPLY ↩

Perhaps here would be a good place to start:

<https://machinelearningmastery.com/start-here/#timeseries>

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee December 8, 2017 at 5:36 am #

REPLY ↩

This post will help you make predictions that are out of sample:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Desmond December 9, 2017 at 7:52 pm #

REPLY ↩

Hey Jason thanks so much for the clarification! But just to clarify, when I run the example above, my inputs are the past records for the past 3 years grouped by month. Then, how the code actually plot out the forecasted graph is basically takes in those input and plot, am I right? So, can I assumed that the graph that plotted out is meant for the prediction of next year?



Jason Brownlee December 10, 2017 at 5:24 am #

REPLY ↩

I don't follow, sorry. You can plot anything you wish.

Desmond December 10, 2017 at

Start Machine Learning

Sorry but what does the expected and predicted means actually?



Jason Brownlee December 11, 2017 at 5:21 am #

The expected value is the real observation from your dataset. The predicted value is the value predicted by your model.

Desmond December 10, 2017 at 4:25 pm #

Also, why the prediction has to have 12 months? Looking forward to h



Jason Brownlee December 11, 2017 at 4:12 pm #

I arbitrarily chose to make predictions for 13 months.

You're right, it would have been clearer if I had said 12 months.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Desmond December 11, 2017 at 4:12 pm #

Hey Jason, thanks so much for the replies! But just to check with you, which line of the code should I modify so that it will only predict for the next 12 months instead of 13?

Also, just to be sure, if I were to predict for the profit for next year, the value that I should take should be the predicted rather than expected, am I right?

Thanks!!



Jason Brownlee December 11, 2017 at 4:55 pm #

Sorry, I cannot prepare a code example for you, the URLs I have provided show you exactly what to do.

Desmond December 11, 2017 at 6:24 pm #

Hey Jason, thanks so much for your reply.
The model above aims to make a prediction for the next 13 months.

Start Machine Learning

on what you do not have?

Also, may I check with you on how it works? Because I downloaded the sample dataset and the dataset contains the values of past 3 years grouped by months. So, can I assume the prediction takes all the values from past years into account in order to calculate for the prediction value? Or it simply takes the most recent one and calculate for the prediction?

Thanks!

Desmond December 11, 2017 at 4:17 pm #

REPLY ↗

Hey Jason, I am so sorry for the spams. But if I put some zero value for the profit, will it break the forecast? I am getting a non-zero value. Because sometimes I am getting “numerical error” message and I not sure if it is the zero values that cause the error.



Jason Brownlee December 11, 2017 at 4:56 pm #

Good question, it might depend on the model.

Perhaps spot check some values and see how the model handles them.

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

REPLY ↗

Desmond December 11, 2017 at 8:33 pm #

May I know what kind of situation will cause the error above? Is it because of drastic up and down from 3 different dataset?

Sushil Namdeo Raut December 13, 2017 at 10:26 am #

REPLY ↗

Hi Jason,

Thanks for this post. I am getting following error while running the very first code:

ValueError: time data ‘1901-Jan’ does not match format ‘%Y-%m’



Jason Brownlee December 13, 2017 at 4:13 pm #

REPLY ↗

Ensure your data is in CSV format and that the footer was removed.

Start Machine Learning

REPLY ↗

Denise December 13, 2017 at 7:07 pm #

Hi Jason, thanks so much for the share! The tutorial was good! However, when I am using my own data set, I am getting the same error message as one of the guy above. The error message is ‘numpy.linalg.linalg.LinAlgError: SVD did not converge’.

I tried to crack my head out trying to observe the data sets that caused the error message but I could not figure out anything. I tried with 0 value and very very very drastic drop or increase in the data, some seems okay but at some point, some data set will just fail and return the error message.

May I know what kind of data or condition will trigger the error above so I can take extra precaution when preparing the data?

**Jason Brownlee** December 14, 2017 at 5:36 am #

Perhaps try manually differencing the data

Perhaps there are a lot of 0 values in your data tha

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)
Denise December 14, 2017 at 2:38 pm #

I tried with multiple set of data without a single zero. I realized a problem but I not sure if my observation is correct as I am still trying to figure out how the code above works, for that part I might need your enlightenment.

Let's say the data is 1000, 100, 10000 respectively to first, second and third year. This kind of data will throw out the error message above. So can I assume that, as long as there is a big drastic drop/increase in the data set, in this case from 100 to 10000, this kind of condition will execute with error?

**Jason Brownlee** December 14, 2017 at 4:46 pm #

REPLY ↗

Sorry Denise, I'm not sure I follow.

anand February 27, 2018 at 4:40 am #

REPLY ↗

Hey Denise, i got the same issue. did you get any solution for this problem??

Kelly December 17, 2017 at 8:00 am #

Start Machine Learning

Hi Jason,

Thank you for the tutorial, it's great! I have a question about stationarity and differencing. If time series is non stationary but is made stationary with simple differencing, are you required to have $d=1$ in your selected model? Can I choose a Model with no differencing for this data if it gives me a better root mean square error and there is no evidence of autocorrelation?



Jason Brownlee December 17, 2017 at 8:57 am #

REPLY ↩

Yes, you can let the ARIMA difference or perform it yourself.

But ARIMA will do it automatically for you which makes it easier.

Satyajit Pattnaik December 20, 2017 at 10:53 pm

@Jason, This article has helped me a lot for the time series analysis earlier too, but could you help me with the future forecasting part. I am working on a project for 2017 and i want to predict the values for the next one year.

If we get help for this, that would be amazing 😊

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee December 21, 2017 at 5:26 am #

REPLY ↩

See this post on how to make predictions:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Satyajit Pattnaik December 21, 2017 at 2:23 am #

REPLY ↩

@Jason,

For future predictions, let say i have data till 10th November, and based on your analysis as shown above, can you help me with the future predictions for a week or so, need an idea of how to predict future data..



Jason Brownlee December 21, 2017 at 5:27 am #

REPLY ↩

Yes, see this post:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Start Machine Learning

Shariq Suhail December 27, 2017 at 4:29 pm #

REPLY ↗

Great post Jason!

I have a question:

– We need to ensure that the residuals of our model are uncorrelated and normally distributed with zero mean.

What if the residuals are not normally distributed?

It would be very grateful if you could explain how to approach in such scenario.

Thanks

Shariq



Jason Brownlee December 28, 2017 at 5:18 am #

It may mean that you could improve your model like a boxcox?

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Namrata Nayak December 28, 2017 at 5:12 pm #

@Jason, What if we don't want Rolling forecast, which means, my forecast should only be based on the training data, and it should predict the test data..

I am using the below code:

```
X = ts.values
size = int(len(X) * 0.75)
train, test = X[0:size], X[size:len(X)]
model = ARIMA(train, order=(4, 1, 2))
results_AR = model.fit(disp=0)
preds=results_AR.predict(size+1,size+16)
pyplot.plot(test[0:17])
pyplot.plot(preds, color='red')
pyplot.show()
```

This prediction is giving me really bad results, need urgent help on this.



Jason Brownlee December 29, 2017 at 5:19 am #

REPLY ↗

This is called a multi-step forecast and it is very challenging. You may need a different model.

More here:

<https://machinelearningmastery.com/multi-step-time-series-forecasting/>

Start Machine Learning

Vadim Pliner December 29, 2017 at 3:17 am #

REPLY ↗

Hi Jason, I have two questions.

1. Let's say I want to estimate an AR model like this: $x(t)=a*x(t-2) + e$. If I use ARIMA(2,0,0), it will add the term $x(t-1)$ as well, which I don't want. In SAS I would use $p=(2)$ on the estimate statement of proc arima rather than $p=2$.
2. How do I incorporate covariates? For example, a simple model like this: $x(t)=a*x(t-2) + b*f(t) + e$, where $f(t)$ e.g. is 1 if it's the month of January and 0 otherwise.

Thanks.



Jason Brownlee December 29, 2017 at 5:24 am #

Re the first question, it's good. I don't know what you mean by "google searchers are needed".

Re multivariates, you may need to use ARIMAX or similar models.

Fawad January 3, 2018 at 6:16 pm #

Hi,

I am getting the following error when loading the series dataframe in python
 "ValueError: time data '190Sales of shampoo over a three year period' does not match format '%Y-%m'"
 Ive just copy pasted the code from this website but its not working. Any suggestions? Im using Sypder



Jason Brownlee January 4, 2018 at 8:08 am #

REPLY ↗

Ensure you remove the footer from the data file.

Jelly January 9, 2018 at 1:58 pm #

REPLY ↗

Hi, may I know what are the yhat, obs and error variable are for? As for the error, is it better with greater value or the other way around? Thanks!



Jason Brownlee January 9, 2018 at 3:19 pm #

REPLY ↗

Start Machine Learning

yhat are the predictions. obs are the observations or the actual real data.

Jelly January 9, 2018 at 4:11 pm #

REPLY ↗

Thanks! Then what about the MSE? Is it the greater the better or the other way around?



Jason Brownlee January 10, 2018 at 5:20 am #

REPLY ↗

A smaller error is better.

Satyajit Pattnaik January 17, 2018 at 10:13 pm #

Could you please have a blog on Anomaly detection example itself.



Jason Brownlee January 18, 2018 at 10:09 am #

Thanks for the suggestion.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Omar Irbaihat January 23, 2018 at 1:51 am #

REPLY ↗

hey sir , thanks for that , Is ARIMA good for predictions of currencies exchange rate or not ?



Jason Brownlee January 23, 2018 at 8:05 am #

REPLY ↗

I don't know about currency exchange problems sorry. Try it and see.

Chintan January 25, 2018 at 7:20 am #

REPLY ↗

Hello,

Is it possible to predict hourly temperature for upcoming 5 years based on hourly temperature data of last 5 years ?

I am trying this out with ARIMA model, its giving me vrey bad output (attenuating curve).

Start Machine Learning



Jason Brownlee January 25, 2018 at 9:10 am #

REPLY ↗

You could model that, but I expect the skill to be very poor. The further in the future you want to predict, the worse the skill.

Jing February 2, 2018 at 9:24 am #

REPLY ↗

if the time series corresponds to brownian motion time series generated with different Hurst value (let's say H1 = 0.6 and H2 = 0.7), is this model a good fit to classify if it is H1 or H2?

Rajan R G February 12, 2018 at 1:25 am #

Hi Jason,

I have followed all of your posts related to Time Series parameter optimization also. The same code is working computed initial AR coefficients are not stationary You should induce stationarity, choose a different model pass your own start_params". The python version is so common?

Start Machine Learning

X

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee February 12, 2018 at 8:30 am #

REPLY ↗

Sorry, I don't know about "running code in kaggle".

Sofia May 22, 2019 at 7:59 pm #

REPLY ↗

I get the same error when I run the code in my local PC. Not for every p and q though, but for higher values.



Jason Brownlee May 23, 2019 at 6:00 am #

REPLY ↗

Perhaps try using a "d" term to make the data stationary.

Deepu Raj March 10, 2018 at 6:53 pm #

Start Machine Learning

Hello, may I know what is the purpose for these two lines?

```
size = int(len(X) * 0.66)
train, test = X[0:size], X[size:len(X)]
```

Thanks!

Deepu Raj March 10, 2018 at 7:05 pm #

REPLY ↩

Also, just to double confirm with you on my understanding, basically what the algorithm does is, take in all input in csv and fit into model, perform a forecast, append the forecast value into the model, then go thru the for loop again to recreate a forecast value, then go thru the for loop again?

In addition, the next row prediction is always depe



Jason Brownlee March 11, 2018 at 6:23 am #

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

REPLY ↩



Jason Brownlee March 11, 2018 at 6:21 am #

To split the dataset into train and test sets.

Deepu Raj March 11, 2018 at 7:38 pm #

REPLY ↩

Is there a specific reason for you to multiply with 0.66? Thanks!



Jason Brownlee March 12, 2018 at 6:28 am #

REPLY ↩

No reason, just an arbitrarily chosen 66%/37% split of the data.

James Neligan March 13, 2018 at 6:56 am #

REPLY ↩

I need to forecasting the next x hour. How can i do this?

Start Machine Learning



Jason Brownlee March 13, 2018 at 3:03 pm #

REPLY ↗

This post might help:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Ajay March 15, 2018 at 2:28 am #

REPLY ↗

Thanks Jason for making it simple. I run the program but getting error

1st error :

TypeError: Cannot cast ufunc subtract output from dtype('float32') to dtype('float64') according to the casting rule 'same_kind'

After changing code , i got 2nd error

model = ARIMA(series.astype(float), order=(5,1,0))

I m getting following error

LinAlgError: SVD did not converge

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee March 15, 2018 at 6:32 am #

Looks like the data might have some issues. Perhaps calculate some summary stats, visualizations and look at the raw data to see if there is anything obvious.

Ajay Verma March 16, 2018 at 2:16 am #

REPLY ↗

Thanks Jason for the quick response. Now i tried for Sampoo dataset, getting following error :

ValueError: time data '1901-Jan' does not match format '%d-%m'

Code :

```
def parser(x): return datetime.strptime('190'+x, '%d-%m')
```

```
series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True,
date_parser=parser)
print(series.head())
series.plot()
pyplot.show()
```



Jason Brownlee March 16, 2018 at 6:21 am #

REPLY ↗

Start Machine Learning

Perhaps your data contains the footer. Here is a clean version of the data ready to go:
<https://raw.githubusercontent.com/jbrownlee/Datasets/master/shampoo.csv>

Satyajit Pattnaik March 19, 2018 at 7:28 pm #

REPLY ↗

When we use a recursive model for ARIMA, let say like saw in one of your examples:

```

1 for t in range(len(test)):
2     try:
3         model = ARIMA(history, order=(4,0,2))
4         model_fit = model.fit(disp=0)
5         output = model_fit.forecast()
6         yhat = output[0]
7         predictions.append(yhat)
8         obs = test[t]
9         history.append(obs)
10    except (ValueError, LinAlgError):
11        pass
12    print('predicted=%f, expected=%f' % (yhat, obs))

```

Why my final test vs predicted graph is coming as if, the test is following a pattern, predictions is following similar isn't working properly, i hope you got my point.

For example: if test[0] keeps increasing till test[5] and predictions[0] and decreases..

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee March 20, 2018 at 6:14 am #

REPLY ↗

It suggests the model is not skilful and is acting like a persistence model.

It may also be possible that persistence is the best that can be achieved on your problem.

Satyajit Pattnaik March 21, 2018 at 5:46 pm #

REPLY ↗

Does that mean, ARIMA isn't giving good results for my problem?

What are different ways of solving this problem by ARIMA, can differencing or Log approach be a good solution?



Jason Brownlee March 22, 2018 at 6:19 am #

REPLY ↗

You can use ACF/PACF plots to help choose ARIMA parameters, or you can grid search ARIMA parameters on your test set.

Start Machine Learning

REPLY ↗

Mihir Ranade March 21, 2018 at 12:28 am #

Hello! Thank you for this great tutorial. It'd be a great help if you guide me through one of my problems.

I want to implement a machine learning model to predict(forecast) points scored by each player in the upcoming game week.

Say I have values for a player (Lukaku) for 28 game weeks and I train my model based on some selected features for those 28 weeks. How do I predict the outcome of the 29th week?

I am trying to predict total points to be scored by every player for the coming game week.

So basically what should be the input to my model for 29th game week? Since the game assigns points as per live football games happening during the week, I w

Thank you 😊

**Jason Brownlee** March 21, 2018 at 6:37 am #

I would recommend looking into rating system
https://en.wikipedia.org/wiki/Elo_rating_system

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

REPLY ↗

Raphael March 30, 2018 at 2:20 am #

Hi Jason,
Great tutorial once again!

I have a question on your Rolling Forecast ARIMA model.

When you are appending obs (test(t)) on each step to history, aren't we getting data leakage?
The test set is supposed to be unseen data, right? Or are you using the test set as a validation set?

**Jason Brownlee** March 30, 2018 at 6:42 am #

REPLY ↗

In this case no, we are assuming the real observation is available at the end of each iteration.

You can change the assumptions and therefore the test setup if you like.

Raphael April 2, 2018 at 6:09 am #

REPLY ↗

oh I see, i misunderstood this assumption, sorry. But how can I predict multiple steps? I used the predict() method from ARIMA model



Jason Brownlee April 2, 2018 at 2:44 pm #

REPLY ↗

Yes, you can use the predict() function. Performance may be poor as predicting multiple steps into the future is very challenging.

Ftima April 2, 2018 at 6:57 pm #

REPLY ↗

Hi,

In case we try to introduce more than one input, then how?

Thanks



Jason Brownlee April 3, 2018 at 6:32 am #

We don't fit one point, we fit a series of points.

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Hsiang April 9, 2018 at 9:35 am #

Hi Jason,

Very nice introduction! Thank you very much for always bringing us excellent ML knowledge.

Can you further explain why you chose $(p,d,q) = (5,1,0)$? Or you did grid search (which you show in other blogs) using training/test sets to find minimum msg appears at $(5,1,0)$? Did you know any good reference for diagnostic plots for the hyper-parameters grid searching?

Meanwhile, I am interested in both time-series book and LSTM book. If I purchased both, any further deal?



Jason Brownlee April 10, 2018 at 6:09 am #

REPLY ↗

I recommend using both a PACF/ACF interpretation and grid searching approaches. I have tutorials on both.

Sorry, I cannot create custom bundles of books, you can see the full catalog here:
<https://machinelearningmastery.com/products>

Hsiang April 12, 2018 at 6:05 pm #

REPLY ↗

Start Machine Learning

Hi Jason,

Thank you for your answer. I have purchased time series book.

I still have few more questions on ARIMA model:

(1) The shampoo sale data obviously shows non-stationary; strictly speaking, we should transform data until it becomes stationary data by taking logarithm and differencing (Box-Cox transformation), and then apply to ARIMA model. Is it correct?

(2) Does the time series data with first-order differencing on ARIMA ($p, 0, q$) give the similar results to the time series data without differencing on ARIMA($p, 1, q$)? i.e. $d = 1$ in ARIMA(p, d, q) equivalently process data with first-order difference?

(3) In this example, we chose ARIMA (5,1,0) and $p=5$ came from the autocorrelation plot. However, what I read from the book <https://www.otexts.org/fpp/2/10> is that we can check ACF and PACF for correlation. So, I am confused about how to choose ARIMA model. I am not sure if I should use ACF or PACF for choosing ARIMA model. Are there any things I should pay attention to?



Jason Brownlee April 13, 2018 at 6:06 am #

The shampoo data is non stationary, so we need to difference the data before modeling or as part of the ARIMA.

No, 0 and 1 for d mean no differencing and one differencing respectively.

Yes, you can check ACF and PACF for correlation.

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Marco April 11, 2018 at 6:06 am #

REPLY ↗

Hi Jason,

In your code you use :

`yhat=output[0]`

So you take the first element of output, what are the other elements of output represent?

Thank you



Jason Brownlee April 11, 2018 at 6:42 am #

REPLY ↗

You can see all of the returned elements here:

http://www.statsmodels.org/dev/generated/statsmodels.tsa.arima_model.ARMAResults.forecast.html

Mutasem April 22, 2018 at 12:48 pm #

Start Machine Learning

Thank you for your efforts ... i have question

i'm using the following code as mentioned above

```
def parser(x):
```

```
    return datetime.strptime('190' + x, '%Y-%m')
```

but the error appears :

ValueError: time data '1902-Jan' does not match format '%Y-%m'

could you please help me



Jason Brownlee April 23, 2018 at 6:13 am #

REPLY ↗

It looks like you downloaded the dataset i

You can get the correct dataset here:

<https://raw.githubusercontent.com/jbrownlee/Datasets/master/365day.csv>

Harshil April 24, 2018 at 8:26 pm #

Hey Jason,

Best article I have ever seen. Currently I am working on

ARIMA model. I have data of appliance energy which depends on 26 variables over period of 4 months. My question is how can I use 26 variables to forecast the future value?

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee April 25, 2018 at 6:19 am #

REPLY ↗

Thanks.

Sorry, I don't have an example of ARIMA with multiple input variables.

Harshil April 26, 2018 at 5:33 pm #

REPLY ↗

Hello Jason,

Thanks for your reply.

Can I solve my problem with ARIMA model?



Jason Brownlee April 27, 2018 at 6:02 am #

REPLY ↗

Perhaps a variant that supports multiple s

Start Machine Learning

Muhammad May 8, 2018 at 10:42 am #

REPLY ↗

Hey Jason, I am new to data analytics. From the chart, may I know how you determined it is stationary or non-stationary as well as how do you see whether it has a lagged value?

Thanks!



Jason Brownlee May 8, 2018 at 2:53 pm #

REPLY ↗

Yes, you can learn more about ACF and PACF here:
<https://machinelearningmastery.com/gentle-introduction-random-walk-times-series-forecasting-python/>

Sven May 20, 2018 at 8:40 am #

Hello Jason,

can Autoregression model be used for forecasting stock price?

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee May 21, 2018 at 6:23 am #

REPLY ↗

Yes, but it will likely do worse than a persistence model.

Learn more here:

<https://machinelearningmastery.com/gentle-introduction-random-walk-times-series-forecasting-python/>

Randal Michnovicz May 30, 2018 at 7:21 am #

REPLY ↗

Hello! I think you may have made a mistake in the following paragraph.

"If we used 100 observations in the training dataset to fit the model, then the index of the next time step for making a prediction would be specified to the prediction function as start=101, end=101. This would return an array with one element containing the prediction."

Since python is zero-indexed, the index of the next time step for making a prediction should be 100, I think.



Jason Brownlee May 30, 2018 at 3:06 pm #

REPLY ↗

Start Machine Learning

Not in this case. Try it and see.

Franky Philip June 7, 2018 at 2:19 am #

REPLY ↗

Hello Jason!

I'm stuck at this error when i execute these lines of code:

```
from pandas import read_csv
from pandas import datetime
from matplotlib import pyplot

def parser(x):
    return datetime.strptime('190'+x, '%Y-%m')

series = read_csv('shampoo_time_series.csv', header=0,
date_parser=parser)
print(series.head())
series.plot()
pyplot.show().
```

Error:-

time data '19001-Jan' does not match format '%Y-%m'

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee June 7, 2018 at 6:33 am #

REPLY ↗

Perhaps you downloaded a different version of the dataset. Here is a direct link:

<https://raw.githubusercontent.com/jbrownlee/Datasets/master/shampoo.csv>

Does that help?

bakhouche June 13, 2018 at 6:40 pm #

REPLY ↗

hi dear,

can ask you please what is the meaning of the arrow that cant be copied, thank you.



Jason Brownlee June 14, 2018 at 5:59 am #

REPLY ↗

Sorry, what arrow?

Arsim June 17, 2018 at 8:24 am #

Start Machine Learning

Hi Jason,

great tutorial, as always! Thank you very much for providing your excellent knowledge to the vast community! You really helped me to get a better understanding of this ARIMA type of models.

Do you plan to make a tutorial on nonlinear time-series models such as SETAR? Would be great, because I could not really find anything in this region.



Jason Brownlee June 18, 2018 at 6:36 am #

REPLY ↩

Thanks for the suggestion.

I do hope to cover more methods for nonlinear tim

Saloni Patil June 21, 2018 at 6:59 pm #

Hi Jason

I tried the code with my data. ACF, PACF plots aren't showing anything by which I can still try the forecast? What shou



Jason Brownlee June 22, 2018 at 6:04 am #

REPLY ↩

Perhaps try a grid search on ARIMA parameters and see what comes up?

ezgi June 22, 2018 at 10:29 pm #

REPLY ↩

Hi Jason,

Is it possible to make a forecast with xgboost for a time series data with categorical variables?



Jason Brownlee June 23, 2018 at 6:18 am #

REPLY ↩

Yes.

dnyanada June 26, 2018 at 3:48 am #

REPLY ↩

Hello Jason, I have been following your articles and it has been very helpful.
I am running the same code above and get following error:

Start Machine Learning

```
ValueError Traceback (most recent call last)
```

```
in ()
```

```
7 pred=list()
```

```
8 for i in range(len(test)):
```

```
--> 9 model=ARIMA(history,order=(5,1,0))
```

```
10 model_fit=model.fit(disp=0)
```

```
11 output=model_fit.forecast()
```

```
~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in __new__(cls,
```

```
endog, order, exog, dates, freq, missing)
```

```
998 else:
```

```
999 mod = super(ARIMA, cls).__new__(cls)
```

```
-> 1000 mod.__init__(endog, order, exog, dates, freq, missing)
```

```
1001 return mod
```

```
1002
```

```
~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in __new__(cls,
```

```
endog, order, exog, dates, freq, missing)
```

```
1013 # in the predict method
```

```
1014 raise ValueError("d > 2 is not supported")
```

```
-> 1015 super(ARIMA, self).__init__(endog, (p, q), exog)
```

```
1016 self.k_diff = d
```

```
1017 self._first_unintegrate = unintegrate_levels(self.endog, self.exog, self.freq, self.order, self.k_diff)
```

```
~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in __init__(self,
```

```
endog, order, exog, dates, freq, missing)
```

```
452 super(ARMA, self).__init__(endog, exog, dates, freq, missing=missing)
```

```
453 exog = self.data.exog # get it after it's gone through processing
```

```
-> 454 _check_estimable(len(self.endog), sum(order))
```

```
455 self.k_ar = k_ar = order[0]
```

```
456 self.k_ma = k_ma = order[1]
```

```
~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in _check_estimable(self, nobs, n_params)
```

```
438 def _check_estimable(nobs, n_params):
```

```
439 if nobs < n_params: raise ValueError("Insufficient degrees of freedom to estimate")
```

```
441
```

```
442
```

ValueError: Insufficient degrees of freedom to estimate

the code used

```
from sklearn.metrics import mean_squared_error
```

```
size = int(len(df) * 0.66)
```

```
train,test=df[0:size],df[size:len(df)]
```

```
print(train.shape)
```

```
print(test.shape)
```

```
history=[x for x in train]
```

```
pred=list()
```

Start Machine Learning

You can master applied Machine Learning

without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Start Machine Learning

```

for i in range(len(test)):
    model=ARIMA(history,order=(5,1,0))
    model_fit=model.fit(disp=0)
    output=model_fit.forecast()
    yhat=output[0]
    pred.append(yhat)
    obs=test[i]
    history.append(obs)
    print('predicted = %f,expected = %f',(yhat,obs))
    error=mean_squared_error(test,pred)
    print('Test MSE: %.3f' % error)

plt.plot(test)
plt.plot(pred,color='red')
plt.show()

```

Only change I have made in code is date index. I have

`dt=pd.date_range("2015-01-01", "2017-12-1", freq="M")`

Can you explain what is wrong?

also,

I was under impression that you use `auto_corr` function in your code when you call ARIMA why have you used (5) to be (0,1,5)?

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee June 26, 2018 at 6:45 am #

REPLY ↗

I have more on the ACF/PACF plots and how to interpret them here:

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Dnyanada Arjunwadkar June 26, 2018 at 9:28 am #

REPLY ↗

Hello Jason, I posted a problem earlier today that I have successfully resolved. thanks for your help.



Jason Brownlee June 26, 2018 at 2:26 pm #

REPLY ↗

Glad to hear it.

Dnyanada Arjunwadkar June 27, 2018 at 4:13 am #

Start Machine Learning

Hello Jason,

Thanks for the helpful article.

My question is :

"A rolling forecast is required given the dependence on observations in prior time steps for differencing and the AR model."

can you please elaborate?

How do we decide when to use Rolling forecast and when not to use rolling forecast?

what are the factors do you consider?

Thanks



Jason Brownlee June 27, 2018 at 8:22 am #

I believe I mean a walk-forward validation

<https://machinelearningmastery.com/backtest-machine-learning-predictions/>

mithril July 5, 2018 at 1:30 am #

Hello,

My company is supermarket , which have 30 stores and over 2000 products. My boss want me to predict each product sale number in next 7 days.

I think below features would affect sales count much

1. a day is festival
2. a day is weekend
3. a day's weather
4. a day is coupon day

But I don't know how to embed above features with ARIMA model.

And also our data is from 2017-12 to now, there is no history season data.

Could you please give me a some advice?

Thank you.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 5, 2018 at 7:57 am #

REPLY ↩

They could be exogenous binary variables that the statsmodels ARIMA does support.

Paola July 22, 2018 at 8:13 am #

Start Machine Learning

Great article! But I have a question. I have a daily time series, and I am following the steps from the time series forecasting book. How do I obtain the acf and pacf visually (for the Manually Configured ARIMA)? because I will have more than 1000 lag values (as my dataset is for many years), and after this I will need to search for the hyperparameters. I will really appreciate the help



Jason Brownlee July 23, 2018 at 6:00 am #

REPLY ↩

An ARIMA might not be appropriate for 1000 lags.

Luisa July 22, 2018 at 8:15 am #

Great



Jason Brownlee July 23, 2018 at 6:00 am #

Thanks.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

X

Ivan July 22, 2018 at 11:56 pm #

REPLY ↩

thank you very much, Jason.

However. I have some problem. Whenever I adopt your code for forecasting when no validation data is available,

```
for t in range(93):
    model = ARIMA(history, order=(5,1,0))
    model_fit = model.fit(disp=0)
    output = model_fit.forecast()
    yhat = output[0]
    predictions.append(yhat)
    history.append(yhat)
    print('predicted=%f' % (yhat))
```

my series converge to a constant number after a certain number of iterations, which is not right. What is the mistake?



Jason Brownlee July 23, 2018 at 6:12 am #

REPLY ↩

You can fit a final model and make a prediction by calling forecast().

Start Machine Learning

Here's an example:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Siddharth August 3, 2018 at 3:52 pm #

REPLY ↗

Hi Jason,

Your articles are great to read as they give just the right amount of background and detail and are practical oriented. Please continue writing.

I have a question though, being not from the statistical background, i am having difficulty in interpreting the output that is displayed after the summary of the fit model. This summarizes the coefficient values used as well as observations.

Can you please provide some explanation on their attributes interpretation of the results

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 4, 2018 at 5:59 am #

Thanks.

Perhaps focus on the skill of the model and using the forecast of the model?

Anna August 5, 2018 at 8:57 am #

REPLY ↗

Hi Jason,

Thanks a lot for this awesome tutorial.

I am training on a dataset where I have to predict Traffic and Revenue during a campaign (weeks 53,54,55) driven by this marketing campaigns. I think I can only use data preceding the campaigns (weeks 1 to 52) even though I have the numbers for campaign and post campaign.

I have a file as follows:

```
week// campaign-period // TV-traffic // Revenue Traffic
1 //pre-campaign // 108567 // 184196,63
2 //pre-campaign // 99358 // 166628,38
...
53 // Campaign // 135058 //240163,25
54 // Campaign // 129275 //238369,88
...
56 // post-campaign //94062 // 141284,88
...
62 // post-campaign // 86695 // 130153,38
```

Start Machine Learning

It seems like a statistical problem and I don't know whether ARIMA is suitable for this use case (very few data, only 52 values to predict the following one). Do you think I can give it a shot with ARIMA or do you think there are other models that could be more suitable for such a use case please?

Thanks a lot for your help.



Jason Brownlee August 6, 2018 at 6:23 am #

REPLY ↗

Perhaps list out 10 or more different framings of the problem, then try fitting models to a few to see what works best?

Anna August 12, 2018 at 4:31 am #

Hi Jason,
Thanks a lot for this awesome tutorial.

I am training on a dataset where I have to predict (weeks 53,54,55) driven by this marketing campaigns. (weeks 1 to 52) even though I have the numbers

I have a file as follows:

```
week// campaign-period // TV-traffic // Revenue Traffic
1 //pre-campaign // 108567 // 184196,63
2 //pre-campaign // 99358 // 166628,38
...
53 // Campaign // 135058 //240163,25
54 // Campaign // 129275 //238369,88
...
56 // post-campaign //94062 // 141284,88
...
62 // post-campaign // 86695 // 130153,38
```

It seems like a statistical problem and I don't know whether ARIMA is suitable for this use case (very few data, only 52 values to predict the following one). Do you think I can give it a shot with ARIMA or do you think there are other models that could be more suitable for such a use case please?

Thanks a lot for your help.

Thank you for your help

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 12, 2018 at 6:36 am #

REPLY ↗

Perhaps try it and see how you go

Start Machine Learning

Nii Anyetei August 7, 2018 at 5:46 am #

REPLY ↗

Hi Jason, the constant updates are great and very helpful. I need a bit of help with my work. Im trying to forecast solid waste generation in using ANN. But I'm finding challenges with data and modeling my problem. If you could at least get me a headway that can help me produce something in 2weeks I will be grateful. I want to consider variables such as already generated solid waste, population, income levels, educational levels, etc. I hope to hear from you soon.



Jason Brownlee August 7, 2018 at 6:35 am #

This is a good place to start for deep learning:
<https://machinelearningmastery.com/start-here/#deep>

Wen Ge August 8, 2018 at 7:32 pm #

Many thanks Jason, it's really helpful!

Just one question, my data set contains some sales values with zero values. Is there a way to deal with this in a neural network model? if there will be issues, anyway I can deal with the zero values in my data set? Thanks in advance for your advice!

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 9, 2018 at 7:36 am #

REPLY ↗

It can deal with zero values.

Brian Stephans August 15, 2018 at 1:55 am #

REPLY ↗

Hello Jason,

Any idea why I am having issues with datetime?

This is the error that I have received

Traceback (most recent call last):

File "/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site-packages/pandas/io/parsers.py", line 3021, in converter
 date_parser(*date_cols), errors='ignore')

File "/Users/Brian/PycharmProjects/MachineLearningMasteryTimeSeries1/ARIMA nv" line 9, in parser

Start Machine Learning

```
return datetime.strptime('190' + x, '%Y-%m')
TypeError: strptime() argument 1 must be str, not numpy.ndarray
```

During handling of the above exception, another exception occurred:

Thank You
Brian



Jason Brownlee August 15, 2018 at 6:07 am #

REPLY ↗

Perhaps your data file, try this one instead:

<https://raw.githubusercontent.com/jbrownlee/Data>

Anton Petrov August 17, 2018 at 1:32 am #

The formating of csv seems different for everyone. I think it's better to copy/paste the data from Jason (just copy pasted this into a shampoo-sales.csv file).

– thanks to the person above for the tip

1-1,266
1-2,145.9
1-3,183.1
1-4,119.3
1-5,180.3
1-6,168.5
1-7,231.8
1-8,224.5
1-9,192.8
1-10,122.9
1-11,336.5
1-12,185.9
2-1,194.3
2-2,149.5
2-3,210.1
2-4,273.3
2-5,191.4
2-6,287
2-7,226
2-8,303.6
2-9,289.9
2-10,421.6
2-11,264.5
2-12,342.3
3-1,339.7

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Start Machine Learning

3-2,440.4
 3-3,315.9
 3-4,439.3
 3-5,401.3
 3-6,437.4
 3-7,575.5
 3-8,407.6
 3-9,682
 3-10,475.3
 3-11,581.3
 3-12,646.9



Jason Brownlee August 17, 2018 at 6:31 am

It is also available on my github:
<https://github.com/jbrownlee/Datasets>

SA August 17, 2018 at 8:08 am #

Hello Jason

I'm trying to divide time series dataset into several dataset and select the best one as preprocessing dataset. I would like to use RMSE to evaluate each subset. In other word to select the window size and frame size before I do the training . Please let me know if you have any article on rows selection not column selection

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 17, 2018 at 2:04 pm #

REPLY ↗

Yes, this post will help tune the parameters of ARIMA that will include tuning the size of the window for each aspect of the ARIMA model:

<https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/>

SA August 18, 2018 at 8:07 am #

REPLY ↗

Hello Jason

Many thanks for your reply. I have tried the code on the following data set and got "Best ARIMANone MSE=inf"

date price
 0 20160227 427.1

Start Machine Learning

1 20161118 750.9
 2 20160613 690.9
 3 20160808 588.7
 4 20170206 1047.3

RangelIndex: 657 entries, 0 to 656

Data columns (total 2 columns):

date 657 non-null int64
 price 657 non-null float64
 dtypes: float64(1), int64(1)
 memory usage: 10.3 KB

SA August 18, 2018 at 8:16 am #

Hello Jason

Just to clarify my previous question that i have 700 rows
 70(window size) rows for prediction and decide on the



Jason Brownlee August 19, 2018 at 6:14 am

Sounds great, let me know how you go!

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



SA August 19, 2018 at 7:11 am #

REPLY ↗

Hi Jason

Please let me know if you have an article help on specifying frame size , frame step and extent of prediction as data pre-processing step using RMSE and SEP.



Jason Brownlee August 20, 2018 at 6:30 am #

REPLY ↗

I do, the grid search of the ARIMA algorithm I linked to above does that.

Perhaps try working through it first?

SA August 21, 2018 at 6:41 am #

Thanks Jason. Your post in Gr

Search and got best Arima model .

Start Machine Learning

Now I want to use the result and train the window in LSTM

ARIMA(1, 0, 0) MSE=39.723
 ARIMA(1, 0, 1) MSE=39.735
 ARIMA(1, 1, 0) MSE=36.148
 ARIMA(3, 0, 0) MSE=39.749
 ARIMA(3, 1, 0) MSE=36.141
 ARIMA(3, 1, 1) MSE=36.131
 ARIMA(6, 0, 0) MSE=39.806
 ARIMA(6, 1, 0) MSE=36.134
 ARIMA(6, 1, 1) MSE=36.128
 Best ARIMA(6, 1, 1) MSE=36.128



Jason Brownlee August 21, 2018

An LSTM is a very different algorithm. Does it need at least 6 time steps as input?

SA August 22, 2018 at 7:29 am #

I have 5 years of time series data. I'm not sure what size I want to get the best optimal window as input to LSTM!

Appreciate your feedback.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 22, 2018 at 1:51 pm #

Test many different sized subsequence lengths and see what works best.

SA August 23, 2018 at 7:13 am #

Can I use Gridsearch for the testing purpose to specify the window size for LSTM? And if yes what would be the parameters equal to 60/90/120 days?



Jason Brownlee August 23, 2018 at 8:04 am #

I would recommend running the grid search yourself with a for-loop.

Try time periods that might make sense.

Start Machine Learning

SA August 24, 2018 at 8:06 am #

So I did the for-loop and manage to get different windows.

Now to calculate the RMSE do I need to do linear regression prediction for each window in order to calculate the RMSE or is there any other way around?



Jason Brownlee August 24, 2018 at 9:16 am #

I would expect that you would calculate the RMSE of the models. The models have different approaches even.

SA August 25, 2018 at 7:38 am #

I got the following as examples

For 360 days

Window start after 0 days with window size 360 have RMSE

734.1743876097737

Window start after 100 days with window size 360 and step 100 have RMSE

369.94549420288877

Window start after 200 days with window size 360 and step 100 have RMSE

105.70778076287142

For 180 days

Window start after 0 days with window size 180 and step 90 have RMSE

653.9070358902835

Window start after 90 days with window size 180 and step 90 have RMSE

326.7832188924093

Window start after 180 days with window size 180 and step 90 have RMSE

135.01118940666115

Window start after 270 days with window size 180 and step 90 have RMSE

38.422587695965746

Window start after 360 days with window size 180 and step 90 have RMSE

60.73374764651785

Window start after 450 days with window size 180 and step 90 have RMSE

52.386817309349176

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 26, 2018 at 6:19 am #

Well done!

Start Machine Learning

SA August 26, 2018 at 7:04 am #

Thanks Jason

Appreciate your support.

Your posts are really great and well organized.

I'm excited to ready your publications 😊



Jason Brownlee August 27, 2018

Thanks for your support!

Waldo August 18, 2018 at 9:36 pm #

Hi Jason! Here client and time series forecast
When forecasting, I very often get this error:

LinAlgError: SVD did not converge

Any ideas how to solve this in general?

Thanks!

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 19, 2018 at 6:20 am #

REPLY ↩

This is common.

Sounds like the linear algebra library used to solve the linear regression equation for a given configuration failed.

Try other configurations?

Try fitting a linear regression model manually to the lag obs?

Try normalizing the data beforehand?

Let me know how you go.

Renato August 23, 2018 at 9:51 am #

REPLY ↩

Hey Jason, what model i can use to equipment fault detection and prediction? So have some variables that correlate with others and i need to identification which are. See you soon

Start Machine Learning



Jason Brownlee August 23, 2018 at 1:54 pm #

REPLY ↗

Try a suite of methods in order to discover what works best for your specific problem.

Romain September 2, 2018 at 7:20 pm #

REPLY ↗

Hello Jason,

There is something that I struggle to understand, it would awesome if you could give me a hand.

In ARIMA models, the optimization fits the MA and AR parameters of linear combination of previous terms for formula could be :

$X_t - a_1 X_{t-1} \dots - a_p X_{t-p} \dots = e_t + b_1 e_{t-1} + \dots$

When the fit method is used, it takes the train values of

When the forecast method is used, it forecast the next train values

When the predict method is used, it forecast the next

Let's say I fit a model on n steps in the train set. Now I

Now I am days n+1 and I have the exact signal value.

In the rolling forecast part of your code, you fit again the model with the expanded train set (up to n+1). But in that case the model parameters are changed. It's not the same model anymore.

Is it possible to train one model and then actualize the signal values (the x and e) without changing the parameters (a and b)?

It seems to me that it is important to keep one unique model and evaluate it against different time steps instead of training n different models for each new time steps we get.

I hope I was clear enough. I miss probably a key to understand the problem.

Thanks

Romain



Jason Brownlee September 3, 2018 at 6:15 am #

REPLY ↗

The model will use the prediction as the input to predict t+2.

Matthew Orehek September 7, 2018 at 7:28 am #

REPLY ↗

Start Machine Learning

Hi Jason – Very helpful post here, thanks for sharing. I'm curious why parameter 'p' should be equal to the number of significant lags from the auto correlation plot? Just was wondering if you could give any more context to this part of the problem. Thanks.



Jason Brownlee September 7, 2018 at 8:11 am #

REPLY ↩

Generally, we want to know how many lag observations have a measurable relationship with the next step so that the model can work on using them effectively.

Christopher September 12, 2018 at 12:44 pm #

I used your code to forecast daily temperature behind, i.e. learning history cannot accurately forecast with AIC.



Jason Brownlee September 12, 2018 at 2:39

X

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Perhaps try alternate configurations?
Perhaps try alternate algorithms?
Perhaps try additional transforms to the data?

This might help:

<https://machinelearningmastery.com/how-to-develop-a-skilful-time-series-forecasting-model/>

Anuradha Chaurasia September 16, 2018 at 11:23 pm #

REPLY ↩

How to use ARIMA model in SPSS with few samples as 6 years data and according to this data for how many years we can forecast the future.



Jason Brownlee September 17, 2018 at 6:31 am #

REPLY ↩

Sorry, I don't have examples of SPSS.

Qianqian September 18, 2018 at 1:25 am #

REPLY ↩

Hi Jason,

Start Machine Learning

Thanks for sharing! Very helpful post.

Recently I am writing the methodology of ARIMA, but I can not find any reference (for example, some ARIMA formulas contain constant but some don't have). So could you please give me some reference (or ARIMA formula information) of "statsmodels.tsa.arima_model import ARIMA" used in Python?

Thank you in advance.



Jason Brownlee September 18, 2018 at 6:20 am #

REPLY ↗

The best textbook on ARIMA is:

<https://amzn.to/2MD9IKw>

Pranab May 26, 2020 at 12:50 pm #

Hi Qianqian,

Prof. Hyndman's textbook: <https://otexts.com/fpp/>
Hope this helps.

Milind Mahajani September 20, 2018 at 12:44 am #

If one has a time series where the time steps are not uniform, what should be done while fitting a model such as ARIMA? I have price data for a commodity for about 4 years. The prices are available only for days that a purchase was made. This is often, but not always, every day. So sometimes purchases are made after 2, 3 or even more days and the prices are therefore available only for those days I need to forecast the price for the next week.

Thanks for any advice on this.



Jason Brownlee September 20, 2018 at 8:01 am #

REPLY ↗

Perhaps try modeling anyway?

Perhaps try an alternative model?

Perhaps try imputing the missing values?

Milind Mahajani September 20, 2018 at 8:27 pm #

REPLY ↗

Thank you, Dr Jason!

Start Machine Learning

Kruthika Vishwanath September 25, 2018 at 6:55 am #

REPLY ↗

Hi Jason,

Thanks for this post.

I am working on finding an anomaly using arima. Will I be able to find from the difference in actual & predicted value shown above ?

Thanks,

Kruthika



Jason Brownlee September 25, 2018 at 2:43

Sorry, I don't have examples of using ARI

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Bhadri September 29, 2018 at 6:20 pm #

Hi Jason,

I have couple of questions.

1. is it necessary that we need to have always uni variate data set to predict for time series using ARIMA?
What if i have couple of features that i want to pass along with the date time?

2. is it also necessary that we have a non-stationary data to use time series for modelling? what if the data is already stationary? can i still do the modelling using time series?

Thanks

Bhadri



Jason Brownlee September 30, 2018 at 6:02 am #

REPLY ↗

ARIMA can support exogenous variables, this is called ARIMAX.

If the data is already stationary, you can begin modeling without transforms.

Bhadri September 30, 2018 at 3:13 pm #

REPLY ↗

Thanks Jason!! do u have any examples related to ARIMAX or point me to some articles..

Start Machine Learning



Jason Brownlee October 1, 2018 at 6:24 am #

REPLY ↗

Yes, there are examples here:

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

awa October 17, 2018 at 4:36 pm #

REPLY ↗

Hello sir,

This is a great article. But sir I have couple of questions?

1. Assume that if we have three inputs and one output future value according to the past values to next time value next time interval period is 120min) as a example

6:00:00 63 0 0 63
7:00:00 63 0 2 104
8:00:00 104 11 0 93
9:00:00 93 0 50 177

2. To predict value should I have to do time forecast ac

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees.**

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee October 18, 2018 at 6:25 am #

REPLY ↗

You could treat the other inputs as exogenous variables and use ARIMAX, or you could use another method like a machine learning algorithm or neural network that supports multivariate inputs.

Mohammad October 31, 2018 at 3:07 pm #

REPLY ↗

This is a great post, thank you very much.

I'm new in this field, and I look for simple introduction to ARIMA models in general then an article about multivariate ARIMA.

Could you please help me.



Jason Brownlee November 1, 2018 at 6:01 am #

REPLY ↗

Thanks.

I don't think I have an example of a multivariate ARIMA, maybe ARIMAX/SARIMAX would be useful as a start:

Start Machine Learning

Ramy November 2, 2018 at 10:07 pm #

REPLY ↗

Hey Jason,

I was wondering if you are aware of any auto arima functions to fine tune p,d,q parameters. I am aware that R has an auto.arima function to fine tune those parameters but was wondering if you're familiar with any Python library.



Jason Brownlee November 3, 2018 at 7:06 am #

Yes, I wrote one in Python here:

<https://machinelearningmastery.com/how-to-grid-search-arima-hyperparameters-with-python/>

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Sheldon November 6, 2018 at 1:11 pm #

Hi Jaosn.

Thanks a lot for the great tutorial!

Have followed your post : “How to Grid Search ARIMA Model Hyperparameters with Python” to fine tune the p,q and d value. Have come across the below point in the post.

“The first is to ensure the input data are floating point values (as opposed to integers or strings), as this can cause the ARIMA procedure to fail.”

My initial data is in the below format. Month and #Sales

2014-11 4504794
 2014-12 7656479
 2015-01 9340428
 2015-02 7229578
 2015-03 7092866
 2015-04 14514074
 2015-05 9995460
 2015-06 8593406
 2015-07 8774430
 2015-08 8448562

I applied a log transofrmation on the above data set to convert the numbers to flot as below:-

```
dateparse = lambda dates: pd.datetime.strptime(dates, '%Y-%m')
salesataparsed = pd.read_csv('sales.csv', parse_date
```

Start Machine Learning

```
index_col='Month',date_parser=dateparse)
salessataparsed.head()
ts_log = np.log(salessataparsed['#Sales'])
```

Below is the ts_log.head() output.

```
2014-11-01 15.320654
2014-12-01 15.851037
2015-01-01 16.049873
2015-02-01 15.793691
2015-03-01 15.774600
2015-04-01 16.490560
2015-05-01 16.117632
2015-06-01 15.966517
```

With this log value, applied the grid search approach to get the best parameters. However, I got Best ARIMA(0, 1, 0) MSE=0.023. Looks like it's acceptable. Please confirm.

Next, I have 37 Observations from Nov 2014 to 31-Dec-2015. etc. How to do this?

Also, do you have any Youtube videos explaining each step of forecasts available? It would be great if you can share.

Once again thanks a lot for the article and your help!

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee November 6, 2018 at 2:21 pm #

REPLY ↗

You can discover if your model is skillful by comparing its performance to a naive model:

<https://machinelearningmastery.com/faq/single-faq/how-to-know-if-a-model-has-good-performance>

Perhaps try standardizing or normalizing the data as well.

I don't make videos, only text-based tutorials, sorry.

I show how to use an ARIMA model to make forecasts here:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

REPLY ↗

Sheldon November 6, 2018 at 5:33 pm #

Great ! Thanks a ton Jason.

Kindly confirm if the p,q value is 0 is an acceptable scenario.

Perhaps try standardizing or normalizing the data as well : I am not sure how to proceed with this?

It would be great if you can share related article if you have any. 😊

Start Machine Learning

For now, I am going to implement the future forecasting using the above link with this ARIMA(0,1,0) and will check how it behaves. 😊



Jason Brownlee November 7, 2018 at 6:00 am #

REPLY ↗

See this post:

<https://machinelearningmastery.com/machine-learning-data-transforms-for-time-series-forecasting/>

Michelle November 13, 2018 at 1:08 am #

Hi Jason, thanks for the tutorial i am new to this and would like to predict when a customer is likely to make next purchase based on transaction amount.

Will this tutorial help me or is there any suggestion on what I can do?
Could you please advice

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee November 13, 2018 at 5:48 am #

REPLY ↗

I recommend following this process:

<https://machinelearningmastery.com/start-here/#process>

Ayub November 28, 2018 at 3:44 am #

REPLY ↗

Hi Jason,

Used your epic tutorial to forecast bookings.

I used the whole of 2017 as my data set and after applying everything in your post the predicted graph seems to be one day off i.e. prediction graph looks spot on with each data point very close to the what it should be, the only thing is is that it's a day late...is this normal? Is there something within the code that causes something like this?

Thanks



Jason Brownlee November 28, 2018 at 7:46 am #

REPLY ↗

This is a common problem, I explain more in this post:

<https://machinelearningmastery.com/faq/single-step-guide-arima-time-series-forecasting/>

Start Machine Learning

actual-time-series**Kris** November 29, 2018 at 9:09 pm #

REPLY ↗

Hi, i have had a question for a while, now this might be silly but I can't figure out what's wrong here...

So I have a timeseries data and when I used `order=(0,1,0)` that is, differencing is 1 then I get a timeseries that is ahead of time by one.

example:

input: 10, 12, 11, 15

output: 8, 9.9, 12.02, 11.3, 14.9

Now if I shift the resulting series by one timeperiod, it'll

Also, similar output can be seen is `(0,2,1)` that is, differ-

Could someone explain why is this happening and what
[numbers in example are representative not actual]

**Jason Brownlee** November 30, 2018 at 6:30 am #

X

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)

It suggests that the model is using the input as the output, this is called a persistence model:
<https://machinelearningmastery.com/faq/single-faq/why-is-my-forecasted-time-series-right-behind-the-actual-time-series>

Kris December 3, 2018 at 6:29 am #

REPLY ↗

Thanks Jason, I went through the link and it helps me see a clear picture which should have been obvious to notice but I missed it.

If you please, could also share some thoughts on...

- My model uses `order(0,1,0)`. i.e. differencing is 1. Does such model make sense for a practical scenario where we are trying to predict inventory requirement for a part (based on past consumption) that may fail in coming future (where failing of a part is totally a random act of nature).
- Also, `(0,2,1)` and `(0,1,0)` give very similar results. Is this expected in some sense. Is there any concept that I am missing here.

Thanks a lot again, for your help.

Jason Brownlee December 3, 2018
[Start Machine Learning](#)



I generally recommend using the model that gives the best performance and is the simplest.

[REPLY ↩](#)
Dhananjai Sharma December 4, 2018 at 8:08 pm #

Hello Jason!

Thank you for the tutorial. It's a good start to implementing an ARIMA model in Python. I have a question: You have used the actual data samples to update your training dataset after each prediction as given in "history.append(obs)". Now let's take a real life example when you don't have any further actual data and you use your predictions only to update your training dataset which looks like "history.append(yhat)". What will happen in this case? I am working on air quality pr the seasonal pattern in the test set but the latter does know what's your take on this.

Regards,
Dhananjai

—

Start Machine Learning X

You can master applied Machine Learning **without math or fancy degrees**.
Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)

Jason Brownlee December 5, 2018 at 6:15 am #

You can re-fit the model using predictions as obs and/or predictions as inputs for subsequent predictions (recursive).

Perhaps evaluate a few approaches on your dataset and see how it impacts model performance.

[REPLY ↩](#)
Beshoy Akram December 8, 2018 at 3:19 am #

Hi Jason ,

Thank you for the tutorial.

I have two questions :

first : why you set moving average "q" Parameter by 0 ?

second : why you set Lag value To 5 not 7 for example?

Thanks.

[REPLY ↩](#)

Jason Brownlee December 8, 2018 at 7:12 am #

They are an arbitrary configuration.

Perhaps try other configurations and compare results.

[Start Machine Learning](#)

ben December 10, 2018 at 6:31 pm #

REPLY ↗

Thank you for your great tutorial.

I know that the third output from `model_fit.forecast()` consists of the confidence interval. But how can I plot the confidence interval on the whole range automatically?

Thanks



Jason Brownlee December 11, 2018 at 7:40 am #

REPLY ↗

I believe this tutorial will help:

<https://machinelearningmastery.com/time-series-forecasting-with-python/>

Glenn Dalida December 11, 2018 at 2:06 am #

What's the difference of predicted and expect

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee December 11, 2018 at 7:48 am #

REPLY ↗

“Predicted” is what is output by the model.

“Expected” or “actual” are the true observations.

Ronald December 23, 2018 at 2:17 am #

REPLY ↗

Hey Jason,

Amazing blog, subscribed and loving it. I had a question about how you would send the output of the model to a data table in CSV?

Ramy



Jason Brownlee December 23, 2018 at 6:07 am #

REPLY ↗

You can save a Numpy array as a csv directly:

<https://docs.scipy.org/doc/numpy-1.14.0/reference/generated/numpy.savetxt.html>

Start Machine Learning

REPLY ↗

Benny Late December 23, 2018 at 6:01 am #

Hi Jason, man I love this blog.

I'm running this with a separate data set. I've shaped my dataset, but when I run the error line, I'm getting this:

`ValueError: Found array with dim 3. Estimator expected <= 2.`

What are you thoughts?

Thanks,
Benny

Shaping:

```
X_train = np.reshape(X_train, (len(X_train), 1, X_train.shape[2]))
X_test = np.reshape(X_test, (len(X_test), 1, X_test.shape[2]))
```

Code:

```
history = [x for x in X_train]
predictions = list()
for t in range(len(X_test)):
    model = ARIMA(history, order=(10,0,3))
    model_fit = model.fit(disp=0)
    output = model_fit.forecast()
    yhat = output[0]
    predictions.append(yhat)
    obs = X_test[t]
    history.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))

error = mean_squared_error(X_test, predictions)
print('Test MSE: %.3f' % error)
```

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

**Jason Brownlee** December 23, 2018 at 6:10 am #

REPLY ↗

You're data has too many dimensions. It should be 2D, but you have given it 3D data, perhaps change it to 2d!

Benny Late December 23, 2018 at 6:16 am #

REPLY ↗

Oh. I thought that's what I did with reshaping. Whoops =)

I'll hunt up some code. Thank you.

Start Machine Learning

Walid December 24, 2018 at 7:12 pm #

REPLY ↗

Hi Jason,

Thanks for this great work!

If you allow me, I have a question: how was the confidence interval calculated in the above example? I know its equation, but I do not know what are the values to be used for (sigma) and (number of samples). Thank you once more.



Jason Brownlee December 25, 2018 at 7:20 am #

REPLY ↗

You can review the statsmodels source code documentation may also be helpful.

Start Machine Learning

X

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Saravana Ayyappa January 1, 2019 at 6:44 am #

Thanks a lot Jason!

I am preparing a time series model for my capstone project. The data is different for each item, how can i deploy this as a tool? items?

Thanks in advance.



Jason Brownlee January 1, 2019 at 11:12 am #

REPLY ↗

Perhaps model each series separately?

Avd January 10, 2019 at 6:02 pm #

REPLY ↗

How many minimum data points do we require for creating accurate prediction using ARIMA model. We are predicting future cut-off values of colleges using previous records, how many years of records would we need to predict just the cutoff value of next year.



Jason Brownlee January 11, 2019 at 7:41 am #

REPLY ↗

I recommend testing with different amounts of history on your specific dataset and discover the right amount of data for modeling.

Start Machine Learning

Renu Kalra January 16, 2019 at 10:58 pm #

REPLY ↗

If I am not wrong, ACF plot is used to get MA value for ARIMA. But here, you have taken AR value as 5 using ACF plot?

Nauman Naeem January 22, 2019 at 1:07 am #

REPLY ↗

Hi Jason Brownlee!

I have been following your blog since some time and the concepts and code snippets here often come handy.

I'm totally new to time series analysis and have read some course questions from stackoverflow.

What confuses me is, to make a series stationary we can remove trend and seasonal components. Both trend and seasonal both are present in the series. Now while performing ARIMA, how do we remove both? Number of times we have performed differencing or lagged the series to remove trend and seasonal components (lag value of 12 for seasonal, lag value of 1 for trend).

For example, let say there is a dataset of monthly average temperature over a year (global warming). Now there is seasonality (lag value of 12) and trend both are present in the series. Now before performing ARIMA I need to make the series stationary. To do that I Difference twice like this:

```
differenced = series - series.shift(1) # to remove trend
```

```
double_differenced = differenced - differenced.shift(12) # to remove seasonality.
```

Now what should be passed as 'l' to ARIMA?

2? As we did double(2) differencing

or

1 or 12 as that's the value we used for shifting.

Also if you're kind enough, can you elaborate more how *exactly* did you decide the value of 'p' and 'q' from acf and pacf plots.

Or link me to some post if you have already explained that in layman terms somewhere else!

Extremely thankful for your time and effort!

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee January 22, 2019 at 6:25 am #

REPLY ↗

It might be better to let the ARIMA model perform the differencing rather than do it manually.

And, if you have seasonality, you can use SARIMA to difference the trend and seasonality for you.

If you difference manually, you don't need the model to do it again.

jaideep January 29, 2019 at 9:54 am #

Start Machine Learning

The computed initial MA coefficients are not invertible

You should induce invertibility, choose a different model order, or you can pass your own start_params.

How do I fix this error? Best ARIMA params are (4,1,3)



Jason Brownlee January 29, 2019 at 11:41 am #

REPLY ↗

Perhaps try a different configuration or try to prepare the data before modeling.

SKN January 30, 2019 at 1:21 am #

Do we have a similar function in python like w



Jason Brownlee January 30, 2019 at 8:14 am #

I wrote one here:

<https://machinelearningmastery.com/how-to-grid-series-forecasting-in-python/>

And another here:

<https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/>

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

SKN January 31, 2019 at 12:19 am #

REPLY ↗

Thank you very much, your blogs really come in handy for a beginner in python. when I run the ARIMA forecasting using above codes, getting some format error. I have tried to use Shampoo sales data too. below is the error note,

File "", line 1, in

```
runfile('C:/Users/43819008/untitled2.py', wdir='C:/Users/43819008')
```

File "C:\ProgramData\Anaconda3\lib\site-packages\spyder\utils\site\sitecustomize.py", line 880, in runfile
execfile(filename, namespace)

ValueError: time data '19019-01-2019' does not match format '%Y-%m'

I have tried all the format in excel and saved as CSV. but nothing helped me. hope you can help me.



Jason Brownlee January 31, 2019 at 5:34 am #

Start Machine Learning

REPLY ↗

Looks like an issue loading the data.

You could try removing the date column and changing the load function call to not use the custom function?

Abid Mehmod February 14, 2019 at 11:24 pm #

REPLY ↩

Hello Everyone , I want to implement ARIMA model but this error is not leaving me.

```
from . import kalman_loglike
ImportError: cannot import name 'kalman_loglike'
```



Jason Brownlee February 15, 2019 at 8:06 am #

Looks like you're trying to import a modul

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Abid Mehmod February 16, 2019 at 10:10 am #

I got that.
Thank you very very much ,

Barry A. February 21, 2019 at 3:37 am #

REPLY ↩

Hi Jason, I recently came accross your blog and really like the things I have learned in a short period of time. Machine learning and AI are still relatively new to me, but I try to catch up with your information. As the ARIMA Model comes from the statistics field and predicts from past data, could it be used as the basis of a machine learning algorithm? For example: if you would create a system that would update the predictions as soon as the data of a new month arrives, can it be called a machine learning algorithm? Or are there better standarized machine learning solutions to make sales predictions?



Jason Brownlee February 21, 2019 at 8:16 am #

REPLY ↩

Sure.

Yes, ARIMA is a great place to start.

Fredrick Ughimi February 25, 2019 at 7:55 am #

Start Machine Learning

Hello Al,

>>the last line of the data set, at least in the current version that you can download, is the text line “Sales of shampoo over a three year period”. The parser barfs on this because it is not in the specified format for the data lines. Try using the “nrows” parameter in read_csv.

```
series = read_csv('~/Downloads/shampoo-sales.csv', header=0, nrows=36, parse_dates=[0], index_col=0,
squeeze=True, date_parser=date_parser)
```

worked for me.

Thank you for posting this. I was having the same issue. This solved it.

Thanks Jason for another great tutorial.



Jason Brownlee February 25, 2019 at 2:09 pm

Thanks, I'm glad it helped.

Mo March 1, 2019 at 6:27 am #

Jason,

thank you it was very helpful in many different ways. I just want to know how you predict and how far you can predict in the future.



Jason Brownlee March 1, 2019 at 2:18 pm #

REPLY ↗

Thanks, good question. This post will show you how to predict:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

JY March 6, 2019 at 6:33 pm #

REPLY ↗

Hi Jason,

Thanks for your write-up. I've tried all the suggestions here but still getting these two errors.

```
in parser(x)
5 def parser(x):
--> 6 return datetime.strptime('190'+x, '%Y-%m')
7
```

TypeError: strptime() argument 1 must be str, not numpy.ndarray

ValueError: time data '1901-Jan' does not match format

Start Machine Learning

I removed the footer, tried with your csv file , tried with nrows but nothing worked. Please give me your valuable feedback.Thanks.



Jason Brownlee March 7, 2019 at 6:44 am #

REPLY ↗

Perhaps confirm that you downloaded the dataset in the correct format?

Charlie March 16, 2019 at 1:36 am #

REPLY ↗

i use R to get the p,q but it does work in the s
not converge even i set the p,q very small



Jason Brownlee March 16, 2019 at 7:57 am #

Hmm, maybe the R version is preparing th

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

cryptoripple March 20, 2019 at 10:56 pm #

how can I get future forecast value with arima?



Jason Brownlee March 21, 2019 at 8:15 am #

REPLY ↗

See this tutorial:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

AJIT MUNJULURU March 28, 2019 at 3:42 am #

REPLY ↗

Hi Jason,

Your materials on Time Series have been extremely useful. I want to clarify a basic question on Model results. For an ARMA(3,0) , the statsmodel prints the output as
coef P>Z

```
const c 0.00
ar.L1 x1 0.003
ar.L2 x2 0.10
ar.L3 x3 0.0001
```

Start Machine Learning

And the Data is:

Actual Daily Traffic Predicted Traffic

Jan7 100
Jan8 95
Jan9 85
Jan10 105

If I want to convert the output to a linear equation will the Predicted Traffic for Jan10 be :Pred= c+ x1*85 + 0*x2 + x3*100 ?? Appreciate your thoughts



Jason Brownlee March 28, 2019 at 8:22 am #

Great question, I have an example of making a manual prediction in my post:
<https://machinelearningmastery.com/make-manual-predictions-machine-learning-model/>

AJIT MUNJULURU March 29, 2019 at 7:45 pm #

Thank you very much Jason. That post has given me the exact result for prediction. i.e.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee March 29, 2019 at 8:49 am #

REPLY ↗

Glad it hear it.

Jay March 30, 2019 at 8:35 am #

REPLY ↗

i am newbie and trying to learn time series. getting following error, please help.

```
series = read_csv('sales.csv', delimiter=',', header=0, parse_dates=[0], index_col=0, squeeze=True,
date_parser=date_parser)
```

Traceback (most recent call last):

File "", line 1, in

```
series = read_csv('sales.csv', delimiter=',', header=0, parse_dates=[0], index_col=0, squeeze=True,
date_parser=date_parser)
```

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 678, in parser_f
return _read(filepath_or_buffer, kwds)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 446, in _read
data = parser.read(nrows)

Start Machine Learning

```

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1036, in read
ret = self._engine.read(nrows)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1922, in read
index, names = self._make_index(data, alldata, names)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1426, in _make_index
index = self._agg_index(index)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1504, in _agg_index
arr = self._date_conv(arr)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 3033, in converter
return generic_parser(date_parser, *date_cols)

File "E:\ProgramData\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 270, in generic_parser
results[i] = parse_func(*args)

File "", line 2, in parser
return datetime.strptime('190'+x, '%Y-%m')

File "E:\ProgramData\Anaconda3\lib\_strptime.py", line 557, in strptime
tt, fraction = _strptime(data_string, format)

File "E:\ProgramData\Anaconda3\lib\_strptime.py", line 528, in _strptime
(data_string, format))

```

ValueError: time data '1901-Jan' does not match format '%Y-%m'

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee March 31, 2019 at 9:22 am #

REPLY ↗

Looks like you need to download the data with numeric date format, or change the data parsing string.

Jay April 6, 2019 at 6:56 am #

REPLY ↗

Thanks, it is resolved, i have to download another file.



Jason Brownlee April 7, 2019 at 5:22 am #

REPLY ↗

Glad to hear that.

Joker Ho March 31, 2019 at 6:46 pm #

Start Machine Learning

Hi Jason!

I have a compile error: insufficient degree of freedom to estimate, when finishing my program on ARIMA in Python. Could you tell me what leads to this error? Cuz I found little answer in other solution website like stack overflow.

Hoping to hear from you!

Thank you, Jason!



Jason Brownlee April 1, 2019 at 7:48 am #

REPLY ↗

Perhaps your data requires further preparation – it can happen if you have lots of zero values or observations with the same value.

Start Machine Learning

X

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee April 9, 2019 at 2:41 pm #

REPLY ↗

Perhaps confirm that you have loaded your data correctly, as a floating point values?

Orsola April 14, 2019 at 8:20 am #

REPLY ↗

Hi Jason,

Do you know how predict from estimated ARIMA model with new data, preserving the parameters just fitted in the previous model?

I'm trying to accomplish in python something similar to R:

```
# Refit the old model with newData
new_model <- Arima(as.ts(Data), model = old_model)
```



Jason Brownlee April 15, 2019 at 7:48 am #

REPLY ↗

Yes, you can use the forecast() or predict() functions.

Start Machine Learning

More here:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Naveksha Sood April 22, 2019 at 5:53 pm #

REPLY ↗

Jason, great tutorial! I follow your blogs and book regularly and they help me a lot!
However I have some conceptual doubts that I hope you can help me with.

1. If you don't do a rolling forecast and only use the predict function, it gives us various predicted values (number of predicted values are equal to length of training data). How are the predictions made in this case? Does it use the previous predicted values to make next predictions?
2. When I validate a neural network made of one or more layers, I use the validation function and hence it uses that data to make predictions. Is this redundant there?



Jason Brownlee April 23, 2019 at 7:53 am #

Good question, ideally you want to fit the model on all data and then make a prediction.

So, in a walk-forward validation you might want to

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Karl April 24, 2019 at 1:46 am #

REPLY ↗

Hi Jason, thank you so much for all your tutorials. They have been of great help to me.

I had a question about the ARIMA model in statsmodels. If I want to select certain lags for the parameter p instead of all lags up until p how would I do it? I have not seen functionality for this in statsmodels, I wondered if you knew.

Whenever you find the time. Kind regards Karl



Jason Brownlee April 24, 2019 at 8:06 am #

REPLY ↗

You might have to write a custom implementation I'm afraid.

Naveksha Sood April 25, 2019 at 3:28 pm #

REPLY ↗

Start Machine Learning

Yes, I totally understand why we use walk forward validation, but I see a major drawback of it i.e it works great with shorter time series, however when you have a longer time series and multiple variables, it takes a really really long time to re-fit a SARIMAX model and get the predictions.

That's why what I intended to ask in the second point is, if instead of a SARIMA model, I use an LSTM model, do I still need to do walk forward validation, since it already uses the actual values up to the point of prediction.



Jason Brownlee April 26, 2019 at 8:23 am #

REPLY ↗

Yes. But you may not have to refit the model each step. I often do not.

Yarong April 26, 2019 at 6:51 am #

Hi Jason, thanks for the great post. My time series is large and inconsistent. For example, I want to know we will use to fulfill this order. We might not know the first order even yesterday since they might not be fulfilled yet. Do you have any suggestion on this? Thank you so much.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee April 26, 2019 at 8:40 am #

REPLY ↗

That sounds like a great problem.

I recommend using this framework to help think about different ways you can frame the problem for prediction:

<http://machinelearningmastery.com/how-to-define-your-machine-learning-problem/>

Naveksha Sood April 26, 2019 at 8:03 pm #

REPLY ↗

Ok, Have you covered it in any of your articles? Can you refer me to it?

Luis Zarate May 4, 2019 at 8:30 am #

REPLY ↗

Hi Jason. Thank You very much for teach how to make Forecast. Butaca i have a doubt, in this example only we have 12 prediction for 12 observations (or expected values).

On this case, i would like to know. What is the prediction to the short future.

Thank so much.

Start Machine Learning

Atte. Luis



Jason Brownlee May 5, 2019 at 6:19 am #

REPLY ↩

Perhaps this post will help:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Mitchyuuu May 8, 2019 at 12:39 pm #

REPLY ↩

Hi! Thank You for your teach.I have a problem with my multivariate data,but appear some error"TypeError:must be 1-dimensional array-like, not tuple".The history data is a list of 500*2.



Jason Brownlee May 8, 2019 at 2:12 pm #

Sounds like you might not have loaded the data correctly.

Perhaps confirm it was loaded as real values, not strings.

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

me May 8, 2019 at 9:03 pm #

REPLY ↩

Hi can you please show us some plots ,specific to ARIMA ?
thank you



Jason Brownlee May 9, 2019 at 6:41 am #

REPLY ↩

Like what exactly?

Anindya Sankar Chattopadhyay May 12, 2019 at 7:31 pm #

REPLY ↩

Hi Jason:

Thanks for this tutorial.

Just wondering how was a value of 0 was decided for q? For that don't you need the PACF plot?

Any help will be much appreciated.

Start Machine Learning

Regards,
Anindya



Jason Brownlee May 13, 2019 at 6:45 am #

REPLY ↗

I may have configured the model in this tutorial based on a trial and error.

Abhishek Mishra May 13, 2019 at 1:44 pm #

REPLY ↗

Hey man, great tutorial. I just wanted to ask you about time series analysis, I mean i am not able to understand the concepts. I am still in the learning phase.



Jason Brownlee May 13, 2019 at 2:34 pm #

Thanks, we expect the residual error to be large. It's likely your model is missing something important.

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Greg Houston May 14, 2019 at 9:19 am #

REPLY ↗

Hi Jason,

I'm considering buying your book. Will the code examples be up to date seeing as it is now 2019? Also, what success have you had forecasting several time series, lets say 30, with the same model. Would you suggest more of an ensemble approach?

gregory houston May 14, 2019 at 9:23 am #

REPLY ↗

Oh, is any other reading material you would suggest? We did not cover time series in my masters program, so I'm a newbie.



Jason Brownlee May 14, 2019 at 2:28 pm #

REPLY ↗

Yes, you can get started with the basics here:

<https://machinelearningmastery.com/start-here/#timeseries>

Start Machine Learning

Advanced topics here:

https://machinelearningmastery.com/start-here/#deep_learning_time_series



Jason Brownlee May 14, 2019 at 2:28 pm #

REPLY ↗

Yes, I update the books frequently. After purchasing, you can email me any time to get the latest version.

Hmm, 30 is not a large number, it might be best developing a separate model for each and compare the results to any model that tries to learn across the series.

mbelahcen May 22, 2019 at 8:01 pm #

Hello Jason,

I still don't understand why the forecast is one step ahead expected, If for instance my model predicts very well that the model is good or I should go on tuning to take off the lag. In the case of the lag, the line print('predicted=%f, expected=%f') is not representative of the actual comparison?

Thanks

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee May 23, 2019 at 6:00 am #

REPLY ↗

I think you are describing a persistence forecast, this might help:

<https://machinelearningmastery.com/faq/single-faq/why-is-my-forecasted-time-series-right-behind-the-actual-time-series>

Bankole Akinuli May 29, 2019 at 6:52 pm #

REPLY ↗

Dear Prof. Kindly help to write the equation for ARIMA (0,0,0); (0,1,0); (1, 0,1), VARMA (1,1), and ARMA (5,4)

Thanks



Jason Brownlee May 30, 2019 at 8:59 am #

REPLY ↗

I cannot write equations for you, this would be trivial though, start with the ARIMA equation and add the terms you need.

Start Machine Learning

Perhaps get a good textbook on the topic.

Bankole Akinuli May 31, 2019 at 11:22 am #

REPLY ↗

I appreciate your view and advice, sir. Please, suggest relevant textbook on ARIMA and how or where i can get one. Warmest regards.



Jason Brownlee May 31, 2019 at 2:44 pm #

REPLY ↗

Here are some suggestions:

<https://machinelearningmastery.com/books-on-time-series-forecasting/>

Mithlesh Patel June 3, 2019 at 8:46 pm #

Thanks Jason for overview of ARIMA model very well explained. In below code, are you creating model again and fitting it on test dataset. In other algorithms we generally create model and fit on training dataset and then predict on test dataset.

```
for t in range(len(test)):
    model = ARIMA(history, order=(5,1,0))
    model_fit = model.fit(disp=0)
```

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee June 4, 2019 at 7:51 am #

REPLY ↗

Yes, this is called walk forward validation and it is the preferred way for evaluating time series models.

You can learn more here:

<https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>

mickael June 4, 2019 at 3:17 am #

REPLY ↗

Hi,

Great post, and blog in general!

I have a question regarding the practical use of ARIMA. Is it possible to use it (after fitting on some dataset), to test the prediction from any new input data, just like [this](#)?
For instance, I have one year of temperature data on w...

Start Machine Learning

point per day) for autoregression. Then, to use the model in production, I want to simply store the last 7 days and use them to predict the next one. (Without the need to fit my model again and again each day)

Many thanks,

Mickaël



Jason Brownlee June 4, 2019 at 7:57 am #

REPLY ↗

Yes, here is an example:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Dav June 8, 2019 at 2:35 am #

Hi

I am new to Auto Regression and Python. Great article

My question is around how much historical time series prediction? For example, if I have two years worth of data, does it really make a difference if I use all of it in a training set for a model? (assuming lags would be less than 10?)

Also, how should I think about accounting for seasonality. I understand I would need to remove it from the time series in order to get a reasonable prediction. Should I then have an overlay on top of predicted values to implement the impact of seasonality?

Thanks

Dav

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee June 8, 2019 at 7:03 am #

REPLY ↗

It depends on the dataset, try different amounts of history to see how sensitive your model is to dataset size.

You can remove seasonality or let the model remove it for you in the case of SARIMA. Any structure removed must be added to predictions, it is easier to let the model do it for you perhaps.

Dav June 10, 2019 at 6:23 pm #

REPLY ↗

got it, thanks

Start Machine Learning

baktr_ June 15, 2019 at 8:19 pm #

REPLY ↗

Hi Jason, thanks for your blog, i'm newbie, i have a question: model ARIMA is machine learning?



Jason Brownlee June 16, 2019 at 7:12 am #

REPLY ↗

It was developed in statistics and borrowed in machine learning.

The intent makes it machine learning, more here:

<https://machinelearningmastery.com/faq/single-faq/how-are-statistics-and-machine-learning-related>

Adi June 28, 2019 at 11:58 pm #

Hi Jason,

Thank you for this. What are some good strategies to know consecutive zeros can be a problem for AR algo multiplicative version.. Is there any useful resource you denormalizing?

Also, if I have a lot of time series to forecast for, where indicators that will be helpful to describe the time serie

Thanks

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee June 29, 2019 at 6:56 am #

REPLY ↗

Good question – it is probably going to be domain specific how to best handle it.

Test many things.

Try small random values?

Try impute with mean value?

Try alternate methods, like neural nets?

...

staph July 3, 2019 at 6:23 pm #

REPLY ↗

Hi Jason how long does it take to fit a model, code is taking ages at the fit line

```
model_fit = model.fit(disp=0)
```

Start Machine Learning



Jason Brownlee July 4, 2019 at 7:42 am #

REPLY ↗

It really depends on the size of the dataset.

sundus July 7, 2019 at 5:23 am #

REPLY ↗

much appreciated..



Jason Brownlee July 7, 2019 at 7:55 am #

REPLY ↗

You're welcome.

Niloofar July 11, 2019 at 7:19 pm #

Hi, thanks Mr.Brownlee for your great posts. I forecast NA values in a dataset? I mean can it handle

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 12, 2019 at 8:34 am #

REPLY ↗

No.

More on missing values here:

<https://machinelearningmastery.com/handle-missing-timesteps-sequence-prediction-problems-python/>

Hugo Pire July 20, 2019 at 7:06 am #

REPLY ↗

Hello Jason and thank you for your great posts

I am trying to fit an ARIMA model to an company invoices timeseries. It has a timestamp (not regulary spaced) and a value that can be negative or positive – with a large interval.

Do I have to interpolate in order to have regular intervals? If I use a naive solution, as group by day, I get a lot of zero values.

Could you help me?



Jason Brownlee July 20, 2019 at 10:59 am #

REPLY ↗

Start Machine Learning

I have some suggestions here:

<https://machinelearningmastery.com/faq/single-faq/how-do-i-handle-discontiguous-time-series-data>

Toraskar August 5, 2019 at 5:38 pm #

REPLY ↗

Hello Jason and thank you for your posts
can you please make same project for stock price prediction using Arima model??



Jason Brownlee August 6, 2019 at 6:30 am #

Sorry, I choose not to give examples for the waste of time.

Amin Nasri August 6, 2019 at 1:45 am #

Hi,

How can I specify which lags the model uses, for instance?
Thanks in advance for your reply.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 6, 2019 at 6:41 am #

REPLY ↗

It will use all lags in between.

To use otherwise, you may have to develop your own implementation.

Nir August 6, 2019 at 3:36 pm #

REPLY ↗

Hi Jason,

Thanks for writing such a detailed tutorial.
In your text, you mentioned that “A crude way to perform this rolling forecast is to re-create the ARIMA model after each new observation is received.” Is there another way to do so without retraining the model? Is there a way just to update the inputs (and not the parameters)? After our first prediction, we get the true value and the prediction error we now use the new information to predict the next step (without retraining)?

Thanks!

Start Machine Learning



Jason Brownlee August 7, 2019 at 7:41 am #

REPLY ↗

Yes, you can forecast for the future interval directly without updating the model, e.g. `model.forecast()` or `model.predict()`

Is that what you mean?

Nir August 20, 2019 at 9:08 am #

REPLY ↗

Hi Jason,

Thanks for the rapid reply, and sorry for not being clear. If I understood it right, `model.forecast()` will forecast the next 4 months' worth of data sampled every 1 minute (or 10 minutes). If my training dataset ends at t+1, it will not be available and can help to predict t+2. I see three ways:

1. Use `model.predict()` for 2 samples, but then retrain the model every time the parameters have changed.
2. As in your example, retrain the model every time it is running this in real-time and don't want to retrain the model by adding new observation but without updating the parameters have changed.
3. Update the model input without retraining the model by adding new observation but without updating the parameters have changed.

Thanks,

Nir

Start Machine Learning

X

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 20, 2019 at 2:10 pm #

REPLY ↗

Not quite.

You can use `forecast()` and specify the number of steps required.

You can use `predict()` to specify an interval of dates or time steps.

See this post:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Yes, perhaps try with and without refitting the model, and try refitting every hour, day, week and compare.

Amin Nasri August 6, 2019 at 6:59 pm #

REPLY ↗

Thanks for your response. In Matlab, you can choose specific lags.

When I am trying to use all the lags in between it takes

Start Machine Learning



Jason Brownlee August 7, 2019 at 7:45 am #

REPLY ↗

Yes, the statsmodel implementation could use some improvement.

Marcel August 14, 2019 at 9:20 pm #

REPLY ↗

Hey Jason, I'm currently doing my thesis on forecasting electricity load and am also using the ARIMA from statsmodels.

You mentioned, that the reestimation you are doing for compute a new ARIMA for every step. What would be on the training data and after each forecasting step ap forecasting the next step (without having to fit the model) might this work with the initialize() function of the ARIM Btw, thanks a lot for this excellent tutorial, it's really we



Jason Brownlee August 15, 2019 at 8:07 am #

Ideally fitting the model only when needed refit is required.

A fit model can forecast any future period, e.g. see forecast() and predict().

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Rittick Sinha Roy August 18, 2019 at 12:53 am #

REPLY ↗

Hey Jason I'm doing a project on crime prediction and wanted to use ARIMA model could you help me in understanding what kind of factors would predict the trend



Jason Brownlee August 18, 2019 at 6:47 am #

REPLY ↗

If you are using an ARIMA, it will remove the trend via differencing. Perhaps try different d values.

Or, perhaps try a grid search of different model parameters;

<https://machinelearningmastery.com/grid-search-arima-hyperparameters-with-python/>

Asieh August 28, 2019 at 5:24 am #

Start Machine Learning

Hi Jason,

What made you choose 5 lags for this dataset? In other words, what is the threshold we should choose for autocorrelation? Is it above 0.5? What about negative correlation? So in this example, the absolute value of the negative correlation is <0.5. How would we choose the number of lags (p) if it was say -0.52?

Thanks,
Asieh



Jason Brownlee August 28, 2019 at 6:43 am #

REPLY ↗

Perhaps test a range of values and see w

Dongchan Christopher Kim September 7, 2019 at 10:30 pm #

Loving the post! It definitely helps me grab off sales of an object where the growth path jumps up and have some smoothing ways for the projection rather than simple and efficient well enough. Thank you very much!

Start Machine Learning

X

You can master applied Machine Learning **without math or fancy degrees.**

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee September 8, 2019 at 5:15 am #

REPLY ↗

Thanks, I'm happy that it heled!

mark patzelt September 11, 2019 at 5:35 pm #

REPLY ↗

Hello Jason. I am new to Series Forecasting in Python. I would like to dig into it and learn how to forecast time series. I have recreated your ARIMA sample with my own data and it worked. I have a unix time series and would need to forecast the next 5 future values. I have not fully grasped the concept of predicted/expected and how I can get these future values. Did I misunderstand the model? I will buy your ebook, but maybe your response will help me proceed fast.



Jason Brownlee September 12, 2019 at 5:15 am #

REPLY ↗

Perhaps this post will help:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Start Machine Learning

mark patzelt September 12, 2019 at 6:42 pm #

REPLY ↗

One more question: I am using a time series with a frequency of 1 minute. The series is correctly setting a DateTimeIndex in col 0 and there seem to be no values missing. When I call ARIMA I get this message: "ValueWarning: No frequency information was provided, so inferred frequency T will be used. % freq, ValueWarning)". None of your examples are based in 1 minute frequencies. Is it not possible to work with 1 minute time series with ARIMA?



Jason Brownlee September 13, 2019 at 5:40 am #

REPLY ↗

Good question, I don't have an example about the frequency as long as it is consistent.

mark patzelt September 13, 2019 at 6:00 pm #

All sorted. I am using the unix timestamp with your book and have already included the frequency information. I have implemented the grid search and have assumed that the order combination is the best. I assume that the order combination is the best (considering that the dataset has been sorted).

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee September 14, 2019 at 6:12 am #

Correct. Test different orders and see what works well/best for your specific dataset.

mark patzelt September 12, 2019 at 4:34 pm #

REPLY ↗

Thank you for the link. This will definitely help. I am half way through your book on Time Series Forecasting and will get there too, I guess. Your book is well written, hands on. Ta



Jason Brownlee September 13, 2019 at 5:38 am #

REPLY ↗

Thanks Mark.

Start Machine Learning

REPLY ↗

Brian McAAdams October 3, 2019 at 6:37 am #

Hey Jason,

Interested if you know what type of correlation pandas.plotting.autocorrelation_plot is using. I get a different result with this data set using pandas.Series.autocorr over 35 lags than I do from autocorrelation_plot.

This is a copy paste of the autocorrelation_plot code to retrieve the data:

```
from pandas.compat import lmap
series = shampoo_df.Sales

n = len(series)
data = np.asarray(series)
mean = np.mean(data)
c0 = np.sum((data - mean) ** 2) / float(n)

def r(h):
    return ((data[:n - h] - mean) *
           (data[h:] - mean)).sum() / float(n) / c0

x = np.arange(n) + 1
y = lmap(r, x)
```

There isn't any information I can find about why they would be different. It could be it, but it isn't. And mathematically float(n) can't be zero so that wasn't caught.

Anyway, if you could shed any light on why pandas.Series.autocorr is different than pandas.plotting.autocorrelation_plot that would be very helpful!

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

**Jason Brownlee** October 3, 2019 at 6:55 am #

REPLY ↗

I believe it is simple linear correlation, i.e. pearsons.

Minor differences in implementation can cause differences in result, e.g. rounding errors, choice of math libs, etc.

Karan Sehgal November 4, 2019 at 12:54 am #

REPLY ↗

Hi Jason,

1) ARIMA model works on three parameters – Auto-regression, Differencing and Moving average. So does the ARIMA model makes three separate columns like – one for AR, another for Differencing and and other for Moving average separately or it does only one column and does all the above operations on the same column only (AR, I, MA) ?

Start Machine Learning

2) If ARIMA makes separate columns like (AR, I and MA) for forecasting, then should we also do the same thing to forecast time series using supervised machine learning or we can create only one column with all the operation (AR, I and MA) done on that column only.

Thanks.



Jason Brownlee November 4, 2019 at 6:47 am #

REPLY ↗

It does not create different columns, it creates “transformed inputs” to a linear model that is fit.

Karan Sehgal November 4, 2019 at 1:39 pm #

so there are different columns for tran



Jason Brownlee November 5, 2019 at 1:39 pm #

I don't follow your question, sorry

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Karan Sehgal November 6, 2019 at 2:10 am #

I mean to say if in ARIMA model our values for (p,d,q) is (2,1,2) then it will create variables – two variables for Auto Regression i.e for Lag 1 & Lag 2 and two variables for Moving Average i.e MA1 and MA2 and all the variables created i.e AR1, AR2, MA1 and MA2 will be differenced one time as value of d is 1.

do we need to difference the value for y variable also ?



Jason Brownlee November 6, 2019 at 6:42 am #

From memory, yes I believe so. Perhaps confirm.

Karan Sehgal November 6, 2019 at 5:00 am #

REPLY ↗

Why ACF and PACF plots applied on stationary Series only ?

Start Machine Learning



Jason Brownlee November 6, 2019 at 6:47 am #

REPLY ↗

To help you see the signal that the model will learn, and not get distracted by trend/seasonality which get in the way.

Karan Sehgal November 6, 2019 at 5:07 am #

REPLY ↗

to make the series stationary, we deseasonalize the series by dividing it with seasonal index and difference it for detrending it. Now the same stationary series need to be used for both x and y variables (Dependent and independent variables) and if so then ~~we have to reverse the above process to get the original data.~~

I am following the below approach to make the series
The process of deseasonalizing – dividing it with season



Jason Brownlee November 6, 2019 at 6:48 am #

Correct.

Examples here:

<https://machinelearningmastery.com/remove-trends-seasonality-difference-transform-python/>

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Karan Sehgal November 6, 2019 at 6:36 am #

REPLY ↗

Why we require a stationary data series for time series forecasting ?



Jason Brownlee November 6, 2019 at 6:49 am #

REPLY ↗

To learn the signal in the data.

Paul November 8, 2019 at 6:03 pm #

REPLY ↗

Hi Jason,

It seems that the most of methods described in this tutorial are meant for testing on the data we already have. How about multi step prediction? Is there any simple way to extend some of the methods to perform multi step forecast? Thanks in advance.

Start Machine Learning



Jason Brownlee November 9, 2019 at 6:11 am #

REPLY ↗

Call model.forecast()

Here is an example:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Ahmed November 20, 2019 at 4:30 am #

REPLY ↗

Hi, Jason,

What is the reason here by multiplying the length with 0
(size = int(len(X) * 0.66))
train, test = X[0:size], X[size:len(X)]



Jason Brownlee November 20, 2019 at 6:22 am #

To split the data into train and test sets.

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Alexander Dautzenberg November 24, 2019 at 2:51 am #

REPLY ↗

Hi Jason,

for my bachelors thesis I need to generate “Sample months of solar radiation”. The problem is as follows: I've got ten years of historic hourly solar radiation. It's in a DataFrame where every column represents an hour of a day and every row is one day (so starting with the first of January of the first year and ending with the 31st of January of the last year). Now I need to feed my data into a SARIMA Model for each month of the year so that I can use it to generate a fictive month of solar radiation. I want to generate 1000 years and they should each be a bit different, have some kind of a random component to them.

Do you have any idea how to do this?

If I just feed it the DataFrame as is, it returns the “Invalid value for design matrix. Requires a 2- or 3-dimensional array, got 1 dimensions” error.

If I flatten the DataFrame (.values.flatten()), I think it doesn't “see” the seasonality and returns an array as long as the input data when the predict() method is called



Jason Brownlee November 24, 2019 at 9:22 am #

REPLY ↗

This might help you think about the data:

<https://machinelearningmastery.com/time-series-fo...>

Start Machine Learning

An ARIMA/SARIMA/etc model expects one sequence of observations. It will transform the data into a supervised learning problem for you.

Does that help?

Matt November 29, 2019 at 3:06 pm #

REPLY ↩

In your example you re-fit the model every timestep to do a rolling forecast. This is horrendously inefficient of course—how can this be avoided? Can't the model be applied to a window of the test set data, and a prediction of next step generated, without re-training it?



Jason Brownlee November 29, 2019 at 6:18 pm #

Yes, you could fit the model once and use the most recent obs in the selection of coefficients.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Matt December 2, 2019 at 4:38 am #

Thanks a lot for your reply. How do you do the same way I would with e.g. RNN-train it up, then feed it arbitrary subsequences from a test dataset and generate the predicted next item in each case. As obvious as a use case as this is, I haven't been able to see how to effectively do this. The predict method, for example, doesn't seem to take in the current subsequence, which seems bizarre. I must not be understanding something, but what?



Jason Brownlee December 2, 2019 at 6:10 am #

REPLY ↩

You can fit the model once on the training dataset and make predictions by calling predict() and specifying the interval in the future (beyond the end of the training set) to predict.

Predict will take any future contiguous sequence of steps to predict.

Also this might help:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Matt December 2, 2019 at 12:29 pm #

Thanks for the link to your other tutorial, that also was very helpful. But it seems to confirm that Python's ARIMA can only

Start Machine Learning

that example, let's suppose I want to predict days 8-14 past the end of the training set—at this point I'd want to take the real data from days 1-7 into account. But apparently I'd have to retrain the model, with the training data now extended to include these days 1-7. This makes no sense to me, I would have thought that an ARIMA model, once all its coefficients are determined, could be applied to any arbitrary sequence. A (7,0,1) model should need just the prior seven days to make a prediction right? *Any* prior 7 days. Help.



Jason Brownlee December 2, 2019 at 1:55 pm #

Nice.

Yes, I show how to pull the coefficients
any help:

<https://machinelearningmastery.com/reverse-engineer-arima-model-coefficients/>

I did it to help show how the model works
that makes predictions on demand if you

Matt December 2, 2019 at 3:48 pm #

Wow you have a tutorial for everything!

back I recommended your computer vision book to someone at work (on the strength of what I'd seen in the tutorials), who in fact went and bought it from you. Glad now more than ever that I did.

Anyway I'm surprised that what I want here isn't more of a standard use case. But the idea of course is that to get a good sense of how well the model does, and compare with other models, then I need to generate lots of short-term predictions from a test set, and do so efficiently.

Suggestion—put a link to this tutorial in the original ARIMA blogpost that launched this thread.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee December 3, 2019 at 4:47 am #

Thanks!

Appreciate the suggestion.

Khan December 3, 2019 at 1:08 am #

How we can use ARIMA with multiple input variables?

Start Machine Learning

REPLY ↗



Jason Brownlee December 3, 2019 at 4:53 am #

REPLY ↗

It is called VARIMA, see this:

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

Ef December 14, 2019 at 10:50 pm #

REPLY ↗

How we can remove AutoCorrelation of our data?



Jason Brownlee December 15, 2019 at 6:05 am #

Differencing can remove trends, seasonal

Start Machine Learning



You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Shubh December 20, 2019 at 6:07 pm #

PLS HELP!

I have a datetime stamp column and the Power consumed against it. I used ARIMA to forecast Load consumption using:

```
results_ARIMA = results_AR.forecast(steps=24)
```

I'm getting the result as:

```
(array([2.29239839, 2.26938029, 2.25877423, 2.25559929, 2.25846445,
       2.26670267, 2.27794598, 2.28892255, 2.29774307, 2.30384135,
       2.30707076, 2.30747149, 2.30555066, 2.30218792, 2.29826336,
       2.29444756, 2.29121391, 2.28885976, 2.28748131, 2.28698461,
       2.28715699, 2.28774348, 2.28849448, 2.2891958 ]),
array([0.02200684, 0.05321806, 0.08660913, 0.11822268, 0.14836925,
       0.17610274, 0.19985402, 0.21907622, 0.23444159, 0.24686324,
       0.2570923 , 0.26579143, 0.2735913 , 0.28101168, 0.28840908,
       0.29599757, 0.30388675, 0.31209504, 0.32055845, 0.32915964,
       0.33776759, 0.34626676, 0.35457151, 0.36263193]),
array([[2.24926578, 2.335531 ],
       [2.16507481, 2.37368577],
       [2.08902345, 2.428525 ],
       [2.02388709, 2.48731149],
       [1.96766606, 2.54926284],
       [1.92154764, 2.6118577 ],
       [1.8862393 , 2.66965266],
       [1.85954104, 2.71830405],
```

Start Machine Learning

```
[1.838246 , 2.75724014],  
[1.81999828, 2.78768442],  
[1.80317912, 2.81096241],  
[1.78652986, 2.82841311],  
[1.76932157, 2.84177976],  
[1.75141515, 2.85296069],  
[1.73299196, 2.86353476],  
[1.71430298, 2.87459214],  
[1.69560682, 2.88682099],  
[1.67716473, 2.90055479],  
[1.6591983 , 2.91576432],  
[1.64184358, 2.93212565],  
[1.62514468, 2.94916929],  
[1.6090731 , 2.96641385],  
[1.59354709, 2.98344188],  
[1.57845029, 2.99994131]]))
```

Why is that in that format? I only want a single column

shubh December 20, 2019 at 9:20 pm #

I figured it out. It was showing the arrays
`results_ARIMA = results_AR.forecast(steps=24).values`

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.
 Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee December 21, 2019 at 7:11 am #

REPLY ↗

Yes, correct! Well done.



Jason Brownlee December 21, 2019 at 7:08 am #

REPLY ↗

I believe it returns point forecasts and a prediction interval.

This tutorial will help:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Mauro Júnior December 26, 2019 at 6:35 am #

REPLY ↗

Hello Jason Brwnlee! Congratulations on the article, it is very well explained and easy to understand! I know this is not the purpose of this publication, but I would like to share with you my lines of code, which I adapted from yours, trying to develop th

Start Machine Learning

successful at forecasting and am unable to find what is missing so I can reproduce this study with AutoArima. Can you help me?

```
X = series.values
size = int(len(X) * 0.66)
train, test = X[0:size], X[size:len(X)]
history = [x for x in train]
predictions = list()
for t in range(len(test)):
    model = auto_arima(train, trace = True, error_action='ignore', suppress_warnings=True)
    model_fit = model.fit(train)
    output = model_fit.predict(n_periods=len(test) + 7)
    yhat = output[0]
    predictions.append(yhat)
    obs = test[t]
    history.append(obs)
    print('predicted=%f, expected=%f' % (yhat, obs))
error = mean_squared_error(test, predictions)
print('Test MSE: %.3f' % error)
#plot
pyplot.plot(test)
pyplot.plot(predictions, color='red')
pyplot.show()
```

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

START MY EMAIL COURSE



Jason Brownlee December 26, 2019 at 7:43 am #

REPLY ↗

Thanks.

Perhaps this will help with making a prediction:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

clare antony January 2, 2020 at 9:11 pm #

REPLY ↗

Can i able to predict the job's waiting time , i took a dataset from grid5000 and i feel really hard to handle the correlation of the dataset , Can you able to give any suggestion



Jason Brownlee January 3, 2020 at 7:28 am #

REPLY ↗

Perhaps try and compare results of different methods?

Start Machine Learning

vania todorova January 7, 2020 at 10:05 am #

REPLY ↗

hey Jason! when i do a prediction on my arima model it simply shows a straight line and no ups and downs. i tried both exponential smoothing and arima and same.. any ideas? thanks!



Jason Brownlee January 7, 2020 at 1:46 pm #

REPLY ↗

Perhaps try scaling the data?

Perhaps try alternate configurations of the model?

Perhaps try other models?

Perhaps compare results to a naive model?

Perhaps the series is not predictable?

vania todorova January 8, 2020 at 2:50 am #

ya , its very random data..

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee January 8, 2020 at 8:30 am #

It might be a random walk or simply random.

vania todorova January 9, 2020 at 8:41 am #

REPLY ↗

exponential smoothing is working well as well as lstm for time series except when i make a prediction beyond the test data is blank..



Jason Brownlee January 9, 2020 at 1:41 pm #

REPLY ↗

Nice work!

vania todorova January 9, 2020 at 10:06 am #

REPLY ↗

i have this from the test file :

```
test['forecast'] = mod_fit.predict(start=1, end=51)
test[['Counts', 'forecast']].plot(figsize=(12, 8)) and that p
```

Start Machine Learning

test file it doesnt work:

```
test['a'] = mod_fit.predict(start=51, end=81)
```

```
test['a'].plot(figsize=(12, 8))
```

what am i doing wrong?



Jason Brownlee January 9, 2020 at 1:50 pm #

REPLY ↗

The predict function would take an index at the end or beyond the end of the training set.

Perhaps try using forecast() instead and see examples here:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

vania todorova January 10, 2020 at 5:04 am #

that worked . a little better with arma but the c
cuz no trend or seasonality is a straight line again. tha



Jason Brownlee January 10, 2020 at 7:29 am #

You're welcome!

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Haitham Fawzy January 25, 2020 at 4:46 am #

REPLY ↗

Hi Jason

How to Create a hybrid ARIMA and SVM Model for Time Series Forecasting in Python?



Jason Brownlee January 25, 2020 at 8:44 am #

REPLY ↗

Perhaps feed ARIMA outputs into an SVR model?

Jake February 4, 2020 at 12:57 am #

REPLY ↗

Hi Jason, you have used test data in observation and used it further to train the algorithm, so, it might not be that helpful in situations where we have to predict values for a further couple of months without new observations.

Start Machine Learning

Is this correct or am I missing something? If yes, can you give a few suggestions on how to proceed in this case?



Jason Brownlee February 4, 2020 at 7:57 am #

REPLY ↗

Not quite, we are using walk forward validation:

<https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>

More on making out of sample predictions:

<https://machinelearningmastery.com/make-sample-forecasts-arima-python/>

Mars February 4, 2020 at 10:28 pm #

Hi Jason!

Many thanks for your helpful articles.

In the above article, you set the AR parameter as 5. In on y-scale. What is AR parameter in that case?

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

REPLY ↗



Jason Brownlee February 5, 2020 at 8:10 am #

Perhaps your fit model is different?

Mars February 6, 2020 at 8:19 am #

REPLY ↗

I haven't fit any model yet, It is just the beginning where I am plotting the autocorrelation of series to get the value of the AR parameter/lag. In the above article, after plotting the autocorrelation graph, you wrote 'A good starting point for the AR parameter of the model maybe 5'. My question is " Did you choose this value from the x-axis where the plot-line enters in the highlighted dashed region?



Jason Brownlee February 6, 2020 at 8:34 am #

REPLY ↗

Exactly!

See this:

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Start Machine Learning

Devarshi Goswami February 12, 2020 at 10:21 pm #

REPLY ↗

Why does this warning come up when i fit the arima model??

ValueWarning: No frequency information was provided, so inferred frequency MS will be used.
% freq, ValueWarning)



Jason Brownlee February 13, 2020 at 5:40 am #

REPLY ↗

I don't know, perhaps try posting to stackoverflow?

Devarshi Goswami February 13, 2020 at 6:00 am #

Hello again , sir
Thanks for the suggestion

<https://imgur.com/a/H1BJdJL>

The auto correlation plot of my data looks like
and what should my p and q be according to you

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee February 14, 2020 at 6:28 am #

REPLY ↗

Perhaps grid search a range of values and discover what works best – it is the most reliable approach I've found.

Devarshi Goswami February 19, 2020 at 6:37 pm #

I do not understand why you are doing the

```
obs = test[t]
history.append(obs)
```

part in the loop that fits and forecasts ARIMA.?



Jason Brownlee February 20, 2020 at 6:08 am #

REPLY ↗

To add the last observation to the history and make it available to the model for the next one-step prediction.

Start Machine Learning

It is an assumption of the test harness that the prior observation is available prior to making a one step forecast.

Ganesh February 15, 2020 at 5:18 pm #

REPLY ↩

Hello Jason,

Did you ever discuss how to predict ETA of production jobs? the jobs might dependent on the other jobs. So before prediction the current job ETA we need to predict the dependent jobs if they are not completed yet.



Jason Brownlee February 16, 2020 at 6:05 am #

That sounds like a nonlinear forecast. I re

Start Machine Learning

X

You can master applied Machine Learning

without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Vikrant Nag February 19, 2020 at 5:02 am #

REPLY ↩



Jason Brownlee February 19, 2020 at 8:07 am #

Perhaps post your question to stackoverflow.

Mujde February 23, 2020 at 5:43 am #

REPLY ↩

Dear Jason,

Thank you very much for this tutorial. I have got single input single output time series data for which I would like to develop a model. I have 4 different samples (single input-single output time series taken at 4 different happenings of the event). I want to model fit to this data. Should I use ARIMA? How will I handle 4 different samples? Many thanks



Jason Brownlee February 23, 2020 at 7:34 am #

REPLY ↩

Perhaps test different models and discover

Start Machine Learning

Mars February 24, 2020 at 7:34 pm #

REPLY ↗

Dear Jason!

I love reading your articles. I wonder if you have written something about classification using CNN?



Jason Brownlee February 25, 2020 at 7:45 am #

REPLY ↗

Thanks.

Yes many examples for time series and text classification
<https://machinelearningmastery.com/cnn-models-for-classification/>

jayasmita mahapatra February 26, 2020 at 4:07

Thank you Jason for such a wonderful explanation

I am new to time series and have some doubt related to it.

I have learnt that the p is estimated from pacf plot and q is estimated from acf plot.

So as per your explanation you have used only autocorrelation which states that lag =5 .But how about MA order which is estimated to 0.Please clarify on the same.

Start Machine Learning

X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee February 26, 2020 at 8:26 am #

REPLY ↗

Learn how to configure them manually here:

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

David Sanchez March 20, 2020 at 11:46 pm #

REPLY ↗

Great tutorial!

I'm applying an ARIMA model to perform predictions of power in an energy plant. When my time series increases or decreases drastically, my model gives predictions that might be lower than zero or higher than the maximum capacity of the energy plant, which is not physically possible.

Is it possible to configure my model so it doesn't predict values lower or higher than certain pre-defined thresholds?

Start Machine Learning



Jason Brownlee March 21, 2020 at 8:25 am #

REPLY ↗

Perhaps write custom code to interpret the predictions made by the model instead?

Sai April 4, 2020 at 5:30 pm #

REPLY ↗

For making a prediction using any model containing residual(MA, ARIMA, etc,) We need actual(True) values, but in real life when we are using the model for forecasting we don't have them (right?), then how do we generate the residuals to make the predictions ??

Thank you



Jason Brownlee April 5, 2020 at 5:41 am #

You make predictions based on the observed future.

Htet Htet April 7, 2020 at 3:01 am #

Hi Jason,

Thanks for your great tutorials and useful posts.

I would like to do stock price prediction with ARIMA model.

After training the model, I would like to do 5 days ahead prediction.

How to do 5 days ahead prediction?

from the following function

```
predictions = model_fit.forecast(steps=5) [0]
```

we will get 5 consecutive values such as [8.01553836, 8.02257257, 8.01886069, 7.85799964, 7.91102623].

the fifth value 7.91102623 is the 5 days ahead prediction?

Thank you so much Jason.



Jason Brownlee April 7, 2020 at 5:56 am #

REPLY ↗

You're welcome.

See this:

<https://machinelearningmastery.com/faq/single-faq/can-you-help-me-with-machine-learning-for-finance-or-the-stock-market>

Start Machine Learning

vivek April 11, 2020 at 5:30 pm #

REPLY ↗

Great tutorial!

But I am continuously getting the below error:

py:512: ConvergenceWarning: Maximum Likelihood optimization failed to converge.
Check mle_retvals "Check mle_retvals", ConvergenceWarning)

I am not sure why it is not converging. I even tried changing solver but no luck.

Any body can please help!

Regards

Vivek



Jason Brownlee April 12, 2020 at 6:15 am #

It is a warning, not an error, and you can s

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Dmitriy April 14, 2020 at 6:22 pm #

Hi, Jason. I've read the article, but I still not sure why did you pick AR(5) as 5 at the start?
Autocorrelation shows significance at first 5 lags, but I'm not sure I see how it is related to the AR?



Jason Brownlee April 15, 2020 at 7:56 am #

REPLY ↗

This can help choose hyperparameters:

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

In practice, it might be better to grid search values.

Dhan April 18, 2020 at 1:14 am #

REPLY ↗

Can we use this method to predict the no. of COVID cases for 1 week/ 1 month ahead making use of total no. of cases for every day for last 2 months.



Jason Brownlee April 18, 2020 at 6:03 am #

REPLY ↗

A simple exponential model via the growth required.

Predicted2020 May 4, 2020 at 7:08 am #

REPLY ↗

Despite you are having many comments but I still wanted to tell you your tutorial is great! Thanks for it Jason!



Jason Brownlee May 4, 2020 at 7:43 am #

REPLY ↗

Thanks!

Prisilla May 5, 2020 at 6:35 pm #

Hi Jason

When i run the first part of the code “def parser(x):
return datetime.strptime('190'+x, '%Y-%m')

```
series = read_csv('April-Stata.csv', header=0, parse_date_parser=parser)
print(series.head())
series.plot()
pyplot.show()
```

I am getting this error

what can be the reason, i searched to debug but no please

```
5 def parser(x):
--> 6 return datetime.strptime('190'+x, '%Y-%m')
7
```

TypeError: strptime() argument 1 must be str, not numpy.ndarray

During handling of the above exception, another exception occurred:

Thanks and waiting for reply

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee May 6, 2020 at 6:23 am #

REPLY ↗

You might need to change the way the date-time is parsed in your data file.

Shashank May 25, 2020 at 7:31 am #

Start Machine Learning

How did you use the ACF plot to determine the order of p at the beginning ? Do you not use a PACF plot to determine the order of p and ACF to determine the order of q?



Jason Brownlee May 25, 2020 at 1:23 pm #

REPLY ↗

See this:

<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

Pranab May 26, 2020 at 12:52 pm #

Hi Jason,

Enjoyed this post, as usual. I believe, AIC and BIC are the value, the better it is). Are there any direct indicate statsmodels or any other library?

Thanks



Jason Brownlee May 26, 2020 at 1:23 pm #

DEDI V ↗

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees.**

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Thanks.

Good question, I don't know for sure, sorry. Perhaps check the API.

Mary Grace Barbecho June 6, 2020 at 1:43 am #

REPLY ↗

Hi. I'm new to python and I got this error msg.

```
C:\Users\HP\Anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:165: ValueWarning: No frequency information was provided, so inferred frequency MS will be used.  
% freq, ValueWarning)
```

What does this mean?



Jason Brownlee June 6, 2020 at 7:56 am #

REPLY ↗

Looks safe to ignore, or you can try posting/searching on stackoverflow.

Marco Alexis June 11, 2020 at 4:39 pm #

Start Machine Learning

Hello Jason,

Thank you for helping a lot of people with your skill. Me in particular learned a lot from you.

I want to ask Jason if you recommend to let ARIMA do the differencing or do the differencing before fitting ARIMA? Does it make a significant difference if I do one over the other or vice versa?

Thank you very much and please continue helping people.



Jason Brownlee June 12, 2020 at 6:08 am #

REPLY ↗

Yes, let ARIMA do the differencing.

Rafi June 17, 2020 at 11:30 pm #

Sir, Is ARIMA, CNN-LSTM, SVM, Random Forest good for forecasting?



Jason Brownlee June 18, 2020 at 6:25 am #

REPLY ↗

They can be for complex datasets, try them and see:

https://machinelearningmastery.com/start-here/#deep_learning_time_series

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Owolabi June 20, 2020 at 5:42 pm #

REPLY ↗

i'm always hungry of reading from your site. you rock!



Jason Brownlee June 21, 2020 at 6:19 am #

REPLY ↗

Thanks.

macro_finance June 24, 2020 at 7:39 am #

REPLY ↗

Hi,

In what package can I find the "mean_squared_error" function?

Thanks,

MF

Start Machine Learning



Jason Brownlee June 24, 2020 at 1:25 pm #

REPLY ↗

The scikit-learn library:

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html

Kevin July 8, 2020 at 12:14 am #

REPLY ↗

Hi there,

I have been stuck for days because I cannot find a way giving multiple time series as input.

I have got thousand trajectories positions of different vehicles. Each sample (trajectory) is composed of 10 positions. (total number of samples).

I want to fit the model with all samples for x coordinates of trajectory that I give in input.

Then I plan to do the same with another model for y coordinates.

I found someone on github who has got the same issue:
<https://github.com/statsmodels/statsmodels/issues/421>

Can someone help me please?

Thank you in advance for your support!

Best regards,

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee July 8, 2020 at 6:32 am #

REPLY ↗

Perhaps use an sklearn model and this function to prepare the data:

<https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/>

Kevin July 8, 2020 at 5:22 pm #

REPLY ↗

Thank you for your answer Jason!

What I do not figure out is that one have to give sklearn models a sequence of observations as a list or a numpy array. But, in my problem, I do have multiple parts of trajectories from different vehicles. If I create a list with all trajectories of all vehicles, the data will be completely mixed up and the model will not be able to fit these data.

What I want to demonstrate is that a number of time series can be (or cannot be) modeled with a single model for my problem (i.e. the same parameters) by fitting them all simultaneously.

Start Machine Learning

Do you think it is possible to resolve the problem I am describing for multiple trajectories (having different length) from multiple vehicles using traditional regression model?

Do you think I have to assign each object a different variable, take trajectories of same length in a fixed window of observations (for example the 10 last values) and give these data to a sklearn model?

Thank you again



Jason Brownlee July 9, 2020 at 6:37 am #

REPLY ↗

Perhaps start with a strong idea of what you want to learn/predict.

E.g. single entity, across entities, etc. Then

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Kevin Burgisser July 9, 2020 at 6:37 am #

I want to forecast the 10th position based on the 40 last positions for all vehicles. But I have a dataset (thousands of trajectories, 40 positions) in statsmodels...

Consider I am only working with the x coordinate (one model for x coordinate and one for y coordinate).



Jason Brownlee July 10, 2020 at 5:53 am #

If I understand correctly, the code I linked previous will do this for you:

<https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/>

Zeina August 7, 2020 at 7:09 pm #

May I ask you Kevin if you found the exact solution to your problem? I am also having the very same issue. Thanks a lot

Keke July 12, 2020 at 7:45 pm #

REPLY ↗

What is the tuple returned by the `model_fit.forecast()` method? Statsmodels documentation is really poor... They just say it is the predicted values but does it return the error values? Or does the method return different values?

Start Machine Learning



Jason Brownlee July 13, 2020 at 5:59 am #

REPLY ↩

Contains predictions for each step and the prediction interval for each step (upper and lower).

Rahul July 14, 2020 at 3:16 am #

REPLY ↩

I am working on Project of prediction and forecasting of pharmacy store sales based on past sales data. Its more likely to forecast future sales , so does ARIMA MODEL is suitable for such according to you ? Whats your view on this? How can I work on this? Any



Jason Brownlee July 14, 2020 at 6:29 am #

I recommend testing a suite of different models to see which works best for your dataset.

Joe July 22, 2020 at 9:25 pm #

Thx Jason!

If I want to predict the recovery of the air travel industry for the following 12 months by predicting the flight passenger data.

May I know how to make the time series with abrupt changes(Covid-19) stationary before using ARIMA? (as first-order differencing isn't working)

What other time series model will you suggest if it remains non-stationary?

Thanks!



Jason Brownlee July 23, 2020 at 6:06 am #

REPLY ↩

Not sure that case is predictable. But what do I know.

You might want to explore nonlinear models that also take multiple input variables:
https://machinelearningmastery.com/start-here/#deep_learning_time_series

Aniket July 25, 2020 at 8:24 pm #

Start Machine Learning

I am working on project to predict ground water level on post ground water level dataset, does Arima model is suitable or there is another model more suitable than Arima model for ground water level ? any link or suggestion??

thanks!!

Aniket July 25, 2020 at 8:26 pm #

REPLY ↩

past ground water level dataset****



Jason Brownlee July 26, 2020 at 6:17 am #

I recommend testing a suite of techniques

Perhaps this will help:

<https://machinelearningmastery.com/time-series-forecasting-with-python/>

Zeina August 7, 2020 at 7:32 pm #

Hello Jason,

Thanks a lot for your article, it's very informative.

Do you know how can I have 2 dimensional input instead of one dimensional?

I have x,y coordinates of a robot and I would like to predict its position after K steps.

Thanks so much

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee August 8, 2020 at 5:59 am #

REPLY ↩

You're welcome.

Yes, you will need to use a different model:

<https://machinelearningmastery.com/time-series-forecasting-methods-in-python-cheat-sheet/>

Zeina August 9, 2020 at 11:11 am #

REPLY ↩

Thanks a lot for your answer.

Can you please refer me to the model that could take a 2 dimensional input (x,y) into account?

Start Machine Learning



Jason Brownlee August 10, 2020 at 5:44 am #

REPLY ↗

The VAR/VARIMA etc linear models are a good start.

cmei September 7, 2020 at 9:32 pm #

REPLY ↗

hi Jason,

Just as a feedback. Mainly because you are able to answer challenger questions on the comment section, you are gaining my full trust. Not sure if there are many tested sources like yours. so pls. keep-up good work!

I hope you provide some Vector autoregression (VAR) ..



Jason Brownlee September 8, 2020 at 6:50 am #

Thanks!

Great suggestion, the example here will get you started:
<https://machinelearningmastery.com/time-series-forecasting-with-python/>

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

cmei September 9, 2020 at 5:36 pm #

REPLY ↗

Yes. I have already benefited greatly from the cheat sheet for VAR and VARMA. there are some details that require deeper understanding particularly tho. particularly in selecting features (as exog or endog, and select correlated ones.) and also correctly forecasting (with the right transformation,etc.)



Jason Brownlee September 10, 2020 at 6:24 am #

REPLY ↗

Great!

jbdatal31 September 15, 2020 at 6:29 pm #

REPLY ↗

Hi and thanks for sharing so many tutorials,

In the dev there's something I don't understand.

obs = test[t]

history.append(obs)

Start Machine Learning

Using real values to feed the model is weird. Predict known values is a test-case, but in a use-case you don't have the value to predict.

For me the dev sould be:

```
history.append(yhat)
```

Then the plot becomes less attractive.



Jason Brownlee September 16, 2020 at 6:22 am #

REPLY ↗

This is called walk forward validation and is the standard way for evaluating time series forecasting models.

You can learn more here:

<https://machinelearningmastery.com/backtest-machine-learning-with-python/>

Fran November 4, 2020 at 10:49 pm #

Hi Jason,

I am interested in exploring the association between 2 vaccination rates).

Start Machine Learning

×

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

I am looking at survey data collected on a yearly basis and I have data available for ~17 time points. I am considering ARIMAX to explore my data but I am not sure whether it is the most appropriate tool. A lot of studies using ARIMAX seem to have more granular data (monthly or quarterly). I also read you would need ~50 observations for aggregated data and this is not my case... I am new to this though so I would welcome some expert opinion.



Jason Brownlee November 5, 2020 at 6:35 am #

REPLY ↗

Perhaps try it and compare results to a naive persistence model.

Mahendra Choudhary November 6, 2020 at 12:23 am #

REPLY ↗

I think there is a fault in your code,
why are you using test observation, in below script

```
for t in range(len(test)):  
    model = ARIMA(history, order=(5,1,0))  
    model_fit = model.fit(disp=0)  
    output = model_fit.forecast()  
    yhat = output[0]
```

Start Machine Learning

```

predictions.append(yhat)
obs = test[t]
history.append(obs)
print('predicted=%f, expected=%f' % (yhat, obs))
error = mean_squared_error(test, predictions)

```

if i dont have test dataset and i want to predict the train dataset only, then i should use
obs = predictions[t] and not test[t]..



Jason Brownlee November 6, 2020 at 5:57 am #

REPLY ↗

It is called walk-forward validation and is used for training and validating forecasting models.

You can learn more about it here:

<https://machinelearningmastery.com/backtest-machine-learning-models-walk-forward-validation/>

Mahendra November 6, 2020 at 5:59 pm #

Thanks for your prompt response,
Jason, i only have 1000 data points(in my train set) and 10000 (in my test set)
10000 values. How can i use your code for the same?

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee November 7, 2020 at 6:26 am #

REPLY ↗

You can fit your model on your data, then call `forecast()` and specify the number of steps to forecast.

Leave a Reply

Start Machine Learning

Name (required)

Email (will not be published) (required)

Website

SUBMIT COMMENT

**Welcome!**

I'm Jason Brownlee PhD
and I **help developers** get results with machine learning.
[Read more](#)

Never miss a tutorial:**Picked for you:**[How to Create an ARIMA Model for Time Series Forecasting in Python](#)[How to Convert a Time Series to a Supervised Learning Problem in Python](#)[11 Classical Time Series Forecasting Methods in Python \(Cheat Sheet\)](#)[Time Series Forecasting as Supervised Learning](#)[How To Backtest Machine Learning Models for Time Series Forecasting](#)**Start Machine Learning**

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

 Email Address

 START MY EMAIL COURSE

The [Time Series with Python](#) EBook is
where you'll find the **Really Good** stuff.

[>> SEE WHAT'S INSIDE](#)

© 2020 Machine Learning Mastery Pty. Ltd. All Rights Reserved.

Address: PO Box 206, Vermont Victoria 3133, Australia. | ACN: 626 223 336.

[LinkedIn](#) | [Twitter](#) | [Facebook](#) | [Newsletter](#) | [RSS](#)

[Privacy](#) | [Disclaimer](#) | [Terms](#) | [Contact](#) | [Sitemap](#) | [Search](#)

Start Machine Learning ×

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

[Start Machine Learning](#)