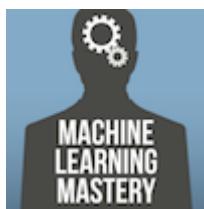


 Navigation

Want help with time series? [Take the FREE Mini-Course](#)



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](#)[G+](#)

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 data with 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.

Let's get started.

Autoregressive Integrated Moving Average Model

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

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

[Your Start in Machine Learning](#)

- **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 between an observation and a residual error from a moving average model applied to lagged observations.

Each of these components are explicitly specified in the model as a parameter. A standard notation is used of ARIMA(p,d,q) where the parameters are substituted with integer values to quickly indicate 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.
- **d:** The number of times that the raw observations have been differenced.
- **q:** The size of the moving average window, also called the order of the moving average model.

A linear regression model is constructed including the raw observations prepared by a degree of differencing in order to make the time series stationary. This removes structures that negatively affect the regression model.

A value of 0 can be used for a parameter, which indicates that the ARIMA model can be configured to perform the function of an AR model 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.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

Stop learning Time Series Forecasting the *slow way!*

Take my free 7-day email course and discover data prep, modeling and more (with sample code).

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

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

Your Start in Machine Learning

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).

Learn more about the dataset and download it from [here](#).

Download the dataset and place it in your current working directory with the filename “shampoo-sales.csv”.

Below is an example of loading the Shampoo Sales dataset with Pandas with a custom function to parse the date-time field. The dataset is baselined in an arbit

```
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('190'+x, '%Y-%m')
7
8 series = read_csv('shampoo-sales.csv', header=0)
9 print(series.head())
10 series.plot()
11 pyplot.show()
```

Running the example prints the first 5 rows of the data

```
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
```

Your Start in Machine Learning

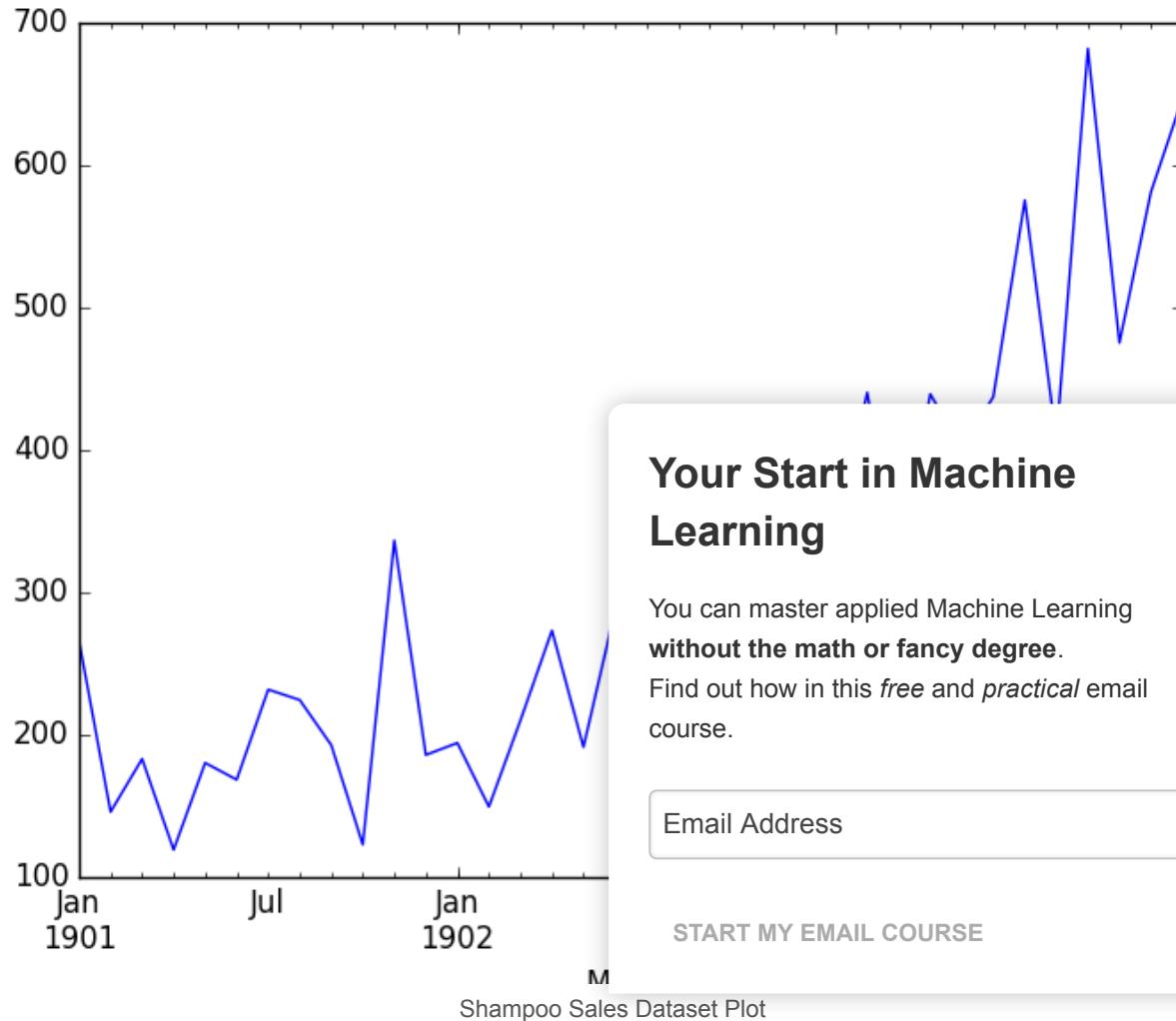
You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning



Your Start in Machine Learning

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

[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.

```

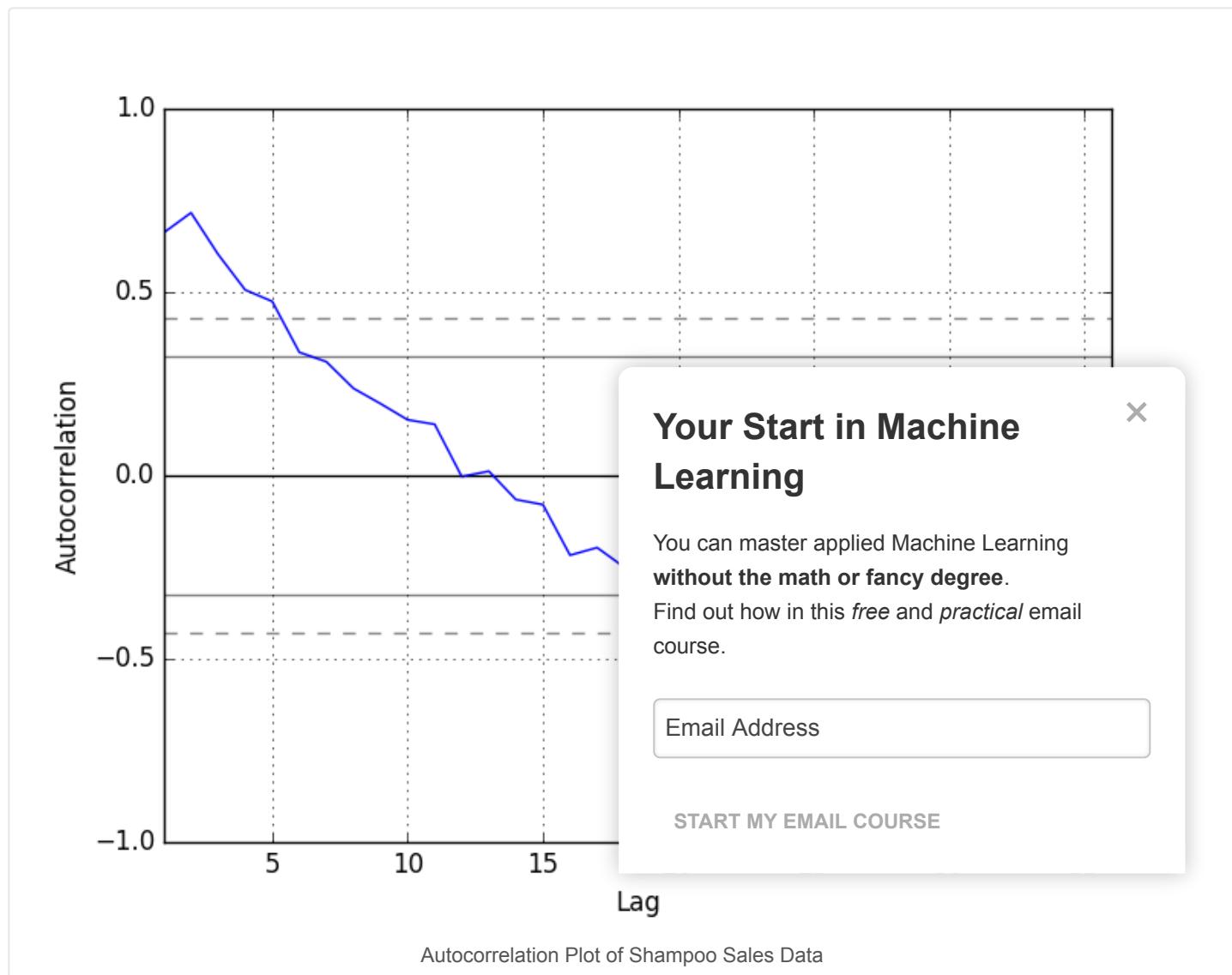
1 from pandas import read_csv
2 from pandas import datetime
3 from matplotlib import pyplot
4 from pandas.tools.plotting import autocorrelation_plot
5
6 def parser(x):
7     return datetime.strptime('190'+x, '%Y-%m')
8
9 series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True,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

Your Start in Machine Learning

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



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 no seasonal component (0).

Your Start in 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, date_parser=parser)
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. Observations:	417.227		
2	Model:	ARIMA(5, 1, 0)	Log Likelihood:	410.098		
3	Method:	css-mle	S.D. Error:			
4	Date:	Mon, 12 Dec 2016	AIC:			
5	Time:	11:09:13	BIC:			
6	Sample:	02-01-1901	HQIC:			
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			Roots			
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.

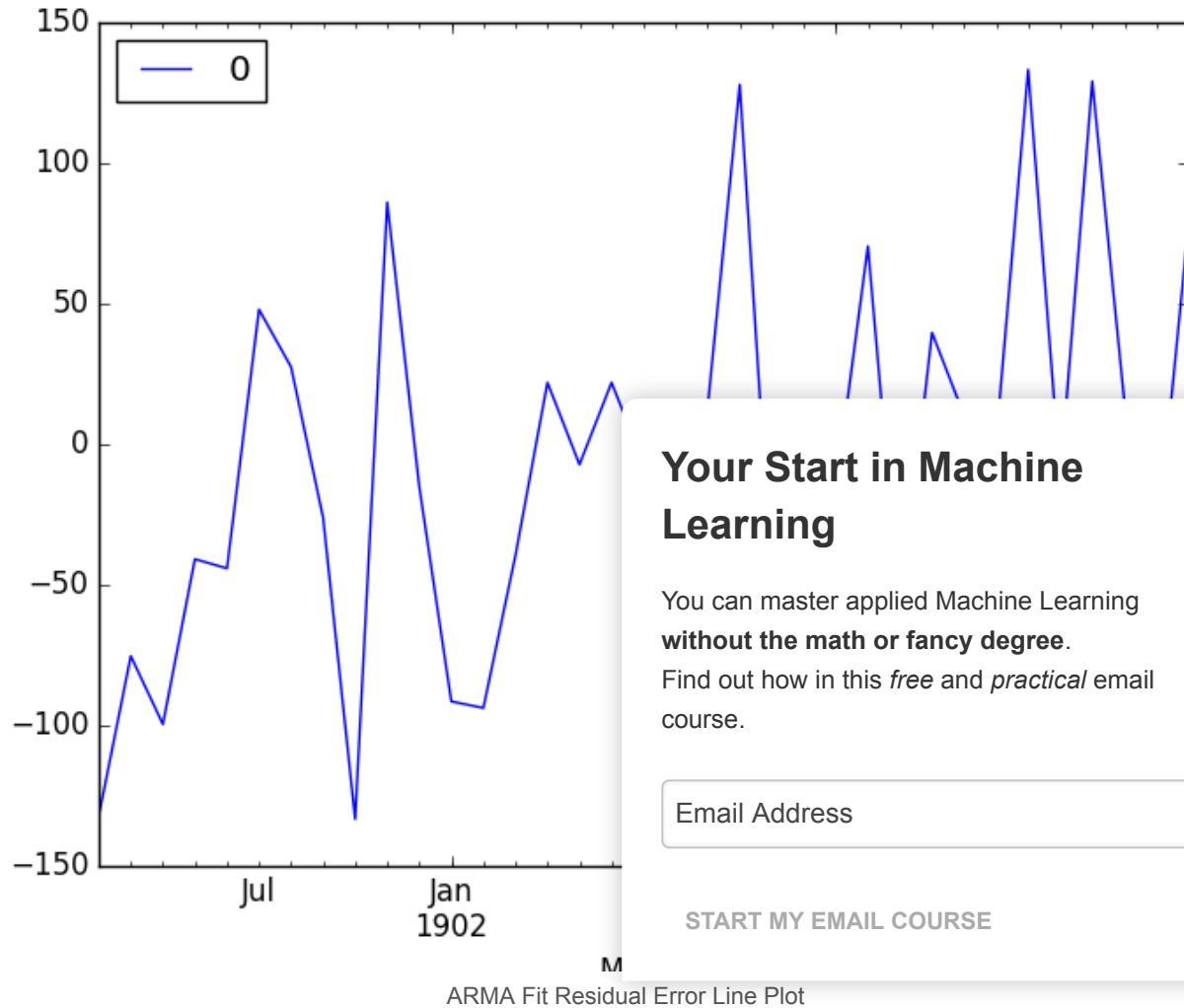
Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

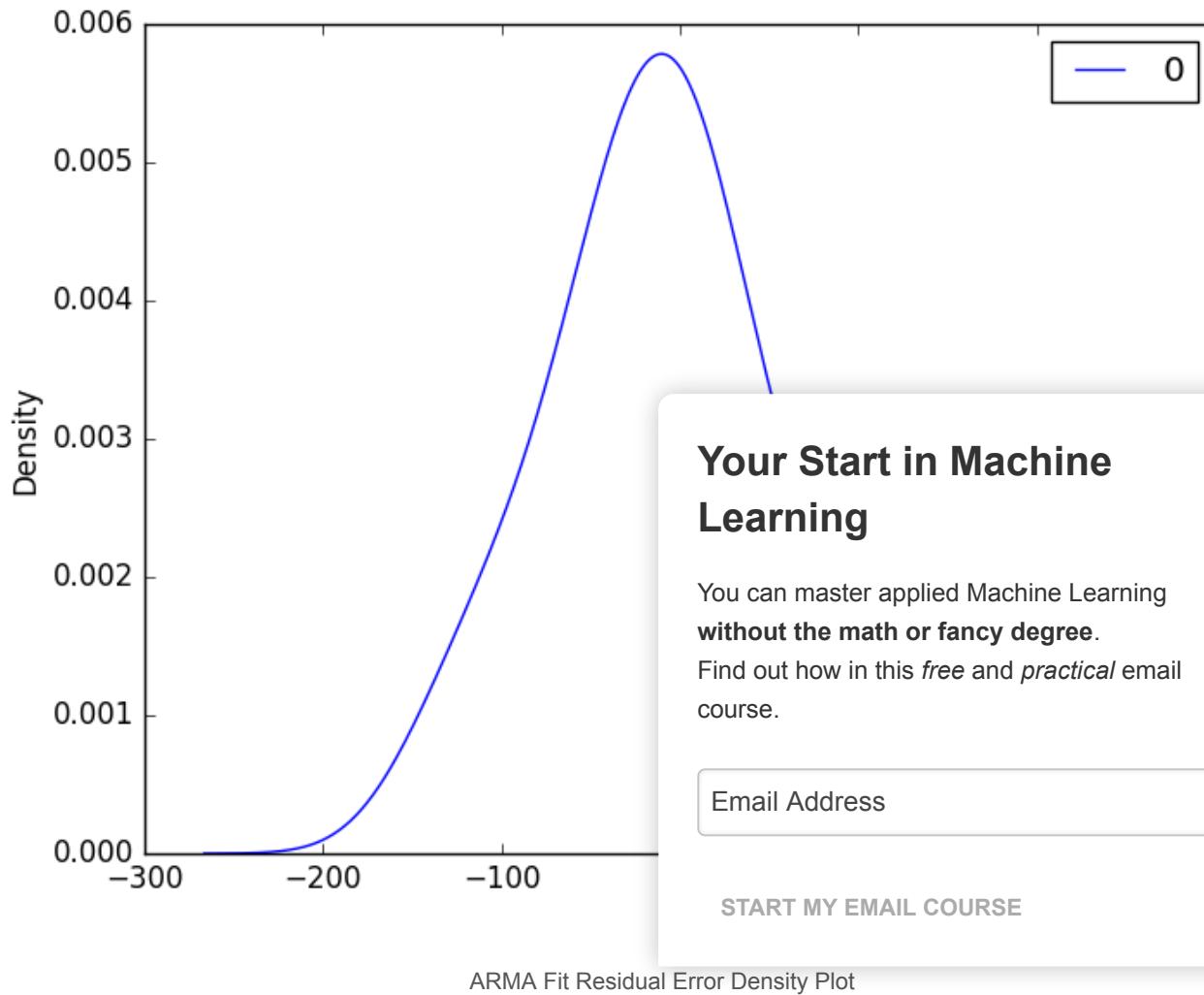
Email Address

START MY EMAIL COURSE



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

Your Start in Machine Learning



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 steps

Your Start in 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 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.

```

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 = []
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

```

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning

```
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.

Your Start in Machine Learning

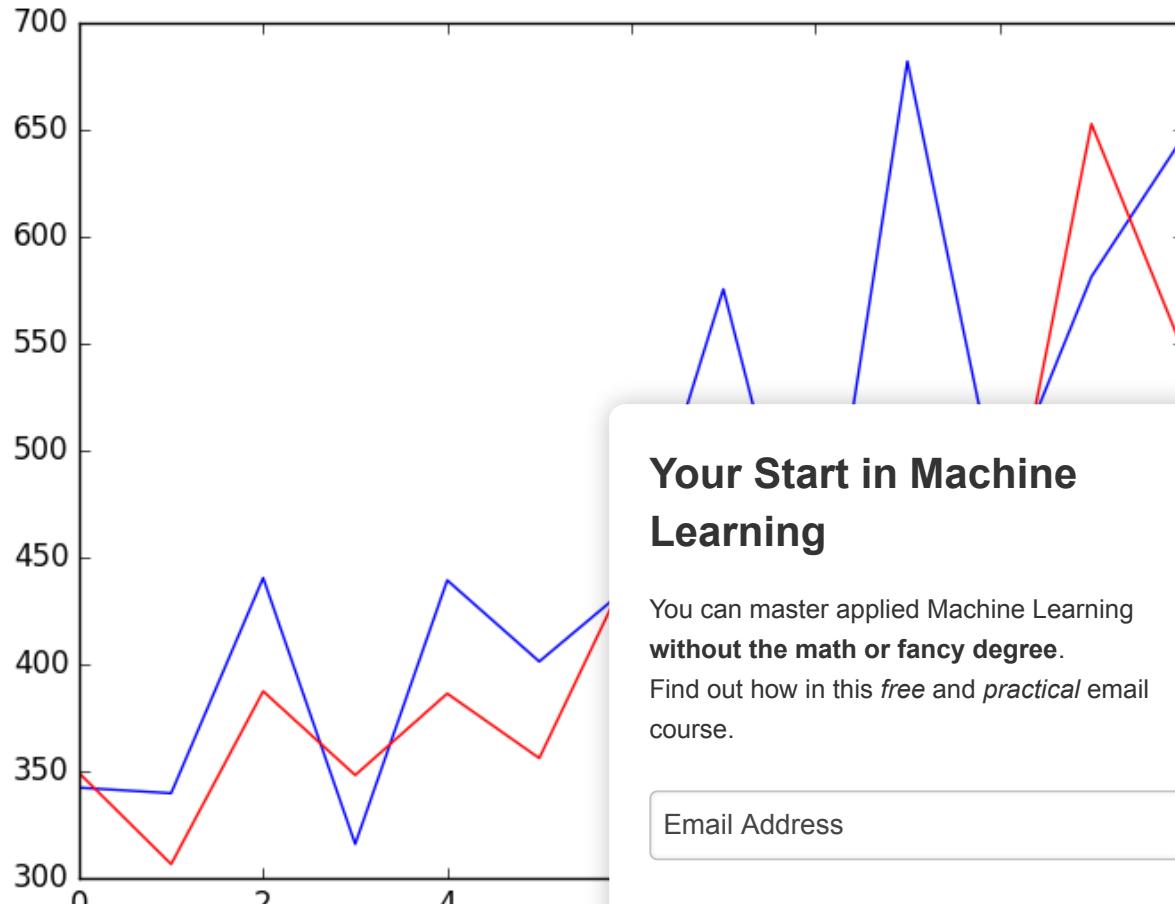
You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning



Your Start in Machine Learning

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

Your Start in 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.

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 used.
- How to perform a quick time series analysis using the statsmodels library.
- How to use an ARIMA model to forecast out of sample data.

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 them.

Your Start in Machine Learning

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

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

Email Address

[START MY EMAIL COURSE](#)

Want to Develop Time Series Models in Python?

Develop Your Own Forecasts in Minutes

...with just a few lines of python code

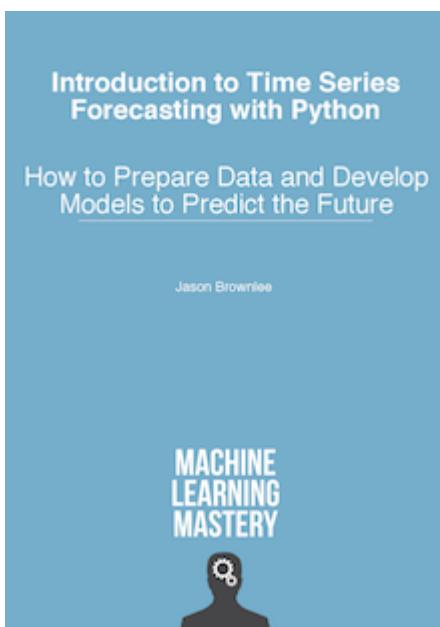
Discover how in my new Ebook:
[Introduction to Time Series Forecasting With Python](#)

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.

[Click to learn more.](#)



[Tweet](#)

[Share](#)

[Share](#)

[G+](#)

[Your Start in Machine Learning](#)



About Jason Brownlee

Jason Brownlee, Ph.D. 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 Residual Errors to Correct Time Series Forecasts with Python >

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



SalemAmeen January 9, 2017 at 7:22 am #

Many thank



Jason Brownlee January 9, 2017 at 7:53 am

You're welcome.



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?



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.

Your Start in Machine Learning

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 with a lot of capacity. You may also require a vast amount of data to learn this problem.



Yash July 25, 2018 at 8:01 pm #

X

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



Jason Brownlee July 26, 2018 at 7:30 pm #

If you have parallel input time series, then you can predict all variables. If you want to predict all variables, then you can use a neural network.

If you want to support multiple series generally, then you can use a recurrent neural network as a start:

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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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?

Your Start in 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/core/tools/parser.py", line 2263, in convert_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: strptime() argument 1 must be str, not number

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/core/tools/parser.py", line 2263, in convert_date_parser(*date_cols), errors='ignore')
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
    series = read_csv('shampoo-sales.csv', header=0, parse_dates=[0], index_col=0, squeeze=True,
                      date_parser=parser)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 562, in parser_f
    return _read(filepath_or_buffer, kwds)
File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 525, in _read
    return parser.read()
```

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

```

File "/Users/kevinoost/anaconda/lib/python3.5/site-packages/pandas/io/parsers.py", line 815, in read
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.p
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 th
Process finished with exit code 1
Help would be much appreciated.

```

Your Start in Machine Learning

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

Find out how in this *free* and *practical* email 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

```

Your Start in 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
```

Your Start in Machine Learning

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



AI January 21, 2018 at 8:56 pm #

REPLY ↗

Your Start in 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



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

Let say I have a time series data with many attributes (e.g. time, date, temperature, tire_pressure), how could we made a model out of this? 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.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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.



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?

Your Start in Machine Learning

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.arima_model.ARIMA()` function documentation says it takes the optional parameter `exog`, which is described in the documentation as 'an optional array of exogenous variables'. This sounds like multivariate analysis to me, would you agree?

I am trying to predict number of cases of a mosquito-borne disease, over time, given weather data. So I believe the ARIMA model should

Thank you!



Jason Brownlee June 2, 2017

I have not experimented with



Jason Brownlee June 2, 2017 at 12:45 pm #

No multivariate examples at this stage.

Yes, any supervised learning method.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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



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.

Your Start in Machine Learning



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 #

REPLY ↗

It is so informative..thankyou



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

X

I'm glad to hear that Muhammad.



Sebastian January 31, 2017 at 3:33 am #

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Great post Jason!

I have a couple of questions:

- Just to be sure. `model_fit.forecast()` is single 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 steps to be very accurate, if that is your question?



Sebastian February 3, 2017 at 9:25 am #

REPLY ↗

Thanks for the reply!

Your Start in Machine Learning

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 10:22 am #

REPLY ↗

Hi Sebastian, I see.

In the case of predicting on the training dataset. the model has access to real observations. For example, if you predict the next 5 obs some predict t+5 rather than prediction(t+4).

In the case of predicting beyond the end of predictions (unless you provide them), it does steps. The result is the errors compound a

Does that make sense/help?



Sebastian February 3, 2017 at 6:

That helped!

Thanks!

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

Glad to hear it Sebastian.



satya May 22, 2017 at 9:19 pm #

Hi Jason,

suppose my training set is 1949 to 1961. Can I get the data for 1970 with using Forecast or Predict function

Thanks
Satya

Your Start in Machine Learning



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 you forecast using something like ARIMA or AR or GARCH in Python? I'm guessing more forecasting mechanism?

To give you a sense of my data, given terms of number of predictions can we jaggged predictions in-sample, but esse semblance of that behavior and instead think. Thanks!



Jason Brownlee July 6, 2018

The skill of AR+GARH (or either) parameters and on the specifics of the .

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.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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?



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

REPLY ↗

Hi Elliot,

Your Start in Machine Learning

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 #

REPLY ↗

I tried the `model.forecast()` at the end of the program.

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

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 in

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee February 2, 2017

Hi Elliot, the `forecast()` function is about it here:

<http://statsmodels.sourceforge.net/stable/generated/statsmodels.tsa.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.

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

Your Start in Machine Learning



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 #

REPLY ↗

Hey Jason,

thank you very much for the post, very good written! I have a question about the rolling forecast. In your code, you change history.append(obs) to history.append(yhat), and I don't understand the reason? and how do you actually do the out-of-sample forecast? Thank you very much!

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

Hi james,

Each loop in the rolling forecast shows you how to

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?

Your Start in Machine Learning

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

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

[START MY EMAIL COURSE](#)
**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?

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

REPLY ↗

Hi James,

Your Start in Machine Learning

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 need to learn a lot from people like you! 😊 nice day!



Jason Brownlee February 17,

I'm here to help James!



Supriya February 16, 2017 at 1:27 am #

when i calculate train and test error , train rms



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

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



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

REPLY ↩

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

Your Start in Machine Learning

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 #

REPLY ↗

I get it. Thanks Jason.

I was thinking, in this particular example, will the prediction change if we keep adding data?



Jason Brownlee February 21, 2017

Great question Matias.

The amount of history is one variable to test.

Design experiments to test if having more



Ubald Kuijpers February 24, 2017 at 10:05 pm #

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 .

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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 be used to detect anomalies in time series. thanks.



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

REPLY ↗

Great suggestion, thanks Jacques.

Your Start in Machine Learning



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 #

REPLY ↗

Are you sure?

See this list of abbreviations:

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

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".

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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 #

REPLY ↗

Glad to hear it Niirkshith.

Try and see.

Your Start in Machine Learning



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)); #ar, diff and ma are the model params
2 model_fit = model.fit()
3 residuals = pd.DataFrame(model_fit.resid)
4 #calculate the std dev of the residuals, we use numpy's std dev function
5 print("the std dev of the residuals = %f" % np.std(residuals))
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 of the fitted model and the std deviation of the residuals of the rolling forecast will



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

No, they are not the same.

See this post on performance measures:

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

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 $y(t) - y(t-12)$ 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 not yearly



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

Your Start in Machine Learning

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



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

You can use ACF and PACF plots to help

See this post:

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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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.



Niirkshith May 12, 2017 at 1:02 pm #

REPLY ↗

Hi Jason,
any updates on the same

Your Start in Machine Learning



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 6:48 am #

REPLY ↗

Sounds like the data is not a good fit for ARIMA. Try fitting an MA model or something else.



David March 30, 2017 at 8:53 am #

Hi Jason,

Thanks for the great post! It was very helpful. I'm currently trying to fit an ARIMA model with order (4, 1, 5) and I'm getting an error message "The model does not appear invertible". This is because the AR parameters should induce invertibility, choose a different model order. I've tried fitting an MA(5) model which works when fitting, but seems to error out when forecasting. I've also tried fitting an AR(4) model which works well when using your parameters of (0, 1, 5) and it errors out when using your suggested parameters. Any ideas on the cause/fix for the error? Any tips would be much appreciated.

Your Start in Machine Learning

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



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.

Your Start in 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 airline routes on daily basis. Now I want to predict for each airline number. How can I fit these time series models. Separate model

Your Start in Machine Learning

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

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

Your Start in Machine Learning



Kashif May 25, 2017 at 6:30 pm #

REPLY ↗

Hi Sir, Do you have tutorial about vector auto regression model (for multi-variate time series forecasting?)



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

REPLY ↗

Not at the moment.



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

Thanks a lot, Dr. Jason. This tutorial explained a lot. I am trying to implement it 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 understand what is wrong?

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

Perhaps consider rescaling your input data and explore other configurations?



Alex June 9, 2017 at 8:01 am #

REPLY ↗

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.

Your Start in Machine Learning

For linear models, you could use vector autoregressions (VAR).

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

I hope that helps as a start.



Donato June 13, 2017 at 10:23 pm #

REPLY ↗

Hi Jeson, it's possible to training the ARIMA with more files? Thanks!



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

Do you mean multiple series?

See VAR:

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



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

"First, we get a line plot of the residual errors, information not captured by the model."

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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

No, the upward trend that appears to exist in the plot of residuals.



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])]
```

Your Start in Machine Learning



Ukesh June 24, 2017 at 12:53 am #

REPLY ↗

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 extreme

Your Start in Machine Learning

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

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

START MY EMAIL COURSE



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

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.

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 ↗

Your Start in Machine Learning

I would focus on the approach that gives the best results on your problem and is robust. Don't get caught up on "proper".



Kashif July 12, 2017 at 11:29 pm #

REPLY ↗

Dear Respected Sir, I have tried to use ARIMA model for my dataset, some samples of my dataset 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

Your Start in Machine Learning

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

Find out how in this *free* and *practical* email 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,

Your Start in Machine Learning

```
def parser(x):
    return datetime.strptime('190'+x, '%Y-%m')
series = read_csv('/home/administrator/Downloads/shampoo.csv', header=0, parse_dates=[0], 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 three year period' does not match format '%Y-%m'

Please help.

Thanks



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

Check that you have deleted the footer in



Kushal July 14, 2017 at 6:53 pm #

Hi Jason

Does ARIMA have any limitations for size of the sample doesn't complete.

Thanks

Kushal

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

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)

Your Start in Machine Learning



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

REPLY ↗

Thanks Olivia, I hope to cover HMMs in the future.



Ben July 19, 2017 at 11:27 am #

REPLY ↗

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 was wondering if there is a function to do this?

Is there a function to do this?

Thank you

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

START MY EMAIL COURSE



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

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.



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 it 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>

Your Start in Machine Learning



Shaun July 27, 2017 at 9:29 am #

REPLY ↗

Hi Jason,

In your example, you append the real data set to the history list- aren't you supposed to append the prediction?

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

in a real example, you don't have access to the real "future" data. if you were to continue your example with dates beyond the data given in the csv, the results are poor. Can you elaborate?



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

We are doing walk-forward validation.

In this case, we are assuming that the real ob is m
the next prediction is required.



Jai July 31, 2017 at 3:59 pm #

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 ARIMA
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/>

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in 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_model, not before
from matplotlib import pyplot
```

REPLY ↗



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

REPLY ↗

Glad to hear it.



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

It looks like statsmodels was not installed environment.

You installed using pip3, are you running a python:

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jai August 1, 2017 at 4:18 pm #

interestingly, under your Rolling Forecasting section, you have statsmodels.

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

i am using jupyter notebook from WinPython-64bit-3.5.3.1Qt5 to run your examples. i keep getting ImportError: No module named 'statsmodels' if i declare import this way in ARIMA with Python 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
```

Your Start in Machine Learning

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

thanks for the replies. case close



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

Glad to hear it.



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

You will need to install statsmode

REPLY ↗



Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Fathi July 31, 2017 at 5:44 pm #

Great explanation

can anyone help me to write code in R about forecasting next 3 hour, kindly help me to write code using R with thanks in advance



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

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/>

REPLY ↗



Fathi August 9, 2017 at 10:49 pm #

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

Your Start in Machine Learning



Quentin August 11, 2017 at 10:37 pm #

REPLY ↗

Can the ACF be shown using bars so you can look to see where it drops off when estimating order of MA model? Or have you done a tutorial on interpreting ACF/PACF plots please elsewhere?



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 #

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-yea
# fit model
model = ARIMA(series, order=(0, 0, 0))
model_fit = model.fit(disp=0)
print(model_fit.summary())
# plot residual errors
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\datetimes.py", line 444, in _convert_listlike
values, tz = tslib.datetime_to_datetime64(arg)
File "pandas\_libs\tslib.pyx", line 1810, in pandas._libs.
```

Your Start in Machine Learning

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

Email Address

START MY EMAIL COURSE

```
(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\tsa_model.py", line 56, in __init__
dates = to_datetime(dates)
File "C:\Python36\lib\site-packages\pandas\core\tools\datetimes.py", line 514, in to_datetime
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.
File "pandas\_libs\tslib.pyx", line 2538, in pandas._libs.
File "pandas\_libs\tslib.pyx", line 2506, in pandas._libs.
File "pandas\_libs\tslib.pyx", line 2500, in pandas._libs.
File "pandas\_libs\tslib.pyx", line 1517, in pandas._libs.
File "pandas\_libs\tslib.pyx", line 1774, in pandas._libs.
pandas._libs.tslib.OutOfBoundsDatetime: Out of bound
```

During handling of the above exception, another exception occurred:

Traceback (most recent call last):

```
File "C:/Users/aamrit/Desktop/untitled1/am.py", line 9,
model = ARIMA(series, order=(0, 0, 0))
File "C:\Python36\lib\site-packages\statsmodels\tsa\arima\processes.py", line 102, in process
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=missing)
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__
raise ValueError("Given a pandas object and the index does not contain dates")
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 #

REPLY ↗

Your Start in Machine Learning

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
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 579, in fit_start_params
start_params = self._fit_start_params_hr(order, start_a
File "C:\Python36\lib\site-packages\statsmodels\tsa\ari
endog -= np.dot(exog, ols_params).squeeze()
TypeError: Cannot cast ufunc subtract output from dtype
'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, pa
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())
```

Your Start in Machine Learning

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

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

[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 #

[REPLY ↗](#)

Your Start in Machine Learning

Great tip.



Vicente Queiroz March 30, 2018 at 8:43 pm #

REPLY ↗

@kyci is correct as you can check in

<https://github.com/statsmodels/statsmodels/issues/3504>.

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 #

How can I add multiple EXOG variables in



Amritanshu August 29, 2017 at 8:00 pm #

Jason, I am able to implement the model but I
how to find the exact values for p,d and q ?

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

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 differencing methods ?



Jason Brownlee August 31, 2017 at 6:18 am #

REPLY ↗

It depends.

The d only does a 1-step difference. You may still want to perform a seasonal difference.

Your Start in Machine Learning



Amritanshu August 31, 2017 at 5:21 pm #

REPLY ↗

Jason, Can I get a link to understand it in a better way ? I am a bit confused on this.



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

REPLY ↗

You can get started with time series here:

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



Amritanshu September 5, 2017 at 11:22 pm #

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 1st August.

My Date format in excel file for the model is 4/22/17 –M

```
output = model_fit.predict(start='2017-01-08',end='2017-01-09')
```

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\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\arima_model.py", line 375, in _validate
    start = _index_date(start, dates)
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 ?

Your Start in Machine Learning

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

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

START MY EMAIL COURSE



Jason Brownlee September 7, 2017 at 12:45

REPLY ↗

Your Start in Machine Learning

Sorry, I'm not sure about the cause of this error. Perhaps try predicting one day and go from there?



Amritanshu September 20, 2017 at 10:20 pm #

REPLY ↗

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', he
squeeze=True, date_parser=parser)
print(series.head())
series.plot()
pyplot.show()
```

ERROR is

```
runfile('C:/Users/kashi/Desktop/prog/Date_time.py', wdir='C:/Users/kashi/Desktop/prog')
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 562, in parser_f
return _read(filepath_or_buffer, kwds)
```

```
File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 325, in _read
return parser.read()
```

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 815, in read
ret = self._engine.read(nrows)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1387, in read
index, names = self._make_index(data, alldata, names)

File "C:\Users\kashi\Anaconda3\lib\site-packages\pandas\io\parsers.py", line 1030, in _make_index
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\parsers.py", line 300, in parse
results[i] = parse_func(*args)

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

File "C:\Users\kashi\Anaconda3\lib_strptime.py", line 5
tt, fraction = _strptime(data_string, format)

File "C:\Users\kashi\Anaconda3\lib_strptime.py", line 3
(data_string, format))

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

I have already removed the footer note from the dataset. I can't remove this error. But when I comment "date_parser=p.parse_date" How to resolve it?

Thanks

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

Find out how in this *free* and *practical* email 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 signature matching types dtype('<U32') dtype('<U32')
dtype('<U32')

I've tried '%Y-%b' but that only gives me the "does not match format" error

Any ideas?

Your Start in Machine Learning

/ Thanks



Jason Brownlee September 22, 2017 at 5:35 am #

REPLY ↗

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.



Alec September 22, 2017 at 5:41 pm #

Hey,

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

The header in the .csv was written as "Month," to "month", "sales" and it worked.

Thanks for putting in the effort to follow up on |



Jason Brownlee September 23, 2017 at 10:15 am #

Glad to hear that Alec!

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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, stores separately, and try models that combine the data. See what works best.

Shud October 23, 2017 at 7:47 pm #

Your Start in Machine Learning



Hey Jason,

You mentioned that since the residuals doesn't have mean = 0, there is a bias. I have same situation. But the spread of the residuals is in the order of 10^5. So i thought it is okay to have non-zero mean. Your thoughts please?



Shud October 23, 2017 at 8:20 pm #

REPLY ↗

Btw my mean is ~400



zhifeng November 4, 2017 at 1:17 am #

For those who came with an error of ValueError
%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

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

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning

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



cuongquyet November 10, 2017 at 9:59 pm #

Dear Jason,

Firstly, I would like to thanks about your sharing

Secondly, I have a small question about ARIMA with P
forecasted with ARIMA model. How Python supports t

For example, I have data of total orders in a country, a
So I need to forecast for each districts (about 700 distr

Thanks you so much

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



Jason Brownlee November 11, 2017 at 9:22 am #

REPLY ↗

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 number means in the result? thank you!



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

REPLY ↗

It may be the confidence interval:

<https://machinelearningmastery.com/time-series-forecasting-uncertainty-using-confidence-intervals-python/>

Your Start in Machine Learning



Chetan November 14, 2017 at 10:32 am #

REPLY ↗

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!



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

Hmm, it might not be the best tool. You may need to design a one-to-many mapping function for data points.



Monsoon November 18, 2017 at 3:23 am #

Hi Jason,

Your tutorial was really helpful to understand the concepts. I have small doubt regarding the steps you followed at the end.

```
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 the elements from the "test" and the forecasted value because in real forecasting we don't have future data to include in test, isn't it? Or is it that you're trying to explain something and I'm not getting it.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

2) Second doubt, aren't you suppose to perform "reverse difference" for that you have used first order differencing in the model?

Kindly, please clear my doubt

Note: I have also went through one of your other tutorial where you have forecasted the average daily 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

In this case, we are assuming the real observation is over time, for example, in a single case, but perhaps over days, weeks, months, etc.

The differencing and reverse differencing were per-



Somayeh November 28, 2017 at 12:39 am #

Hi Jason,

Recently I am working on time series prediction, but my problem is 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 is that I 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.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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>



Desmond December 7, 2017 at 6:04 pm #

REPLY ↗

Hey just a quick check with you regarding the prediction part. I need to do some forecast of future profit based on the data from past profit. Let's say I got the data for the past 3 years, and then I wanted to perform a forecast on the next 12 months in next year.

[Your Start in Machine Learning](#)

Thanks!



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 class above, my inputs are the past records for the past 12 months. I would like to actually plot out the forecasted graph is basically what I am trying to do. I have assumed that the graph that plotted out is mea



Jason Brownlee December 10, 2017 at 1:15 pm #

I don't follow, sorry. You can plot a



Desmond December 10, 2017 at 1:15 pm #

Sorry but what does the expected and predicted means actually?

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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 13 points (start from 0 to 12) when each year only have 12 months? Looking forward to hear from you soon and thanks!



Jason Brownlee December 11, 2017 at 5:02 pm #

Your Start in Machine Learning

I arbitrarily chose to make predictions for 33% of the data which turned out to be 13 months.

You're right, it would have been clearer if I only predicted the final year.



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 4:13 pm #

Sorry, I cannot prepare a code snippet for you exactly what to do.



Desmond December 11, 2017 at 4:17 pm #

Hey Jason, thanks so much for your reply.
The model above aims to make a prediction on what you already have or trying to forecast 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 just a quick check with you again, let's say I have some zero value for the profit, will it break the forecast function? Or the forecast function must take in all non-zero value. Because sometimes I am getting "numpy.linalg.LinAlgError: SVD did not converge" error message and I'm not sure if it is the zero values that is causing the problem. 😊

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

Your Start in Machine Learning



Good question, it might depend on the model.

Perhaps spot check some values and see how the model behaves?



Desmond December 11, 2017 at 8:33 pm #

REPLY ↗

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 10:22 pm #

Hi Jason,
Thanks for this post. I am getting following error while I run the code:
`ValueError: time data '1901-Jan' does not match format`



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

Ensure your data is in CSV format and that the date column is in the correct format.



Denise December 13, 2017 at 7:07 pm #

REPLY ↗

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.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 #

REPLY ↗

Perhaps try manually differencing the data first?

Perhaps there are a lot of 0 values in your data that the model does not like?

Your Start in Machine Learning



Denise December 14, 2017 at 2:38 pm #

REPLY ↗

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

Sorry Denise, I'm not sure I follow



anand February 27, 2018 at 4:40 am #

Hey Denise, i got the same issue. did you



Kelly December 17, 2017 at 8:00 am #

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?

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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 might be easier.



Satyajit Pattnaik December 20, 2017 at 10:53 pm #

REPLY ↗

@Jason, This article has helped me a lot for the training set predictions which i had managed to do earlier too, but could you help me with the future forecasting, let say your date data is till 10th November, 2017 and i want to predict the values for the next one

Your Start in Machine Learning

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



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 can you help me with the future predictions for a week



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

Yes, see this post:

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



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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

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 #

REPLY ↗

It may mean that you could improve your model with some data transform, perhaps something like a boxcox?

Your Start in Machine Learning



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

REPLY ↗

@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 ur help



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

×

This is called a multi-step forecast and it is

More here:

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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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 #

REPLY ↗

Re the first question, it's good. I don't know how to do this with statsmodels off the cuff, some google searchers are needed.

Re multivariates, you may need to use ARIMAX or ~~SARIMAX or similar method~~

Your Start in Machine Learning



Fawad January 3, 2018 at 6:16 pm #

REPLY ↗

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'"
 I've just copy pasted the code from this website but it's not working. Any suggestions? I'm using Spyder



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

REPLY ↗

Ensure you remove the footer from the data.



Jelly January 9, 2018 at 1:58 pm #

Hi, may I know what are the yhat, obs and err? Is it the greater the better or the other way around? Thanks!



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

yhat are the predictions. obs are the observations of 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 #

REPLY ↗

Could you please have a blog on Anomaly detection using timeseries data, maybe from the above example itself.

Your Start in Machine Learning



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

REPLY ↗

Thanks for the suggestion.



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 #

I don't know about currency exchange pr...



Chintan January 25, 2018 at 7:20 am #

Hello,

Is it possible to predict hourly temperature for upcoming years ?

I am trying this out with ARIMA model, its giving me v...

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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 #

REPLY ↗

Hi Jason,

I have followed all of your posts related to Time Series to do my first data science project. I have done the parameter optimization also. The same code is working in my laptop but when i ran in Kaggle it shows "The computed initial AR coefficients are not stationary"

Your Start in Machine Learning

You should induce stationarity, choose a different model order, or you can pass your own start_params". The python version is same in my environment and in Kaggle. Is this common?



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

REPLY ↗

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



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

REPLY ↗

Hello, may I know what is the purpose for these:

```
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 #

Also, just to double confirm with you on my understanding. We take in all input in csv and fit into model, perform a prediction, then go thru the for loop again to recreate a new AR value, then go thru the for loop again?

In addition, the next row prediction is always depends on the past prediction values?

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

Yes, I believe so. Note, this is just one framing of the problem.



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

REPLY ↗

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 m

Your Start in Machine Learning



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?



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

This post might help:

<https://machinelearningmastery.com/make-sample-predictions-time-series-data-python/>



Ajay March 15, 2018 at 2:28 am #

Thanks Jason for making it simple. I run the program and got 1st error :

TypeError: Cannot cast ufunc subtract output from dtype('float64') to '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

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

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 %b %Y'

Your Start in Machine Learning

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 ↗

Perhaps your data contains the fo

<https://raw.githubusercontent.com/jbrownlee/MachineLearningFromScratch/master/datasets/shampoo-sales.csv>



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

When we use a recursive model for ARIMA, le

```
1 for t in range(len(test)):
2     try:
3         model = ARIMA(history, order=(4,0,0))
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 predictions are following the test values, it's like if test is following a pattern, predictions is following similar pattern, hence ultimately our ARIMA predictions isn't working properly, i hope you got my point.

For example: if test[0] keeps increasing till test[5] and decreases, then prediction[1] keeps increasing till predictions[5] and decreases..

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

Find out how in this *free* and *practical* email 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

Your Start in Machine Learning



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.



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

Hello! Thank you for this great tutorial. It'd be problems.

I want to implement a machine learning model to predict upcoming game week.

Say I have values for a player (Lukaku) for 28 game weeks. How do I predict the outcome features for those 28 weeks. How do I predict the outcome?

I am trying to predict total points to be scored by every player. So basically what should be the input to my model for : per live football games happening during the week, I w

Thank you 😊

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

I would recommend looking into rating systems:

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



Raphael March 30, 2018 at 2:20 am #

REPLY ↗

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?

Your Start in Machine Learning



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 but the results were weird.



Jason Brownlee April 2, 2018 at 2:30 pm #

Yes, you can use the predict() function to predict multiple steps into the future. Predicting multiple steps into the future is very challenging.



Ftima April 2, 2018 at 6:57 pm #

Hi,

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

Thanks

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

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



Hsiang April 9, 2018 at 9:35 am #

REPLY ↗

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?

Your Start in Machine Learning



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 ↗

Hi Jason,

Thank you for your answer. I have purchased 1 book from your site. I still have few more questions on ARIMA model.

- (1) The shampoo sale data obviously shows non-stationary data until it becomes stationary data by taking differences and then apply to ARIMA model. Is it correct?
- (2) Does the time series data with first-order differencing the time series data without differencing on ARIMA model equivalently process data with first-order differencing?
- (3) In this example, we chose ARIMA (5,1,0) and what I read from the book <https://www.otexts.org/fpp/2/1.html> is to check the PACF plot, instead ACF. Are there any things I

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee April 13, 2018 at 6:36 am #

REPLY ↗

The shampoo data is non stationary and should be differenced, this can happen before modeling or as part of the ARIMA.

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

Yes, you can check ACF and PACF for configuring the p and q variables, see this post:
<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>



Marco April 11, 2018 at 6:06 am #

REPLY ↗

Hi Jason,

In your code you use :

```
yhat=output[0]
```

Your Start in Machine Learning

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 #

REPLY ↗

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

could you please help me

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



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

REPLY ↗

It looks like you downloaded the dataset in a different format.

You can get the correct dataset here:

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



Harshil April 24, 2018 at 8:26 pm #

REPLY ↗

Hey Jason,

Best article I have ever seen. Currently I am working on data driven time series forecasting with PYTHON by 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?



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

REPLY ↗

Thanks.

Sorry, I don't have an example of ARIMA with mult

Your Start in Machine Learning



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



Muhammad May 8, 2018 at 10:42 am #

Hey Jason, I am new to data analytics. From stationary or non-stationary as well as how do you see
Thanks!

Your Start in Machine Learning

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

Email Address

START MY EMAIL COURSE



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

Yes, you can learn more about ACF and PACF and their interpretation here:
<https://machinelearningmastery.com/gentle-introduction-autocorrelation-partial-autocorrelation/>

REPLY ↗



Sven May 20, 2018 at 8:40 am #

Hello Jason,
can Autoregression model be used for forecasting stock price ?



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/>

Your Start in Machine Learning



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 ↗

Not in this case. Try it and see.



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

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, parse_dates=[0], index_col=0, squeeze=True,
date_parser=parser)
print(series.head())
series.plot()
pyplot.show().
```

Error:-

time data '19001-Jan' does not match format '%Y-%m'

Your Start in Machine Learning

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

Find out how in this *free* and *practical* email 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?

Your Start in Machine Learning



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 #

Hi Jason,
great tutorial, as always! Thank you very much for providing the community! You really helped me to get a better understanding.
Do you plan to make a tutorial on nonlinear time-series? I could not really find anything in this region.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee June 18, 2018 at 6:36 am #

Thanks for the suggestion.

I do hope to cover more methods for nonlinear time series in the future.



Saloni Patil June 21, 2018 at 6:59 pm #

REPLY ↗

Hi Jason

I tried the code with my data. ACF, PACF plots aren't showing me any significant correlations. Is there anything by which I can still try the forecast? What should be one's steps on encounter of such data?



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 ↗

Your Start in Machine Learning

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 article. I am running the same code above and get following error:

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 __init__(self,

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 __init__(self,

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, dates, freq, missing)

1016 self.k_diff = d

1017 self._first_unintegrate = unintegrate_levels(self.endog[:d], d)

~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in __init__(self,

endog, order, exog, dates, freq, missing)

452 super(ARIMA, 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]

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

Your Start in Machine Learning

```

~\AppData\Local\Continuum\anaconda3\lib\site-packages\statsmodels\tsa\arima_model.py in
_check estimable(nobs, n_params)
438 def _check estimable(nobs, n_params):
439 if nobs <= 440 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()
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 done something like this for dates
`dt=pd.date_range("2015-01-01", "2017-12-1", freq="MS")`

Can you explain what is wrong?

also,

I was under impression that you use `auto_corr` function to determine Q parameter in ARIMA model. then in your code when you call ARIMA why have you used (5,1,0) assuming it is (p,d,q)? i thought it was suppose to be (0,1,5)?

Your Start in Machine Learning

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

Find out how in this *free* and *practical* email 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-to-time-series-forecasting/>

Your Start in Machine Learning



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 pm #

X

Hello Jason,

Thanks for the helpful article.

My question is :

"A rolling forecast is required given the dependence on the AR model."

can you please elaborate?

How do we decide when to use Rolling forecast and what are the factors do you consider?

Thanks

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee June 27, 2018 at 8:22 am #

REPLY ↗

I believe I mean a walk-forward validation. More here:

<https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>



mithril July 5, 2018 at 1:30 am #

REPLY ↗

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

Your Start in Machine Learning

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 some advice?

Thank you.



Jason Brownlee July 5, 2018 at 7:57 am #

REPLY ↗

They could be exogenous binary variables.



Paola July 22, 2018 at 8:13 am #

Great article! But I have a question. I have a copy of your time series forecasting book. How do I obtain the acf and pacf plots for my data? I am using statsmodels because I will have more than 1000 lag values (as my data is a time series). I will really appreciate your 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 #

REPLY ↗

Great



Jason Brownlee July 23, 2018 at 6:00 am #

REPLY ↗

Thanks.



Ivan July 22, 2018 at 11:56 pm #

REPLY ↗

thank you very much, Jason.

Your Start in Machine Learning

You can master applied Machine Learning without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE

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 #

You can fit a final model and make a pred

Here's an example:

<https://machinelearningmastery.com/make-sample/>



Siddharth August 3, 2018 at 3:52 pm #

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 under the heading of "ARIMA model results". This summarizes the coefficient values used as well as the skill of the fit on the in-sample observations.

Can you please provide some explanation on their attributes and how the information assists us in the interpretation of the results

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 4, 2018 at 5:59 am #

REPLY ↗

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 ↗

Your Start in Machine Learning

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 Trafic
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 what data, only 52 values to predict the following one). Do you think there are other models that could be more suitable?

Thanks a lot for your help.



Jason Brownlee August 6, 2018 at 6:23 am #

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 #

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 Trafic
1 //pre-campaign // 108567 // 184196,63
2 //pre-campaign // 99358 // 166628,38
...
53 // Campaign // 135058 //240163,25
```

Your Start in Machine Learning

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



Jason Brownlee August 12, 2018 at 10:35 am #

Perhaps try it and see how you go.



Nii Anyetei August 7, 2018 at 5:46 am #

Hi Jason, the constant updates are great and trying to forecast solid waste generation in using ANN. problem. If you could at least get me a headway that would be great. I want to consider variables such as already going to school, age, gender, educational levels, etc. I hope to hear from you soon.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 7, 2018 at 6:35 am #

REPLY ↗

This is a good place to start for deep learning:

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



Wen Ge August 8, 2018 at 7:32 pm #

REPLY ↗

Many thanks Jason, it's really helpful!

Just one question, my data set contains some sales value = 0, would that affect the performance of ARIMA model? if there will be issues, anyway I can deal with the zero values in my data set? Thanks in advance for your advice!

Your Start in Machine Learning



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/
packages/pandas/io/parsers.py", line 3021, in convert
date_parser(*date_cols), errors='ignore')
```

```
File "/Users/Brian/PycharmProjects/MachineLearningN
return datetime.strptime('190' + x, '%Y-%m')
```

TypeError: strftime() argument 1 must be str, not num

During handling of the above exception, another excepti

Thank You

Brian

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 15, 2018 at 6:07 am #

REPLY ↗

Perhaps your data file, try this one instead:

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



Anton Petrov August 17, 2018 at 1:32 am #

REPLY ↗

The formating of csv seems different for everyone who downloads it, here's the format that is used by Jason (just copy pasted this into a shampoo-sales.csv file and save)

– 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

Your Start in Machine Learning

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
 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

Your Start in Machine Learning

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

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

[START MY EMAIL COURSE](#)



Jason Brownlee August 17, 2018 at 6:31 am #

REPLY ↗

It is also available on my github:

<https://github.com/jbrownlee/Datasets>



SA August 17, 2018 at 8:08 am #

REPLY ↗

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

Your Start in Machine Learning



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 "MSE=inf"

```
date price
0 20160227 427.1
1 20161118 750.9
2 20160613 690.9
3 20160808 588.7
4 20170206 1047.3
```

RangefIndex: 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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



SA August 18, 2018 at 8:16 am #

REPLY ↗

Hello Jason

Just to clarify my previous question that i have 700 rows of date and price and I would like select the best 70(window size) rows for prediction and decide on the frame size , frame step and extent of prediction.



Jason Brownlee August 19, 2018 at 6:14 am #

REPLY ↗

Sounds great, let me know how you go!

Your Start in Machine Learning



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 G Search and got best Arima model .

Now I want to use the result and train

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

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 21, 2018 at 2:13 pm #

An LSTM is a very different algorithm. Perhaps difference the series and use at least 6 time steps as input?



SA August 22, 2018 at 7:29 am #

I have 5 years of time series data .Will 6 time steps (6 days) be enough as window size.I want to get the best optimal window as input to LSTM !

Appreciate your feedback.

Your Start in Machine Learning



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 8:15 am #

I would recommend running t

Try time periods that might make sens



SA August 24, 2018 at 8:06 am #

So I did the for-loop and man

Now to calculate the RMSE do I need
order to calculate the RMSE or is there

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 24, 2018 at 9:16 am #

I would expect that you would fit a model for different sized windows and compare the RMSE of the models. The models could be anything you wish, try a few different approaches even.



SA August 25, 2018 at 7:38 am #

I got the following as example for two window size 360 days and 180 days

For 360 days

Window start after 0 days with window size 360 and step 100 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

Your Start in Machine Learning

Window start after 0 days with windwo size 180 and step 90 have RMSE
653.9070358902835

Window start after 90 days with windwo size 180 and step 90 have RMSE
326.7832188924093

Window start after 180 days with windwo size 180 and step 90 have RMSE
135.01118940666115

Window start after 270 days with windwo size 180 and step 90 have RMSE
38.422587695965746

Window start after 360 days with windwo size 180 and step 90 have RMSE
60.73374764651785

Window start after 450 days with windwo size 180 and step 90 have RMSE
52.386817309349176



Jason Brownlee August 26, 2018 at 7:04 am #

Well done!



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



Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

START MY EMAIL COURSE



Jason Brownlee August 27, 2018 at 6:10 am #

Thanks for your support!



Waldo August 18, 2018 at 9:36 pm #

REPLY ↗

Hi Jason! Here client and time series forecaster!

When forecasting, I very often get this error:

LinAlgError: SVD did not converge

Any ideas how to solve this in general?

Thanks!

Your Start in Machine Learning



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 #

Hey Jason, what model i can use to equipment variables that correlate with others and i need to identify



Jason Brownlee August 23, 2018 at 1:54 pm #

Try a suite of methods in order to discover



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. Which can be summed up as parameters of linear combination of previous terms for the AR and previous errors for the MA. A quick math formula could be :

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}$$

When the fit method is used, it takes the train values of the signal to fit the parameters (a and b)

When the forecast method is used, it forecast the next value of the signal using the fitted model and the train values

When the predict method is used, it forecast the next values of the signal from start to stop.

Let's say I fit a model on n steps in the train set. Now I want to make predictions. I can predict step n+1. Now I am days n+1 and I have the exact signal value. I would like to actualize the model to predict n+2.

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 no

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE

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 initial value for the next step. This is equal to the number of significant lags from the auto correlation plot.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Matthew Orehek September 7, 2018 at 7:28 am #

Hi Jason – Very helpful post here, thanks for sharing! I think you are right that this is equal to the number of significant lags from the auto correlation plot. I would like to add some more context to this part of the problem. Thanks.



Jason Brownlee September 7, 2018 at 8:11 am #

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 #

REPLY ↗

I used your code to forecast daily temperature (it has a lag of 365). The forecast is always a day behind, i.e. learning history cannot accurately forecast next day's temperature. I've played with the params with AIC.



Jason Brownlee September 12, 2018 at 2:39 pm #

REPLY ↗

Perhaps try alternate configurations?

Perhaps try alternate algorithms?

Perhaps try additional transforms to the data?

This might help:

<https://machinelearningmastery.com/how-to-develop-forecasting-models-time-series-data/>

Your Start in Machine Learning



Anuradha Chaurasia September 16, 2018 at 11:23 pm #

REPLY ↗

How to use ARIMA model in SPSS with few sample 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 #

Hi Jason,

Thanks for sharing! Very helpful post.

Recently I am writing the methodology of ARIMA, but I ARIMA formulas contain constant but some don't have ARIMA formula information) of "statsmodels.tsa.arima_

Thank you in advance.

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee September 18, 2018 at 6:20 am #

REPLY ↗

The best textbook on ARIMA is:

<https://amzn.to/2MD9IKw>



Milind Mahajani September 20, 2018 at 12:44 am #

REPLY ↗

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

Your Start in Machine Learning

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!



Kruthika Vishwanath September 25, 2018 at 6:5

REPLY ↗

Hi Jason,

Thanks for this post.

I am working on finding an anomaly using arima. Will I predicted value shown above ?

Thanks,
Kruthika

Your Start in Machine Learning

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

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

Email Address

START MY EMAIL COURSE



Jason Brownlee September 25, 2018 at 2:43

Sorry, I don't have examples of using ARIMA for anomaly detection.



Bhadri September 29, 2018 at 6:20 pm #

REPLY ↗

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 ↗

Your Start in Machine Learning

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..



Jason Brownlee October 1, 2018 at 6:24 am #

REPLY ↗

Yes, there are examples here:

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



awa October 17, 2018 at 4:36 pm #

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 period 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 according to the data that I mentioned earlier?

Your Start in Machine Learning

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

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

Leave a Reply

Your Start in Machine Learning

Name (required)

Email (will not be published) (requi

Website

SUBMIT COMMENT

Your Start in Machine Learning

You can master applied Machine Learning
without the math or fancy degree.

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

Email Address

[START MY EMAIL COURSE](#)



Hi, I'm Jason Brownlee, Ph.D.
My goal is to make developers like YOU

[Read More](#)

Time Series Forecasting with Python

Explore time series data and predict the future.

[Your Start in Machine Learning](#)

[Click to Get Started Now!](#)

Introduction to Time Series Forecasting with Python

[How to Prepare Data and Develop Models to Predict the Future](#)

Jason Brownlee

MACHINE
LEARN
MAS

Your Start in Machine Learning

X

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

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

[START MY EMAIL COURSE](#)

POPULAR



[How to Develop a Deep Learning Photo Caption Generator](#)

NOVEMBER 27, 2017



[How to Develop a Neural Machine Translation System](#)

JANUARY 10, 2018



[How to Develop an Encoder-Decoder Model for Sequence-to-Sequence Prediction in Keras](#)

NOVEMBER 2, 2017



[How to Develop a Word-Level Neural Language Model and Use it to Generate Text](#)

NOVEMBER 10, 2017



[Difference Between Classification and Regression in Machine Learning](#)

DECEMBER 11, 2017



[How to Develop an N-gram Multichannel Convolutional Neural Network for Sentiment Analysis](#)

JANUARY 12, 2018



[So, You are Working on a Machine Learning Problem...](#)

APRIL 4, 2018

You might also like...

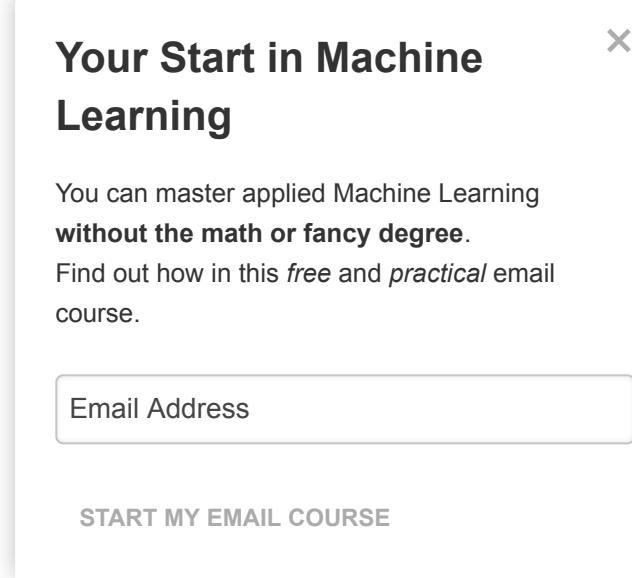
- [How to Install Python for Machine Learning](#)

[Your Start in Machine Learning](#)

- Your First Machine Learning Project in Python
- Your First Neural Network in Python
- Your First Classifier in Weka
- Your First Time Series Forecasting Project

© 2018 Machine Learning Mastery. All Rights Reserved.

[Privacy](#) | [Disclaimer](#) | [Terms](#) | [Contact](#)



Your Start in Machine Learning