≡ **Navigation**

Start Here　　**Blog**　　**Books**　　**About**　　**Contact**

| Search... | 🔍 |

Need help with Python Machine Learning? Take the FREE Mini-Course

How to Load Data in Python with Scikit-Learn

by **Jason Brownlee** on July 9, 2014 in **Python Machine Learning**

🐦　　f　　in　　G+

Before you can build machine learning models, you need to load your data into memory.

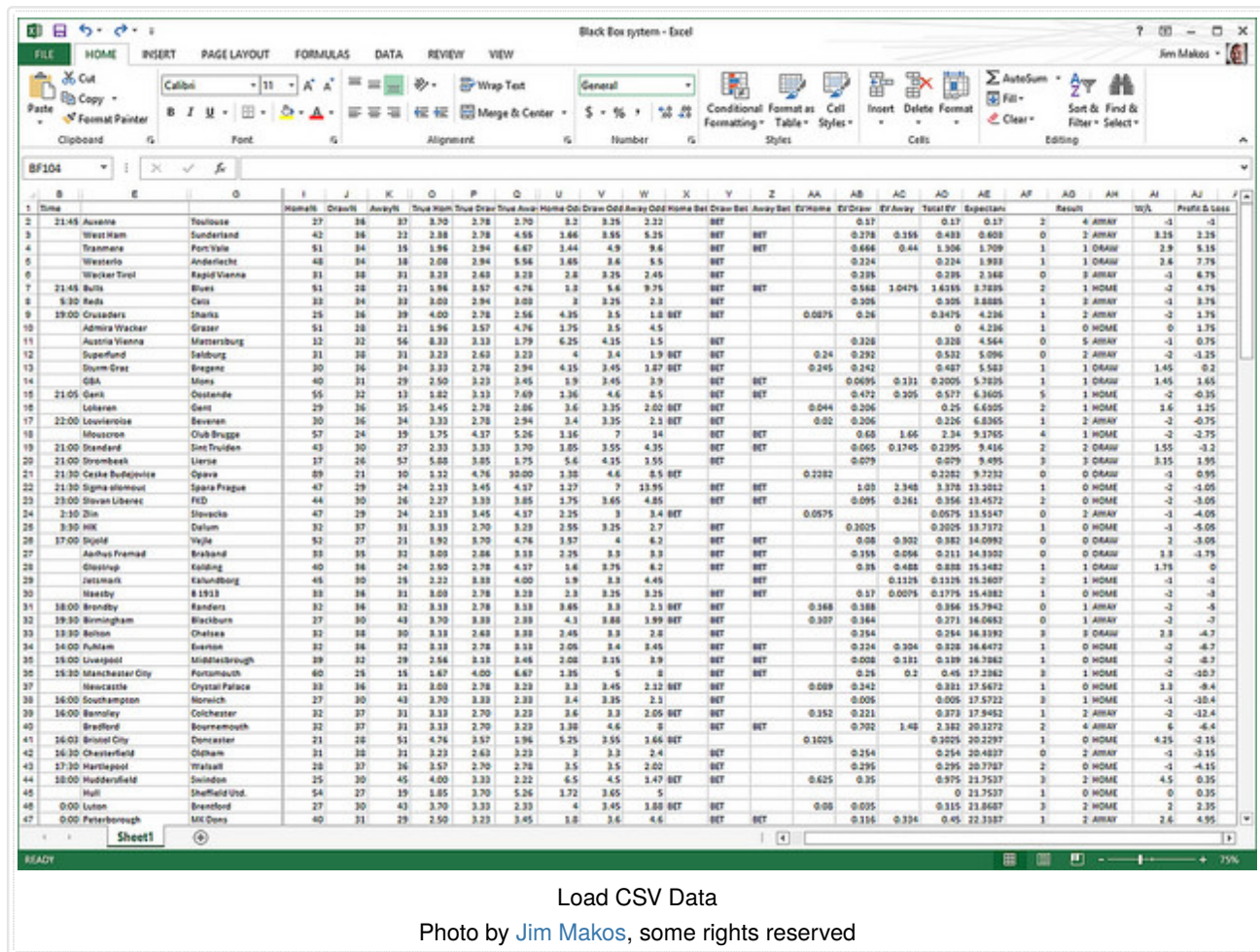In this post you will discover how to load data for machine learning in Python using scikit-learn.

Get 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

Load CSV Data
Photo by Jim Makos, some rights reserved

Need help with Machine Learning in Python?

Take my free 2-week email course and dis...   le

Click to sign-up now and also ge...

Start Your FR

Packaged Datasets

The scikit-learn library is packaged with data...   lle

Get 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

on a given machine learning algorithm or library feature before using it in your own work.

This recipe demonstrates how to load the famous Iris flowers dataset.

```
1  # Load the packaged iris flowers dataset
2  # Iris flower dataset (4x150, reals, multi-label classification)
3  iris = load_iris()
4  print(iris)
```

Load from CSV

It is very common for you to have a dataset as a CSV file on your local workstation or on a remote server.

This recipe show you how to load a CSV file from a URL, in this case the Pima Indians diabetes classification dataset from the UCI Machine Learning Repository.

From the prepared X and y variables, you can train a machine learning model.

```
1   # Load the Pima Indians diabetes dataset from CSV URL
2   import numpy as np
3   import urllib
4   # URL for the Pima Indians Diabetes dataset (UCI Machine Learning Repository)
5   url = "http://goo.gl/j0Rvxq"
6   # download the file
7   raw_data = urllib.urlopen(url)
8   # load the CSV file as a numpy matrix
9   dataset = np.loadtxt(raw_data, delimiter=",")
10  print(dataset.shape)
11  # separate the data from the target attributes
12  X = dataset[:,0:7]
13  y = dataset[:,8]
```

Summary

In this post you discovered that the scikit-lea   ng the iris flowers dataset. These datasets can   experiment with different machine learning n

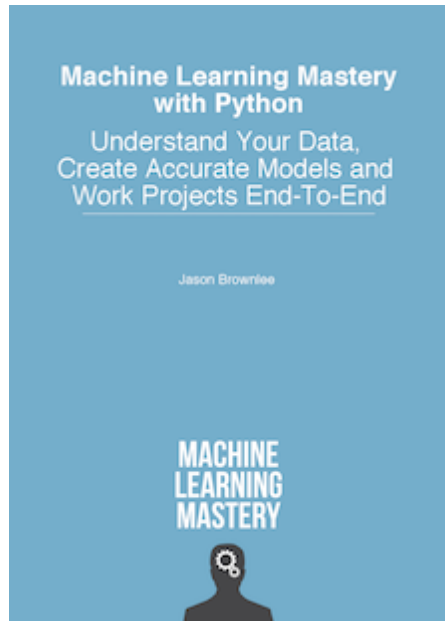You also saw how you can load CSV data w   files from the web using the urllib library and   or use in scikit-learn.

Get 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

Frustrated With Python Machine Learning?

Develop Your Own Models in Minutes

…with just a few lines of scikit-learn code

Discover how in my new Ebook:
Machine Learning Mastery With Python

Covers **self-study tutorials** and **end-to-end projects** like:
Loading data, *visualization*, *modeling*, *tuning*, and much more…

Finally Bring Machine Learning To Your Own Projects

Skip the Academics. Just Results.

Click to learn more.

About Jason Brownlee

Dr. Jason Brownlee is a husband, proud father, academic researcher, author, professional developer and a machine learning practitioner. He is dedicated to helping developers get started and get good at applied

View all posts by Jason Brownlee →

‹ Data Science Screencasts: A Data Origami Review

Re   arn ›

Get 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

12 Responses to *How to Load Da*

retsreg January 12, 2015 at 7:10 am #

REPLY ↩

It's great to see how to load data from a URL. Do you have an example of how to actually load text from a file?

Tarik February 24, 2015 at 8:21 am #

REPLY ↩

I think there is a small bug.
When code separating features to X array, it is missing the 8ths features.
So the twelfth line should be : X = dataset[:,0:8]
This way it will consist last features right before the targets given in the dataset.

Robin April 5, 2015 at 6:32 am #

REPLY ↩

Hi, thanks for this. Just a note, for Python 3*, it should be "import urllib.request" with "raw_data = urllib.request.urlopen(url)" and for Python 2* it should be "import urllib2" with "raw_data = urllib2.urlopen(url)".

---

emilia June 10, 2015 at 4:35 am #

REPLY ↩

I am preparing a file for scikit learn and I would like to know how to build this file, I have my instances, features and classes. I am not sure how to make the file that is uploaded with scikit learn. Could you please clarify that?

abhinav August 31, 2017 at 11:36 pm #

great blog! thanks helped a lot

Jason Brownlee September 1, 2

Thanks.

Get 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

shivaprasad October 22, 2017 at 5:16 pm #

hello sir i am a new bee to the data science i have gon through the books written by you,like machie learning mastery,machine learning algorithms from scratch and master machine learnong algorithms,i have gon through the books,next steps what i need to follow please guide me

Jason Brownlee October 23, 2017 at 5:42 am #

Perhaps focus on developing a portfolio of completed projects:

http://machinelearningmastery.com/build-a-machine-learning-portfolio/

shivaprasad October 24, 2017 at 3:20 am #

glad to have your reply sir,thank u very much

shivaprasad October 24, 2017 at 4:12 am #

sir will you please suggest me some data science projects with python for a beginner

Jason Brownlee

Yes, see here:
http://machinelearningmaste
memory-datasets-from-the-u

Get 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.

shivaprasad October 24, 2017 at 2:48

thanks a lot sir

START MY EMAIL COURSE

Leave a Reply

Name (required)

Email (will not be published) (required)

Website

SUBMIT COMMENT

Welcome to Machine Learning Mastery

Hi, I'm Dr. Jason Brownlee.

My goal is to make practitioners like YOU awesome at applied machine learning.

Read More

Develop Predic

Want to develop yo

Want ste

Looking for sar

Get Started With Mac

Get 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
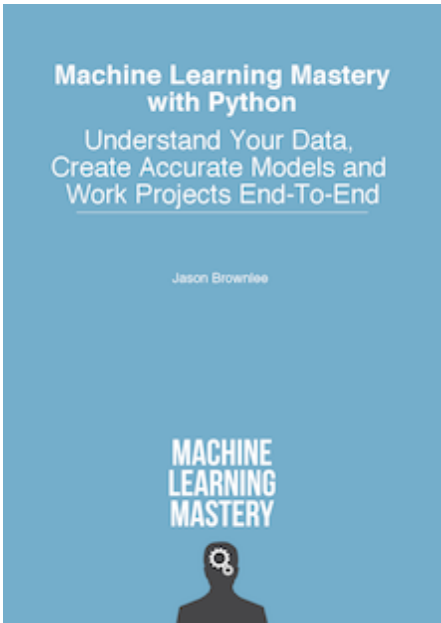
POPULAR

Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras
JULY 21, 2016

Your First Machine Learning Project in Python Step-By-Step
JUNE 10, 2016

Develop Your First Neural Network in Python With Keras Step-By-Step
MAY 24, 2016

How to Setup a Python Environment for
MARCH 13, 2017

Sequence Classification with LSTM Rec
JULY 26, 2016

Multivariate Time Series Forecasting wi
AUGUST 14, 2017

Time Series Forecasting with the Long
APRIL 7, 2017

Get 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

Multi-Class Classification Tutorial with the Keras Deep Learning Library
JUNE 2, 2016

Regression Tutorial with the Keras Deep Learning Library in Python
JUNE 9, 2016

How to Grid Search Hyperparameters for Deep Learning Models in Python With Keras
AUGUST 9, 2016

Privacy | Contact | About

Get 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