## The Top 5 Machine Learning Libraries in Python

May 18, 2018

## 1 The Top 5 Machine Learning Libraries in Python

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to objectoriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal lang-uage for scripting and rapid application development in many areas on most platforms.

## 2 Lesson 1: Intro to pandas data structures

Our first top 5 Library is Pandas. **Pandas** is an open source Python library for data analysis. Pandas makes Python great for analysis.

```
In [5]: # Let's bring in pandas
    import pandas as pd
```

We can import the entire library but that's often not the Pythonic way.

In the code below we are using the from **keyword** to only import what we need.

Interstingly, in most tutorials all of pandas is imported.

At the very basic level, Pandas objects can be thought of as enhanced versions of NumPy structured arrays in which the rows and columns are identified with labels rather than simple integer indices.

There are **three** core Pandas data structures. They are:

Series - A pandas Series is a one-dimensional array of indexed data.

Dataframe - The DataFrame can be thought of either as a generalization of a NumPy array, or as a specialization of a Python dictionary.

Index - This Index object is an interesting structure in itself, and it can be thought of as an ordered set.

```
In [8]: # Let's create a series
        # A series is a one dimensional array like object.
        s = pd.Series([3,5,5,9,6])
        # print the series
Out[8]: 0
             3
        1
             5
        3
             9
             6
        dtype: int64
In [9]: # Prints X elements of the series; s,head(X). Default X = 5
        s.head()
Out[9]: 0
        1
             5
        2
             5
        3
             9
        dtype: int64
   Methods end with parentheses, while attributes don't:
     So head() in the exmple above is a method.
   Let's read some tabular data into our workspace using pandas.
     When you hear the word tabular data think excel spreadsheet.
   Let's read some data into a table and manipulate that data
In [10]: # Reading the data into variable ufo through url.
         ufo = pd.read_table('http://bit.ly/uforeports', sep=',')
In [11]: # Printing first 10 rows.
         ufo.head(10)
Out[11]:
                             City Colors Reported Shape Reported State
                                                                                        Time
                                                          TRIANGLE
                                                                              6/1/1930 22:00
         0
                           Ithaca
                                                NaN
                                                                       NY
                      Willingboro
                                                NaN
                                                              OTHER
                                                                       NJ
                                                                            6/30/1930 20:00
         1
         2
                                                               OVAL
                                                                            2/15/1931 14:00
                          Holyoke
                                                NaN
                                                                       CO
         3
                          Abilene
                                                NaN
                                                                       KS
                                                                              6/1/1931 13:00
                                                               DISK
         4
           New York Worlds Fair
                                                NaN
                                                             LIGHT
                                                                       NY
                                                                            4/18/1933 19:00
         5
                      Valley City
                                                                            9/15/1934 15:30
                                                NaN
                                                               DISK
                                                                       ND
         6
                      Crater Lake
                                                NaN
                                                            CIRCLE
                                                                       CA
                                                                              6/15/1935 0:00
         7
                             Alma
                                                NaN
                                                               DISK
                                                                       MΙ
                                                                              7/15/1936 0:00
         8
                          Eklutna
                                                NaN
                                                              CIGAR
                                                                       AK 10/15/1936 17:00
         9
                          Hubbard
                                                NaN
                                                          CYLINDER
                                                                       OR
                                                                              6/15/1937 0:00
```

## In [12]: # Printing attribute State ufo['State'] Out[12]: 0 NY 1 NJ 2 CO 3 KS 4 NY5 ND 6 CA7 ΜI 8 AK 9 OR 10 CA11 AL 12 SC 13 ΙA 14 $\mathtt{MI}$ 15 CA $\mathsf{C}\mathsf{A}$ 16 17 GA 18 TN19 AK 20 NE21 LA 22 LA 23 ΚY 24 WV25 CAWV26 27 NM28 NM29 UT . . 18211 $\mathsf{M}\mathsf{A}$ 18212 $\mathsf{C}\mathsf{A}$ 18213 CA18214 TX18215 $\mathsf{TX}$ 18216 CA18217 CO 18218 TX18219 CA18220 CA18221 NH18222 ${\tt PA}$ 18223 SC

18224

OK

```
18226
                   CA
         18227
                   CA
         18228
                   TX
         18229
                   IL
         18230
                   CA
         18231
                   CA
         18232
                   WI
         18233
                   AK
         18234
                   CA
         18235
                   AZ
         18236
                   IL
         18237
                   ΙA
         18238
                   WI
                   WI
         18239
         18240
                   FL
         Name: State, Length: 18241, dtype: object
In [13]: # Creating and printing a new attribute Location.
         ufo['Location'] = ufo.City + ', ' + ufo.State
         ufo.head()
Out[13]:
                             City Colors Reported Shape Reported State
                                                                                      Time
                                                          TRIANGLE
                                                                            6/1/1930 22:00
         0
                           Ithaca
                                               NaN
         1
                      Willingboro
                                               NaN
                                                             OTHER
                                                                      NJ
                                                                           6/30/1930 20:00
         2
                          Holyoke
                                               NaN
                                                              OVAL
                                                                      CO
                                                                           2/15/1931 14:00
                          Abilene
                                                              DISK
                                                                            6/1/1931 13:00
         3
                                               NaN
                                                                      KS
            New York Worlds Fair
                                               NaN
                                                             LIGHT
                                                                      NY 4/18/1933 19:00
                             Location
         0
                           Ithaca, NY
         1
                      Willingboro, NJ
         2
                          Holyoke, CO
                          Abilene, KS
         3
            New York Worlds Fair, NY
In [14]: # Description of variable
         ufo.describe()
Out[14]:
                     City Colors Reported Shape Reported
                                                            State
                                                                                Time
         count
                    18216
                                      2882
                                                     15597
                                                            18241
                                                                               18241
         unique
                     6476
                                        27
                                                        27
                                                               52
                                                                               16145
         top
                 Seattle
                                       RED
                                                    LIGHT
                                                               CA
                                                                   11/16/1999 19:00
         freq
                      187
                                       780
                                                     2803
                                                             2529
                                                                                  27
                     Location
                        18216
         count
                         8029
         unique
```

18225

CA

```
Seattle, WA
         top
                         187
         freq
In [15]: # Get the shape of variable(rows, columns)
         ufo.shape
Out[15]: (18241, 6)
In [16]: # Get the datatypes of all the attributes.
         ufo.dtypes
Out[16]: City
                            object
         Colors Reported
                            object
         Shape Reported
                            object
         State
                            object
         Time
                            object
         Location
                            object
         dtype: object
In [17]: # Get the names and datatypes of all the attributes.
         ufo.columns
Out[17]: Index(['City', 'Colors Reported', 'Shape Reported', 'State', 'Time',
                'Location'],
               dtype='object')
In [18]: # Raeading data into variable and printing first 5 rows.
         ufo = pd.read_table('http://bit.ly/uforeports', sep=',')
         ufo.head()
Out[18]:
                            City Colors Reported Shape Reported State
                                                                                   Time
                          Ithaca
                                             NaN
                                                        TRIANGLE
                                                                         6/1/1930 22:00
         0
                                                                    NY
         1
                     Willingboro
                                             NaN
                                                           OTHER
                                                                    NJ 6/30/1930 20:00
         2
                         Holyoke
                                             NaN
                                                            OVAL
                                                                    CO 2/15/1931 14:00
         3
                         Abilene
                                             NaN
                                                            DISK
                                                                       6/1/1931 13:00
                                                                    KS
         4 New York Worlds Fair
                                             NaN
                                                           LIGHT
                                                                    NY 4/18/1933 19:00
In [19]: # Remove attribute Colors reported and printing first 5 rows of the varible afterword
         ufo.drop('Colors Reported', axis=1, inplace=True)
         ufo.head()
Out[19]:
                            City Shape Reported State
                                                                   Time
                                       TRIANGLE
                                                         6/1/1930 22:00
         0
                          Ithaca
         1
                     Willingboro
                                          OTHER
                                                    NJ 6/30/1930 20:00
         2
                                                    CO 2/15/1931 14:00
                         Holyoke
                                           OVAL
                         Abilene
                                           DISK
                                                         6/1/1931 13:00
                                                   KS
         4 New York Worlds Fair
                                                   NY 4/18/1933 19:00
                                          LIGHT
In [20]: # Droping two attributes state and time and printing first 5 rows.
         ufo.drop(['State','Time'], axis=1, inplace=True)
         ufo.head()
```

```
Out [20]:
                             City Shape Reported
                                         TRIANGLE
         0
                           Ithaca
         1
                      Willingboro
                                            OTHER
         2
                          Holyoke
                                             OVAL
         3
                          Abilene
                                             DISK
         4 New York Worlds Fair
                                            LIGHT
In [21]: ufo = pd.read_table('http://bit.ly/uforeports', sep=',')
         ufo.head()
Out [21]:
                             City Colors Reported Shape Reported State
                                                                                      Time
                                                                           6/1/1930 22:00
                                               NaN
                                                          TRIANGLE
                           Ithaca
         1
                                                             OTHER
                                                                      NJ
                                                                          6/30/1930 20:00
                      Willingboro
                                               NaN
         2
                                                                      CO 2/15/1931 14:00
                          Holyoke
                                               NaN
                                                              OVAL
         3
                          Abilene
                                                                           6/1/1931 13:00
                                               NaN
                                                              DISK
                                                                      KS
           New York Worlds Fair
                                               NaN
                                                             LIGHT
                                                                      NY 4/18/1933 19:00
In [22]: # Sorting the data wrt State attribute in Descending order and printing first 25 rows
         ufo.State.sort_values(ascending=False).head(25)
Out[22]: 12079
                  WY
         11490
                  WY
         11333
                  WY
         4866
                  WY
         3326
                  WY
         3328
                  WY
         16594
                  WY
         1177
                  WY
         378
                  WY
         5065
                  WY
         7684
                  WY
         6116
                  WY
         10729
                  WY
         12072
                  WY
         16637
                  WY
         14618
                  WY
         7491
                  WY
         12063
                  WY
         14240
                  WY
         14586
                  WY
         15667
                  WY
         7485
                  WY
         14747
                  WY
         1461
                  WY
                  WY
         1442
         Name: State, dtype: object
```

ufo.sort\_values('City').head()

In [23]: # Sorting the data wrt City names in ascending order (default) and printing first 5 r

```
Out[23]:
                      City Colors Reported Shape Reported State
                                                                                Time
         1761
                Abbeville
                                                      DISK
                                                               SC
                                                                    12/10/1968 0:30
                                        NaN
         4553
                 Aberdeen
                                        NaN
                                                  CYLINDER
                                                               WA
                                                                    6/15/1981 22:00
         16167
                 Aberdeen
                                        {\tt NaN}
                                                   VARIOUS
                                                               OH
                                                                    3/29/2000 3:00
         14703
                 Aberdeen
                                                               WA
                                                                    9/30/1999 21:00
                                        {\tt NaN}
                                                  TRIANGLE
         389
                 Aberdeen
                                    ORANGE
                                                    CIRCLE
                                                               SD 11/15/1956 18:30
In [24]: # Sort the data wrt two attributes first City and then State names in ascending order
         ufo.sort_values(['City', 'State']).head(25)
```

Out[24]:		City	Colors Reported	Shape Reported	State	\
	1761	Abbeville	NaN	DISK	SC	
	2297	Aberdeen	NaN	TRIANGLE	MD	
	9404	Aberdeen	NaN	DISK	MD	
	16167	Aberdeen	NaN	VARIOUS	OH	
	389	Aberdeen	ORANGE	CIRCLE	SD	
	4553	Aberdeen	NaN	CYLINDER	WA	
	12294	Aberdeen	NaN	FIREBALL	WA	
	14703	Aberdeen	NaN	TRIANGLE	WA	
	17809	Aberdeen	GREEN	FIREBALL	WA	
	3	Abilene	NaN	DISK	KS	
	8783	Abilene	NaN	NaN	KS	
	6654	Abilene	NaN	TRIANGLE	TX	
	8357	Abilene	NaN	SPHERE	TX	
	10883	Abilene	NaN	NaN	TX	
	6621	Abingdon	NaN	EGG	IL	
	10861	Abingdon	NaN	LIGHT	VA	
	12229	Acampo	ORANGE	TEARDROP	CA	
	6164	Accokeek	NaN	NaN	MD	
	3534	Accord	RED	TRIANGLE	NY	
	705	Acoma	NaN	CYLINDER	NM	
	7263	Acton	NaN	NaN	CA	
	10695	Acton	ORANGE	OVAL	CA	
	14553	Acton	GREEN	FIREBALL	CA	
	5140	Acton	RED	TRIANGLE	ME	
	6760	Acton to Chelmsford	NaN	TRIANGLE	MA	

Time 1761 12/10/1968 0:30 2297 8/18/1972 1:30 9404 6/15/1996 13:30 16167 3/29/2000 3:00 389 11/15/1956 18:30 4553 6/15/1981 22:00 12294 10/4/1998 4:42 14703 9/30/1999 21:00 17809 10/29/2000 17:25 6/1/1931 13:00