4 - Statistics and Machine Learning

December 1, 2015

1 NumPy

NumPy is the fundamental package for scientific computing with Python.

   Installation instructions: http://www.scipy.org/scipylib/download.html

1.1 NumPy arrays

In [ ]:

In [ ]:

2 PyPlot

PyPlot provides a MATLAB-like plotting framework on top of NumPy. It is part of matplotlib package

   Installation instructions: http://matplotlib.org/users/installing.html

In [ ]:

In [ ]:

3 Pandas

Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

   Installation instructions: http://pandas.pydata.org/pandas-docs/stable/install.html

3.1 Series

In [ ]:

3.2 Dataframe

In [ ]:

3.2.1 Indexing

In [ ]:

3.2.2 Selecting

In [ ]:

3.2.3 Sorting

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

4 Exercise: classification of last 12-month claim data records

In [ ]:

In [ ]:

In [ ]: