project_4_starter

April 6, 2020

1 Project 4: Multi-factor Model

1.1 Instructions

Each problem consists of a function to implement and instructions on how to implement the function. The parts of the function that need to be implemented are marked with a `# TODO` comment. After implementing the function, run the cell to test it against the unit tests we've provided. For each problem, we provide one or more unit tests from our `project_tests` package. These unit tests won't tell you if your answer is correct, but will warn you of any major errors. Your code will be checked for the correct solution when you submit it to Udacity.

1.2 Packages

When you implement the functions, you'll only need to you use the packages you've used in the classroom, like Pandas and Numpy. These packages will be imported for you. We recommend you don't add any import statements, otherwise the grader might not be able to run your code.

The other packages that we're importing are `project_helper` and `project_tests`. These are custom packages built to help you solve the problems. The `project_helper` module contains utility functions and graph functions. The `project_tests` contains the unit tests for all the problems.

1.2.1 Install Packages

You can view the list of packages to install, as mentioned in `requirements.txt` as:

```
! cat requirements.txt
```

```
In [1]: import sys
        !{sys.executable} -m pip install -r requirements.txt
```

```
Collecting alphalens==0.3.2 (from -r requirements.txt (line 1))
  Downloading https://files.pythonhosted.org/packages/a5/dc/2f9cd107d0d4cf6223d37d81ddfbbdbf0d70
    100% || 18.9MB 2.1MB/s eta 0:00:01    54% |              | 10.3MB 30.2MB/s eta 0:00:01    62
Requirement already satisfied: colour==0.1.5 in /opt/conda/lib/python3.6/site-packages (from -r
Collecting cvxpy==1.0.3 (from -r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/a1/59/2613468ffbbe3a818934d06b81b9f4877fe0
    100% || 880kB 15.2MB/s ta 0:00:01
Requirement already satisfied: cycler==0.10.0 in /opt/conda/lib/python3.6/site-packages/cycler-0
Collecting numpy==1.13.3 (from -r requirements.txt (line 5))
  Downloading https://files.pythonhosted.org/packages/57/a7/e3e6bd9d595125e1abbe162e323fd2d06f6f
```

```
    100% || 17.0MB 1.8MB/s eta 0:00:01  5% |                                | 1.0MB 28.6MB/s eta 0
Collecting pandas==0.18.1 (from -r requirements.txt (line 6))
  Downloading https://files.pythonhosted.org/packages/11/09/e66eb844daba8680ddff26335d5b4fead77f
    100% || 7.3MB 8.0MB/s eta 0:00:01    11% |                               | 819kB 24.0MB/s eta 0
Collecting plotly==2.2.3 (from -r requirements.txt (line 7))
  Downloading https://files.pythonhosted.org/packages/99/a6/8214b6564bf4ace9bec8a26e7f89832792be
    100% || 1.1MB 15.6MB/s ta 0:00:01
Requirement already satisfied: pyparsing==2.2.0 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: python-dateutil==2.6.1 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: pytz==2017.3 in /opt/conda/lib/python3.6/site-packages (from -r r
Requirement already satisfied: requests==2.18.4 in /opt/conda/lib/python3.6/site-packages (from
Collecting scipy==1.0.0 (from -r requirements.txt (line 12))
  Downloading https://files.pythonhosted.org/packages/d8/5e/caa01ba7be11600b6a9d39265440d7b3be3d
    100% || 50.0MB 716kB/s eta 0:00:01  7% |                               | 3.8MB 27.4MB/s eta 0:
Requirement already satisfied: scikit-learn==0.19.1 in /opt/conda/lib/python3.6/site-packages (f
Requirement already satisfied: six==1.11.0 in /opt/conda/lib/python3.6/site-packages (from -r re
Collecting tables==3.3.0 (from -r requirements.txt (line 15))
  Downloading https://files.pythonhosted.org/packages/09/e7/72ca83c7bd75db94c23fcaf58debe1be5e98
    100% || 4.6MB 5.9MB/s ta 0:00:011    22% |                         | 1.0MB 23.2MB/s eta 0:00:
Collecting tqdm==4.19.5 (from -r requirements.txt (line 16))
  Downloading https://files.pythonhosted.org/packages/71/3c/341b4fa23cb3abc335207dba057c790f3bb3
    100% || 61kB 10.3MB/s ta 0:00:01
Collecting zipline===1.2.0 (from -r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/15/d3/689f2a940478b82ac57c751a4046059821f
    100% || 665kB 16.3MB/s ta 0:00:01 3% |                               | 20kB 6.2MB/s eta 0:00
Requirement already satisfied: matplotlib>=1.4.0 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: seaborn>=0.6.0 in /opt/conda/lib/python3.6/site-packages (from al
Requirement already satisfied: statsmodels>=0.6.1 in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: IPython>=3.2.3 in /opt/conda/lib/python3.6/site-packages (from al
Collecting osqp (from cvxpy==1.0.3->-r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/6c/59/2b80e881be227eecef3f2b257339d182167b
    100% || 215kB 18.1MB/s ta 0:00:01  24% |                          | 51kB 9.0MB/s eta 0:00:01
Collecting ecos>=2 (from cvxpy==1.0.3->-r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/55/ed/d131ff51f3a8f73420eb1191345eb49f269f
    100% || 153kB 10.8MB/s ta 0:00:01    83% |     | 122kB 18.8MB/s eta 0:00:01
Collecting scs>=1.1.3 (from cvxpy==1.0.3->-r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/1a/72/33be87cce255d4e9dbbfef547e9fd6ec7ee9
    100% || 3.6MB 8.3MB/s eta 0:00:01   8% |                               | 317kB 21.2MB/s eta 0:
Collecting multiprocess (from cvxpy==1.0.3->-r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/58/17/5151b6ac2ac9b6276d46c33369ff814b0901
    100% || 1.6MB 7.7MB/s eta 0:00:01    96% | | 1.5MB 22.2MB/s eta 0:00:01
Requirement already satisfied: fastcache in /opt/conda/lib/python3.6/site-packages (from cvxpy==
Requirement already satisfied: toolz in /opt/conda/lib/python3.6/site-packages (from cvxpy==1.0.
Requirement already satisfied: decorator>=4.0.6 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: nbformat>=4.2 in /opt/conda/lib/python3.6/site-packages (from plo
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /opt/conda/lib/python3.6/site-packages (
Requirement already satisfied: idna<2.7,>=2.5 in /opt/conda/lib/python3.6/site-packages (from re
Requirement already satisfied: urllib3<1.23,>=1.21.1 in /opt/conda/lib/python3.6/site-packages (
```

```
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.6/site-packages (fro
Requirement already satisfied: numexpr>=2.5.2 in /opt/conda/lib/python3.6/site-packages (from ta
Requirement already satisfied: pip>=7.1.0 in /opt/conda/lib/python3.6/site-packages (from ziplin
Requirement already satisfied: setuptools>18.0 in /opt/conda/lib/python3.6/site-packages (from z
Collecting Logbook>=0.12.5 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/2f/d9/16ac346f7c0102835814cc9e5b684aaadea1
    100% || 92kB 12.2MB/s ta 0:00:01
Collecting requests-file>=1.4.1 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/23/9c/6e63c23c39e53d3df41c77a3d05a49a42c4e
Collecting pandas-datareader<0.6,>=0.2.1 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/40/c5/cc720f531bbde0efeab940de400d0fcc95e8
    100% || 81kB 6.3MB/s ta 0:00:011
Requirement already satisfied: patsy>=0.4.0 in /opt/conda/lib/python3.6/site-packages (from zipl
Requirement already satisfied: Cython>=0.25.2 in /opt/conda/lib/python3.6/site-packages (from zi
Collecting cyordereddict>=0.2.2 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/d1/1a/364cbfd927be1b743c7f0a985a7f1f7e8a51
    100% || 143kB 15.3MB/s ta 0:00:01
Collecting bottleneck>=1.0.0 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/5b/08/278c6ee569458e168096f6b51019cc1c81c2
    100% || 92kB 9.1MB/s eta 0:00:01
  Installing build dependencies ... done
Collecting contextlib2>=0.4.0 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/85/60/370352f7ef6aa96c52fb001831622f50f923
Requirement already satisfied: networkx<2.0,>=1.9.1 in /opt/conda/lib/python3.6/site-packages (f
Collecting bcolz<1,>=0.12.1 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/6c/8b/1ffa01f872cac36173c5eb95b58c01040d8d
    100% || 624kB 15.8MB/s ta 0:00:01
Requirement already satisfied: click>=4.0.0 in /opt/conda/lib/python3.6/site-packages (from zipl
Collecting multipledispatch>=0.4.8 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/89/79/429ecef45fd5e4504f7474d4c3c3c4668c26
Requirement already satisfied: MarkupSafe>=0.23 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: Mako>=1.0.1 in /opt/conda/lib/python3.6/site-packages/Mako-1.0.7-
Requirement already satisfied: sqlalchemy>=1.0.8 in /opt/conda/lib/python3.6/site-packages (from
Collecting alembic>=0.7.7 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/60/1e/cabc75a189de0fbb2841d0975243e59bde8b
    100% || 1.1MB 12.5MB/s ta 0:00:01    87% |    | 952kB 21.2MB/s eta 0:00:01
  Installing build dependencies ... done
Collecting sortedcontainers>=1.4.4 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/13/f3/cf85f7c3a2dbd1a515d51e1f1676d971abe4
Collecting intervaltree>=2.1.0 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/e8/f9/76237755b2020cd74549e98667210b2dd54d
Collecting lru-dict>=1.1.4 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/00/a5/32ed6e10246cd341ca8cc205acea5d208e40
Collecting empyrical>=0.4.2 (from zipline===1.2.0->-r requirements.txt (line 17))
  Downloading https://files.pythonhosted.org/packages/84/9e/9506e8b25464ff57ef93b5ba9092b464b44d
    100% || 51kB 6.6MB/s ta 0:00:01
Requirement already satisfied: pygments in /opt/conda/lib/python3.6/site-packages (from IPython>
Requirement already satisfied: traitlets>=4.2 in /opt/conda/lib/python3.6/site-packages (from IP
```

```
Requirement already satisfied: jedi>=0.10 in /opt/conda/lib/python3.6/site-packages (from IPytho
Requirement already satisfied: backcall in /opt/conda/lib/python3.6/site-packages (from IPython>
Requirement already satisfied: simplegeneric>0.8 in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: pickleshare in /opt/conda/lib/python3.6/site-packages (from IPyth
Requirement already satisfied: pexpect; sys_platform != "win32" in /opt/conda/lib/python3.6/site
Requirement already satisfied: prompt-toolkit<2.0.0,>=1.0.15 in /opt/conda/lib/python3.6/site-pa
Requirement already satisfied: future in /opt/conda/lib/python3.6/site-packages (from osqp->cvxp
Collecting dill>=0.3.1 (from multiprocess->cvxpy==1.0.3->-r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/c7/11/345f3173809cea7f1a193bfbf02403fff250
    100% || 153kB 15.5MB/s ta 0:00:01
Requirement already satisfied: ipython-genutils in /opt/conda/lib/python3.6/site-packages (from
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied: jupyter-core in /opt/conda/lib/python3.6/site-packages (from nbfo
Collecting requests-ftp (from pandas-datareader<0.6,>=0.2.1->zipline===1.2.0->-r requirements.tx
  Downloading https://files.pythonhosted.org/packages/3d/ca/14b2ad1e93b5195eeaf56b86b7ecfd5ea2d5
Collecting python-editor>=0.3 (from alembic>=0.7.7->zipline===1.2.0->-r requirements.txt (line 1
  Downloading https://files.pythonhosted.org/packages/c6/d3/201fc3abe391bbae6606e6f1d598c15d3670
Requirement already satisfied: ptyprocess>=0.5 in /opt/conda/lib/python3.6/site-packages (from p
Requirement already satisfied: wcwidth in /opt/conda/lib/python3.6/site-packages (from prompt-to
Building wheels for collected packages: alphalens, cvxpy, pandas, plotly, zipline, scs, multipro
  Running setup.py bdist_wheel for alphalens ... done
  Stored in directory: /root/.cache/pip/wheels/77/1e/9a/223b4c94d7f564f25d94b48ca5b9c53e3034016e
  Running setup.py bdist_wheel for cvxpy ... done
  Stored in directory: /root/.cache/pip/wheels/2b/60/0b/0c2596528665e21d698d6f84a3406c52044c7b4c
  Running setup.py bdist_wheel for pandas ... done
  Stored in directory: /root/.cache/pip/wheels/a3/08/c3/8fdd52954d4b415624cff43c6dd32a22bac90306
  Running setup.py bdist_wheel for plotly ... done
  Stored in directory: /root/.cache/pip/wheels/98/54/81/dd92d5b0858fac680cd7bdb8800eb26c001dd9f5
  Running setup.py bdist_wheel for zipline ... done
  Stored in directory: /root/.cache/pip/wheels/5d/20/7d/b48368c8634b1cb6cc7232833b2780a265d4217c
  Running setup.py bdist_wheel for scs ... done
  Stored in directory: /root/.cache/pip/wheels/df/d0/79/37ea880586da03c620ca9ecd5e42adbd86bc6ea8
  Running setup.py bdist_wheel for multiprocess ... done
  Stored in directory: /root/.cache/pip/wheels/96/20/ac/9f1d164f7d81787cd6f4401b1d05212807d021fb
  Running setup.py bdist_wheel for Logbook ... done
  Stored in directory: /root/.cache/pip/wheels/d2/70/07/68b99a8e05dcd1ab194a8e0ccb9e4d0ac5dd6d8d
  Running setup.py bdist_wheel for cyordereddict ... done
  Stored in directory: /root/.cache/pip/wheels/0b/9d/8b/5bf3e22c1edd59b50f11bb19dec9dfcfe5a479fc
  Running setup.py bdist_wheel for bottleneck ... done
  Stored in directory: /root/.cache/pip/wheels/97/a9/12/41b13e8b44889ab05ec4dcc91f27da21634bacf2
  Running setup.py bdist_wheel for bcolz ... done
  Stored in directory: /root/.cache/pip/wheels/c5/cc/1b/2cf1f88959af5d7f4d449b7fc6c9452d0ecbd86f
  Running setup.py bdist_wheel for alembic ... done
  Stored in directory: /root/.cache/pip/wheels/1f/04/83/76023f7a4c14688c0b5c2682a96392cfdd3ee444
  Running setup.py bdist_wheel for intervaltree ... done
  Stored in directory: /root/.cache/pip/wheels/08/99/c0/5a5942f5b9567c59c14aac76f95a70bf11dccc71
  Running setup.py bdist_wheel for lru-dict ... done
  Stored in directory: /root/.cache/pip/wheels/b7/ef/06/fbdd555907a7d438fb33e4c8675f771ff1cf4191
```

```
  Running setup.py bdist_wheel for empyrical ... done
  Stored in directory: /root/.cache/pip/wheels/10/a4/3b/951bd609878a82fd72b9ea23699daf1eaada4ff6
  Running setup.py bdist_wheel for dill ... done
  Stored in directory: /root/.cache/pip/wheels/59/b1/91/f02e76c732915c4015ab4010f3015469866c1eb9
  Running setup.py bdist_wheel for requests-ftp ... done
  Stored in directory: /root/.cache/pip/wheels/2a/98/32/37195e45a3392a73d9f65c488cbea30fe5bad76a
Successfully built alphalens cvxpy pandas plotly zipline scs multiprocess Logbook cyordereddict
tensorflow 1.3.0 requires tensorflow-tensorboard<0.2.0,>=0.1.0, which is not installed.
moviepy 0.2.3.2 has requirement tqdm==4.11.2, but you'll have tqdm 4.19.5 which is incompatible.
Installing collected packages: numpy, pandas, scipy, alphalens, osqp, ecos, scs, dill, multiproc
  Found existing installation: numpy 1.12.1
    Uninstalling numpy-1.12.1:
      Successfully uninstalled numpy-1.12.1
  Found existing installation: pandas 0.23.3
    Uninstalling pandas-0.23.3:
      Successfully uninstalled pandas-0.23.3
  Found existing installation: scipy 1.2.1
    Uninstalling scipy-1.2.1:
      Successfully uninstalled scipy-1.2.1
  Found existing installation: dill 0.2.7.1
    Uninstalling dill-0.2.7.1:
      Successfully uninstalled dill-0.2.7.1
  Found existing installation: plotly 2.0.15
    Uninstalling plotly-2.0.15:
      Successfully uninstalled plotly-2.0.15
  Found existing installation: tqdm 4.11.2
    Uninstalling tqdm-4.11.2:
      Successfully uninstalled tqdm-4.11.2
Successfully installed Logbook-1.5.3 alembic-1.4.2 alphalens-0.3.2 bcolz-0.12.1 bottleneck-1.3.2
```

```
In [2]: ! cat requirements.txt
```

```
alphalens==0.3.2
colour==0.1.5
cvxpy==1.0.3
cycler==0.10.0
numpy==1.13.3
pandas==0.18.1
plotly==2.2.3
pyparsing==2.2.0
python-dateutil==2.6.1
pytz==2017.3
requests==2.18.4
scipy==1.0.0
scikit-learn==0.19.1
six==1.11.0
tables==3.3.0
```

```
tqdm==4.19.5
zipline===1.2.0
```

1.2.2 Update Packages

Python packages get updated with time. If a compilation issue arises in the below cells, we recommend you to **upgrade** the specific package as:

`!{sys.executable} -m pip install <package-name> --upgrade`

A sample upgrade is done in next cell for `tensorflow` and `numpy` packages.

```
In [3]: !{sys.executable} -m pip install tensorflow --upgrade
        !{sys.executable} -m pip install numpy --upgrade

Collecting tensorflow
  Downloading https://files.pythonhosted.org/packages/de/f0/96fb2e0412ae9692dbf400e5b04432885f67
    100% || 109.2MB 436kB/s eta 0:00:01 7% |                                | 8.7MB 29.8MB/s eta 0:
Collecting google-pasta>=0.1.6 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/a3/de/c648ef6835192e6e2cc03f40b19eeda4382c
    100% || 61kB 8.0MB/s ta 0:00:01
Collecting tensorflow-estimator<1.15.0rc0,>=1.14.0rc0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/3c/d5/21860a5b11caf0678fbc8319341b0ae21a07
    100% || 491kB 17.5MB/s ta 0:00:01
Collecting wrapt>=1.11.1 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/82/f7/e43cefbe88c5fd371f4cf0cf5eb3feccd075
Collecting absl-py>=0.7.0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/1a/53/9243c600e047bd4c3df9e69cfabc1e8004a8
    100% || 112kB 14.3MB/s ta 0:00:01
Collecting termcolor>=1.1.0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/8a/48/a76be51647d0eb9f10e2a4511bf3ffb8cc1e
Collecting keras-applications>=1.0.6 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/71/e3/19762fdfc62877ae9102edf6342d71b28fbf
    100% || 51kB 9.6MB/s eta 0:00:01
Collecting grpcio>=1.8.6 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/cf/7a/9744998129fce7e29c5f2d8b0f545913b738
    100% || 19.5MB 1.9MB/s eta 0:00:01   11% |                              | 2.2MB 28.7MB/s eta 0
Collecting protobuf>=3.6.1 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/57/02/5432412c162989260fab61fa65e0a490c187
    100% || 1.3MB 13.2MB/s ta 0:00:01    39% |                   | 501kB 23.2MB/s eta 0:00:01
Collecting numpy<2.0,>=1.14.5 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/07/08/a549ba8b061005bb629b76adc000f3caaaf8
    100% || 20.2MB 1.4MB/s eta 0:00:01  2% |                                | 604kB 20.2MB/s eta
Collecting tensorboard<1.15.0,>=1.14.0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/91/2d/2ed263449a078cd9c8a9ba50ebd50123adf1
    100% || 3.2MB 11.1MB/s ta 0:00:01    27% |                        | 870kB 35.1MB/s eta 0:00:0
Collecting astor>=0.6.0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/c3/88/97eef84f48fa04fbd6750e62dcceafba6c63
Requirement already satisfied, skipping upgrade: wheel>=0.26 in /opt/conda/lib/python3.6/site-pa
```

```
Collecting gast>=0.2.0 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/d6/84/759f5dd23fec8ba71952d97bcc7e2c9d7d63
Requirement already satisfied, skipping upgrade: six>=1.10.0 in /opt/conda/lib/python3.6/site-pa
Collecting keras-preprocessing>=1.0.5 (from tensorflow)
  Downloading https://files.pythonhosted.org/packages/28/6a/8c1f62c37212d9fc441a7e26736df51ce6f0
    100% || 51kB 7.6MB/s ta 0:00:011
Requirement already satisfied, skipping upgrade: h5py in /opt/conda/lib/python3.6/site-packages
Requirement already satisfied, skipping upgrade: setuptools in /opt/conda/lib/python3.6/site-pac
Requirement already satisfied, skipping upgrade: werkzeug>=0.11.15 in /opt/conda/lib/python3.6/s
Requirement already satisfied, skipping upgrade: markdown>=2.6.8 in /opt/conda/lib/python3.6/sit
Building wheels for collected packages: wrapt, absl-py, termcolor, grpcio
  Running setup.py bdist_wheel for wrapt ... done
  Stored in directory: /root/.cache/pip/wheels/b1/c2/ed/d62208260edbd3fa7156545c00ef966f45f2063d
  Running setup.py bdist_wheel for absl-py ... done
  Stored in directory: /root/.cache/pip/wheels/8e/28/49/fad4e7f0b9a1227708cbbee4487ac8558a733484
  Running setup.py bdist_wheel for termcolor ... done
  Stored in directory: /root/.cache/pip/wheels/7c/06/54/bc84598ba1daf8f970247f550b175aaaee85f68b
  Running setup.py bdist_wheel for grpcio ... done
  Stored in directory: /root/.cache/pip/wheels/00/4d/5f/07d0d4283911d2b917b867a11b1622d9d2cc8c28
Successfully built wrapt absl-py termcolor grpcio
moviepy 0.2.3.2 has requirement tqdm==4.11.2, but you'll have tqdm 4.19.5 which is incompatible.
tensorboard 1.14.0 has requirement setuptools>=41.0.0, but you'll have setuptools 38.4.0 which i
Installing collected packages: google-pasta, tensorflow-estimator, wrapt, absl-py, termcolor, nu
  Found existing installation: numpy 1.13.3
    Uninstalling numpy-1.13.3:
      Successfully uninstalled numpy-1.13.3
  Found existing installation: protobuf 3.5.1
    Uninstalling protobuf-3.5.1:
      Successfully uninstalled protobuf-3.5.1
  Found existing installation: tensorflow 1.3.0
    Uninstalling tensorflow-1.3.0:
      Successfully uninstalled tensorflow-1.3.0
Successfully installed absl-py-0.9.0 astor-0.8.1 gast-0.3.3 google-pasta-0.2.0 grpcio-1.28.1 ker
Requirement already up-to-date: numpy in /opt/conda/lib/python3.6/site-packages (1.18.2)
```

1.2.3 Load Packages

In [4]:
```
# upgrade versions
!sudo pip3 install numpy==1.18
#!{sys.executable} -m pip3 install pandas --upgrade
!sudo pip3 install pandas==0.19.2
#!{sys.executable} -m pip install pandas --upgrade
print('Can you please help me figure out why this error pops up?')
print('Regards')

import cvxpy as cvx
import numpy as np
```

```
        import pandas as pd
        import time
        import project_tests
        import project_helper

        import matplotlib.pyplot as plt
        %matplotlib inline
        plt.style.use('ggplot')
        plt.rcParams['figure.figsize'] = (14, 8)
```

```
Collecting numpy==1.18
  Downloading https://files.pythonhosted.org/packages/eb/ec/d4b7855249ce87ece79783562dd6101b1f0a
    100% || 19.9MB 61kB/s  eta 0:00:01   14% |                          | 3.0MB 25.3MB/s eta 0:
Installing collected packages: numpy
Successfully installed numpy-1.18.0
You are using pip version 8.1.1, however version 20.0.2 is available.You should consider upgradi
Collecting pandas==0.19.2
  Downloading https://files.pythonhosted.org/packages/d8/7f/69354c4ab712acbdf37ece1b9dabdc1bbdb9
    100% || 18.5MB 66kB/s  eta 0:00:01  2% |                                | 378kB 10.2MB/s eta
Collecting pytz>=2011k (from pandas==0.19.2)
  Downloading https://files.pythonhosted.org/packages/e7/f9/f0b53f88060247251bf481fa6ea62cd0d25b
    100% || 512kB 2.6MB/s eta 0:00:01
Requirement already satisfied (use --upgrade to upgrade): numpy>=1.7.0 in /usr/local/lib/python3
Collecting python-dateutil>=2 (from pandas==0.19.2)
  Downloading https://files.pythonhosted.org/packages/d4/70/d60450c3dd48ef87586924207ae8907090de
    100% || 235kB 4.9MB/s eta 0:00:01
Collecting six>=1.5 (from python-dateutil>=2->pandas==0.19.2)
  Downloading https://files.pythonhosted.org/packages/65/eb/1f97cb97bfc2390a276969c6fae16075da28
Installing collected packages: pytz, six, python-dateutil, pandas
Successfully installed pandas-0.19.2 python-dateutil-2.8.1 pytz-2019.3 six-1.14.0
You are using pip version 8.1.1, however version 20.0.2 is available.You should consider upgradi
Can you please help me figure out why this error pops up?
Regards
```

```
        ---------------------------------------------------------------------------

        ImportError                               Traceback (most recent call last)

        <ipython-input-4-ec9367b24b56> in <module>()
          9 import cvxpy as cvx
         10 import numpy as np
    ---> 11 import pandas as pd
         12 import time
         13 import project_tests
```

```
/opt/conda/lib/python3.6/site-packages/pandas/__init__.py in <module>()
 54 # define the testing framework
 55 import pandas.util.testing
---> 56 from pandas.util.nosetester import NoseTester
 57 test = NoseTester().test
 58 del NoseTester


/opt/conda/lib/python3.6/site-packages/pandas/util/nosetester.py in <module>()
 11 import warnings
 12 from pandas.compat import string_types
---> 13 from numpy.testing import nosetester
 14
 15


ImportError: cannot import name 'nosetester'
```

1.3 Data Bundle

We'll be using Zipline to handle our data. We've created a end of day data bundle for this project. Run the cell below to register this data bundle in zipline.

```
In [ ]: import os
        import project_helper
        from zipline.data import bundles

        os.environ['ZIPLINE_ROOT'] = os.path.join(os.getcwd(), '..', '..', 'data', 'project_4_eo

        ingest_func = bundles.csvdir.csvdir_equities(['daily'], project_helper.EOD_BUNDLE_NAME)
        bundles.register(project_helper.EOD_BUNDLE_NAME, ingest_func)

        print('Data Registered')
```

1.4 Build Pipeline Engine

We'll be using Zipline's pipeline package to access our data for this project. To use it, we must build a pipeline engine. Run the cell below to build the engine.

```
In [ ]: from zipline.pipeline import Pipeline
        from zipline.pipeline.factors import AverageDollarVolume
        from zipline.utils.calendars import get_calendar


        universe = AverageDollarVolume(window_length=120).top(500)
        trading_calendar = get_calendar('NYSE')
        bundle_data = bundles.load(project_helper.EOD_BUNDLE_NAME)
        engine = project_helper.build_pipeline_engine(bundle_data, trading_calendar)
```

1.4.1 View Data

With the pipeline engine built, let's get the stocks at the end of the period in the universe we're using. We'll use these tickers to generate the returns data for the our risk model.

```
In [ ]: universe_end_date = pd.Timestamp('2016-01-05', tz='UTC')

        universe_tickers = engine\
            .run_pipeline(
                Pipeline(screen=universe),
                universe_end_date,
                universe_end_date)\
            .index.get_level_values(1)\
            .values.tolist()

        universe_tickers
```

1.5 Get Returns

Not that we have our pipeline built, let's access the returns data. We'll start by building a data portal.

```
In [ ]: from zipline.data.data_portal import DataPortal


        data_portal = DataPortal(
            bundle_data.asset_finder,
            trading_calendar=trading_calendar,
            first_trading_day=bundle_data.equity_daily_bar_reader.first_trading_day,
            equity_minute_reader=None,
            equity_daily_reader=bundle_data.equity_daily_bar_reader,
            adjustment_reader=bundle_data.adjustment_reader)
```

To make the code easier to read, we've built the helper function `get_pricing` to get the pricing from the data portal.

```
In [ ]: def get_pricing(data_portal, trading_calendar, assets, start_date, end_date, field='clos
            end_dt = pd.Timestamp(end_date.strftime('%Y-%m-%d'), tz='UTC', offset='C')
            start_dt = pd.Timestamp(start_date.strftime('%Y-%m-%d'), tz='UTC', offset='C')

            end_loc = trading_calendar.closes.index.get_loc(end_dt)
            start_loc = trading_calendar.closes.index.get_loc(start_dt)

            return data_portal.get_history_window(
                assets=assets,
                end_dt=end_dt,
                bar_count=end_loc - start_loc,
                frequency='1d',
                field=field,
                data_frequency='daily')
```

1.5.1 View Data

Let's get returns data for our risk model using the `get_pricing` function. For this model, we'll be looking back to 5 years of data.

```
In [ ]: five_year_returns = \
            get_pricing(
                data_portal,
                trading_calendar,
                universe_tickers,
                universe_end_date - pd.DateOffset(years=5),
                universe_end_date)\
            .pct_change()[1:].fillna(0)

        five_year_returns
```

2 Statistical Risk Model

It's time to build the risk model. You'll be creating a statistical risk model using PCA. So, the first thing is building the PCA model. ## Fit PCA Implement `fit_pca` to fit a PCA model to the returns data

```
In [ ]: from sklearn.decomposition import PCA


        def fit_pca(returns, num_factor_exposures, svd_solver):
            """
            Fit PCA model with returns.

            Parameters
            ----------
            returns : DataFrame
                Returns for each ticker and date
            num_factor_exposures : int
                Number of factors for PCA
            svd_solver: str
                The solver to use for the PCA model

            Returns
            -------
            pca : PCA
                Model fit to returns
            """
            #TODO: Implement function
            pca = PCA(num_factor_exposures, svd_solver)
            return pca.fit(returns)


        project_tests.test_fit_pca(fit_pca)
```

2.0.1 View Data

Let's see what the model looks like. First, we'll look at the PCA components.

```
In [ ]: num_factor_exposures = 20
        pca = fit_pca(five_year_returns, num_factor_exposures, 'full')

        pca.components_
```

Let's also look at the PCA's percent of variance explained by each factor

```
In [ ]: plt.bar(np.arange(num_factor_exposures), pca.explained_variance_ratio_)
```

You will see that the first factor dominates. The precise definition of each factor in a latent model is unknown, however we can guess at the likely interpretation.

2.1 Factor Betas

Implement `factor_betas` to get the factor betas from the PCA model.

```
In [ ]: def factor_betas(pca, factor_beta_indices, factor_beta_columns):
            """
            Get the factor betas from the PCA model.

            Parameters
            ----------
            pca : PCA
                Model fit to returns
            factor_beta_indices : 1 dimensional Ndarray
                Factor beta indices
            factor_beta_columns : 1 dimensional Ndarray
                Factor beta columns

            Returns
            -------
            factor_betas : DataFrame
                Factor betas
            """
            assert len(factor_beta_indices.shape) == 1
            assert len(factor_beta_columns.shape) == 1

            #TODO: Implement function
            factor_betas = pd.DataFrame(pca.components_.T, factor_beta_indices, factor_beta_colu
            return factor_betas


        project_tests.test_factor_betas(factor_betas)
```

2.1.1 View Data

Let's view the factor betas from this model.

```
In [ ]: risk_model = {}
        risk_model['factor_betas'] = factor_betas(pca, five_year_returns.columns.values, np.aran

        risk_model['factor_betas']
```

2.2 Factor Returns

Implement `factor_returns` to get the factor returns from the PCA model using the returns data.

```
In [ ]: def factor_returns(pca, returns, factor_return_indices, factor_return_columns):
            """
            Get the factor returns from the PCA model.

            Parameters
            ----------
            pca : PCA
                Model fit to returns
            returns : DataFrame
                Returns for each ticker and date
            factor_return_indices : 1 dimensional Ndarray
                Factor return indices
            factor_return_columns : 1 dimensional Ndarray
                Factor return columns

            Returns
            -------
            factor_returns : DataFrame
                Factor returns
            """
            assert len(factor_return_indices.shape) == 1
            assert len(factor_return_columns.shape) == 1

            #TODO: Implement function
            factor_returns = pd.DataFrame(pca.transform(returns), factor_return_indices, factor_
            return factor_returns


        project_tests.test_factor_returns(factor_returns)
```

2.2.1 View Data

Let's see what these factor returns looks like over time.

```
In [ ]: risk_model['factor_returns'] = factor_returns(
            pca,
```

```
            five_year_returns,
            five_year_returns.index,
            np.arange(num_factor_exposures))

        risk_model['factor_returns'].cumsum().plot(legend=None)
```

2.3 Factor Covariance Matrix

Implement `factor_cov_matrix` to get the factor covariance matrix.

```
In [ ]: def factor_cov_matrix(factor_returns, ann_factor):
            """
            Get the factor covariance matrix

            Parameters
            ----------
            factor_returns : DataFrame
                Factor returns
            ann_factor : int
                Annualization factor

            Returns
            -------
            factor_cov_matrix : 2 dimensional Ndarray
                Factor covariance matrix
            """

            #TODO: Implement function
            factor_cov_matrix = np.diag(ann_factor*factor_returns.var(axis=0, ddof=1))
            return factor_cov_matrix


        project_tests.test_factor_cov_matrix(factor_cov_matrix)
```

2.3.1 View Data

```
In [ ]: ann_factor = 252
        risk_model['factor_cov_matrix'] = factor_cov_matrix(risk_model['factor_returns'], ann_fa

        risk_model['factor_cov_matrix']
```

2.4 Idiosyncratic Variance Matrix

Implement `idiosyncratic_var_matrix` to get the idiosyncratic variance matrix.

```
In [ ]: def idiosyncratic_var_matrix(returns, factor_returns, factor_betas, ann_factor):
            """
            Get the idiosyncratic variance matrix
```

```
        Parameters
        ----------
        returns : DataFrame
            Returns for each ticker and date
        factor_returns : DataFrame
            Factor returns
        factor_betas : DataFrame
            Factor betas
        ann_factor : int
            Annualization factor

        Returns
        -------
        idiosyncratic_var_matrix : DataFrame
            Idiosyncratic variance matrix
        """

        #TODO: Implement function
        market_returns = pd.DataFrame(np.dot(factor_returns, factor_betas.T), returns.index,
        residual = returns - market_returns
        idiosyncratic_var_matrix = pd.DataFrame(ann_factor*np.diag(np.var(residual)), return
        return idiosyncratic_var_matrix


    project_tests.test_idiosyncratic_var_matrix(idiosyncratic_var_matrix)
```

2.4.1 View Data

```
In [ ]: risk_model['idiosyncratic_var_matrix'] = idiosyncratic_var_matrix(five_year_returns, ris

    risk_model['idiosyncratic_var_matrix']
```

2.5 Idiosyncratic Variance Vector

Implement `idiosyncratic_var_vector` to get the idiosyncratic variance Vector.

```
In [ ]: def idiosyncratic_var_vector(returns, idiosyncratic_var_matrix):
            """
            Get the idiosyncratic variance vector

            Parameters
            ----------
            returns : DataFrame
                Returns for each ticker and date
            idiosyncratic_var_matrix : DataFrame
                Idiosyncratic variance matrix

            Returns
```