Please refer to the below questions, you are given 24 hours for the test:-

P1. Regression

Attached .csv file has 1000 rows and four columns (from the left, A, B, C, and D). Please build a good model which explains Column D by a function of Column A, Column B and Column C.

P2. Growth – Sales/Earning

*X(t)* is a time series. Growth function is to capture relative change(increase or decrease) of *X(t)* where *t* means this quarter and *t-1* mean last quarter.

* Sales is the amount of money that a company actually receives during a specific period, including discounts and deductions for returned merchandise. Sales are positive number in most cases with few exceptions.
* Earnings are the amount of profit that a company produces during a specific period, which is usually defined as a quarter (three calendar months) or a year. Earnings can be both positive and negative values.

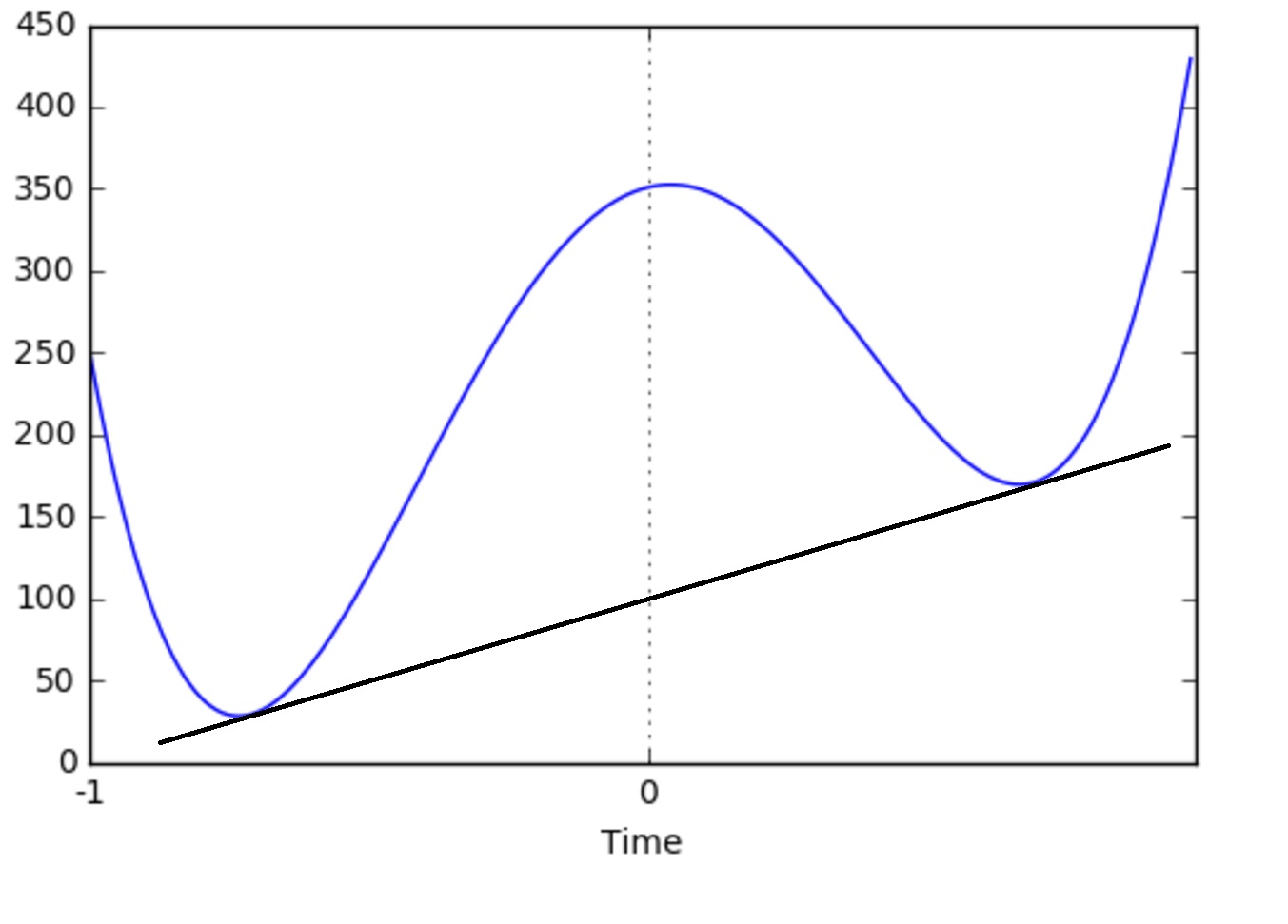
Considering above description, please improve below formula of Sales/Earning Growth to process *dirty* data values.

* Earnings Growth Formula: { Earnings(t)/Earnings(t-1) } – 1
* Sales Growth Formula: { Sales(t)/ Sales(t-1) } – 1

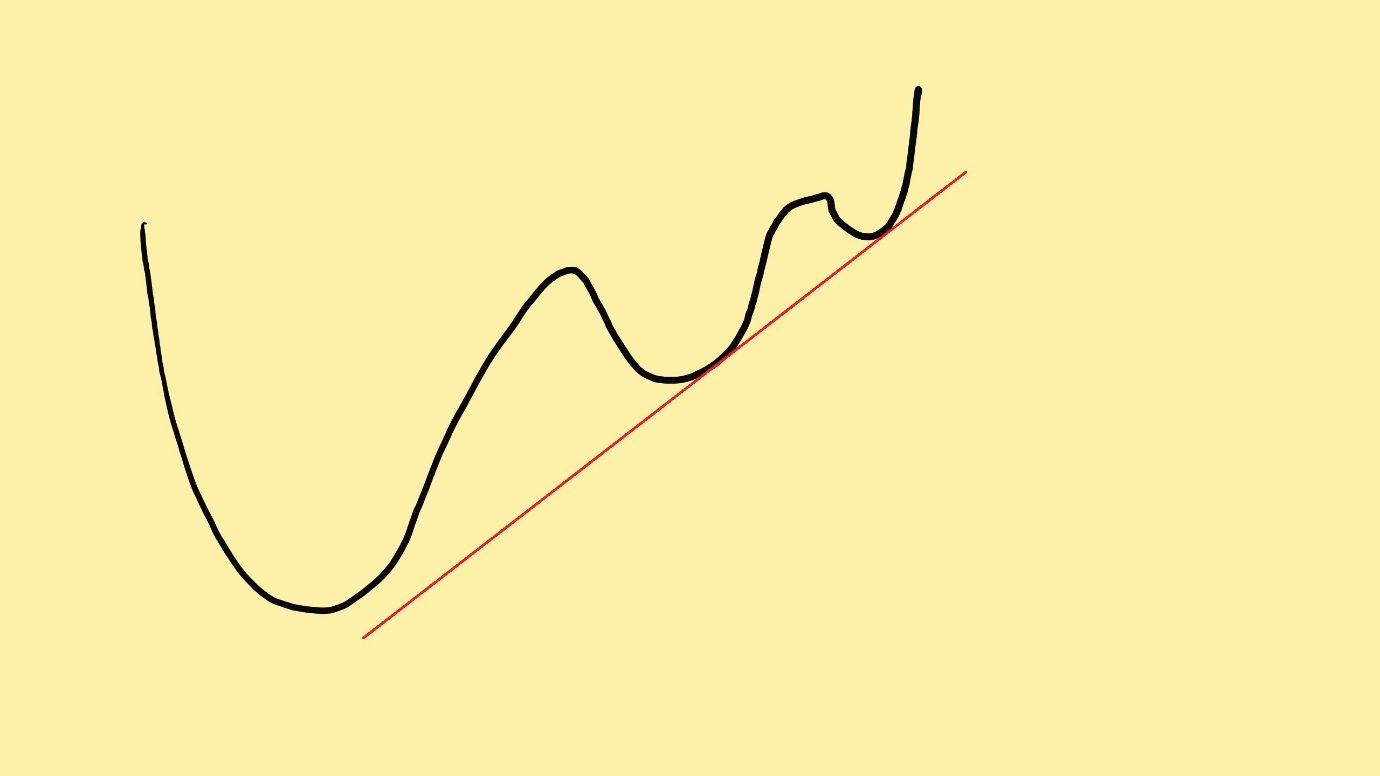
i.e. taking ratios between current quarter’s value and last quarter’s value and subtract by one.

P3. Technical indicator – trend-line

Below plot is daily price curve of 1,000 data-points of daily Price. In other word, Y-axis means Price and X-axis means day(t) – Price(t). For technical analysis, you need to draw a trend-line which defines a Supporting Line. This Supporting Line is a graph of linear function which is located below the daily price curve but intersects the daily price curve twice. Please note that intersections from below plots are not necessarily local minima.



Now please describe (or write a pseudo-code) how to find the linear equation of Supporting Line with t-variable.



M4. vector is a vector whose 2-norm is 1.0. For example, in 2-dimension, ( 1/sqrt(2), -1/sqrt(2) ) is a unit vector.

A. How many uncorrelated unit vectors exist in 3-dimension?

(“Two unit vectors are un-correlated” provided that the angle between two unit vectors is 45 degrees or larger. In the case of 2-dimension, there are eight such unit vectors.\_

B. What if “uncorrelated” means the angle is 60 degrees or larger?

C103. Sorting a number sequence - write a program to sort 1,000,000(one million) numbers, but you can use only 100,000( hundred thousand) memory cells. For this question only, you can use numpy to define arrays and generate random numbers.

a) Generate random 1,000,000 16 bit unsigned integers and save them in the file "hexDec.bin".  Please note that your have a memory limit(100,000).

b) Using an unsigned integer array of size 100,000, sort numbers in "hexDec.bin" in the ascending order and save the result in the file "hexDecSort.bin". Please note that your have a memory limit(100,000).

c) Then sort the file "hexDecSort.bin" using repeated occurrences of numbers. For instance 1,2,2,3,3,3,4,4,5 should be sorted into 3,3,3,2,2,4,4,1,5. The result should be saved in "hexDecSizeSort.bin" Please note that your have a memory limit(100,000).

C201. You are required to use Python 2.7 or C++ as your programming language. You may or may not automate all the process which the question is asking you to solve. If you think  a certain part takes longer time for you to code, you may skip that part and manually compete the task. Task is mainly on data science problem.

Please download daily price data of about 500 stocks in S&P 500 to calculate daily return series of 500 stock in S&P500. Now you can calculate correlation between 500 stocks from return series. This will be written as 500 x 500 matrix with ones along the diagonal. Let's define Corr(A,B) as the correlation coefficient between A and B derived from their return series. Then, build clusters satisfying the following conditions:

Correlation between any pair of stock in the same cluster must be higher than pair of stocks from different cluster. For example, A and B are from one cluster and C are for another cluster; Corr(A,B) >= Corr(A,C) and Corr(A,B) >= Corr(B,C)

Then, you are now supposed to find clusters defined above to optimize the following objective functions.

* Summation over Correlations between stocks in the same cluster is maximized. This sum of correlation is sum of all the elements in correlation matrix. I.e. if you constructed a cluster with 100 stocks. You can calculate 100 x 100 correlation matrix and sum up all the elements to get "sum of correlation". In other word, stocks in the same cluster must be highly correlated.
* The number of clusters is not restricted as long as you have more than four clusters. But standard deviation of the numbers of clusters is to be minimized. For example, clusters of 100 stocks, 100 stocks, 100 stocks and 200 stocks are better than those of 300 stocks, 160 stocks, 35 stocks and 5 stocks.
* You can calculate the average of returns of stocks in the same cluster. If you have four clusters, you will have four average returns, from which 4x4 correlation matrix can be calculated. Summation over Correlation  between average returns of clusters is minimized. In other word, clusters must not be correlated in terms of average return

Useful information

* You can find components of SP500 index from Wikipedia: <https://en.wikipedia.org/wiki/List_of_S%26P_500_companies>

|  |
| --- |
| [List of S&P 500 companies - Wikipedia](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)  en.wikipedia.org  The S&P 500 stock market index, maintained by S&P Dow Jones Indices, comprises 505 common stocks issued by 500 large-cap companies and traded on American stock ... |

|  |
| --- |
| [List of S&P 500 companies - Wikipedia](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)  en.wikipedia.org  The S&P 500 stock market index, maintained by S&P Dow Jones Indices, comprises 505 common stocks issued by 500 mid and large-cap companies and traded on American ... |

|  |
| --- |
| [List of S&P 500 companies](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)  en.wikipedia.org  The S&P 500 stock market index, maintained by S&P Dow Jones Indices, comprises 505 common stocks issued by 500 mid and large-cap companies and traded on... |

* You can download price data from Yahoo: <https://finance.yahoo.com/quote/MON/history?p=MON>

End