JPMQR MINI PROJECT 2

LA ASSIGNMENT

Submitted by-

Shirin Jain

IIT Guwahati

Task 1

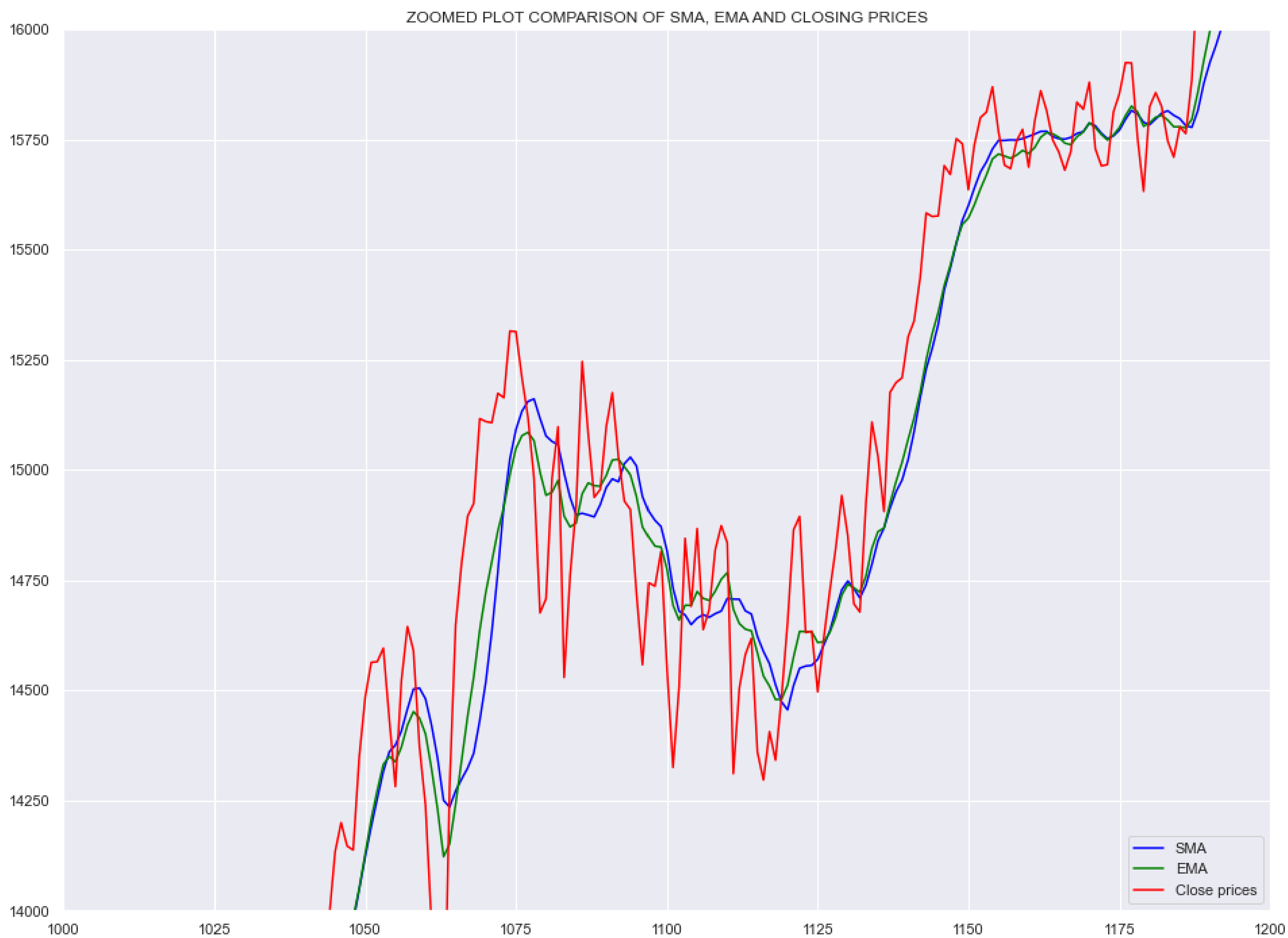
I computed the technical indicators SMA, EMA, MACD, Range, stochastic oscillators and MFI. I plotted each of these technical indicators using pyplot from matplotlib. After analysing the graph we could make some important inferences.

SMA :

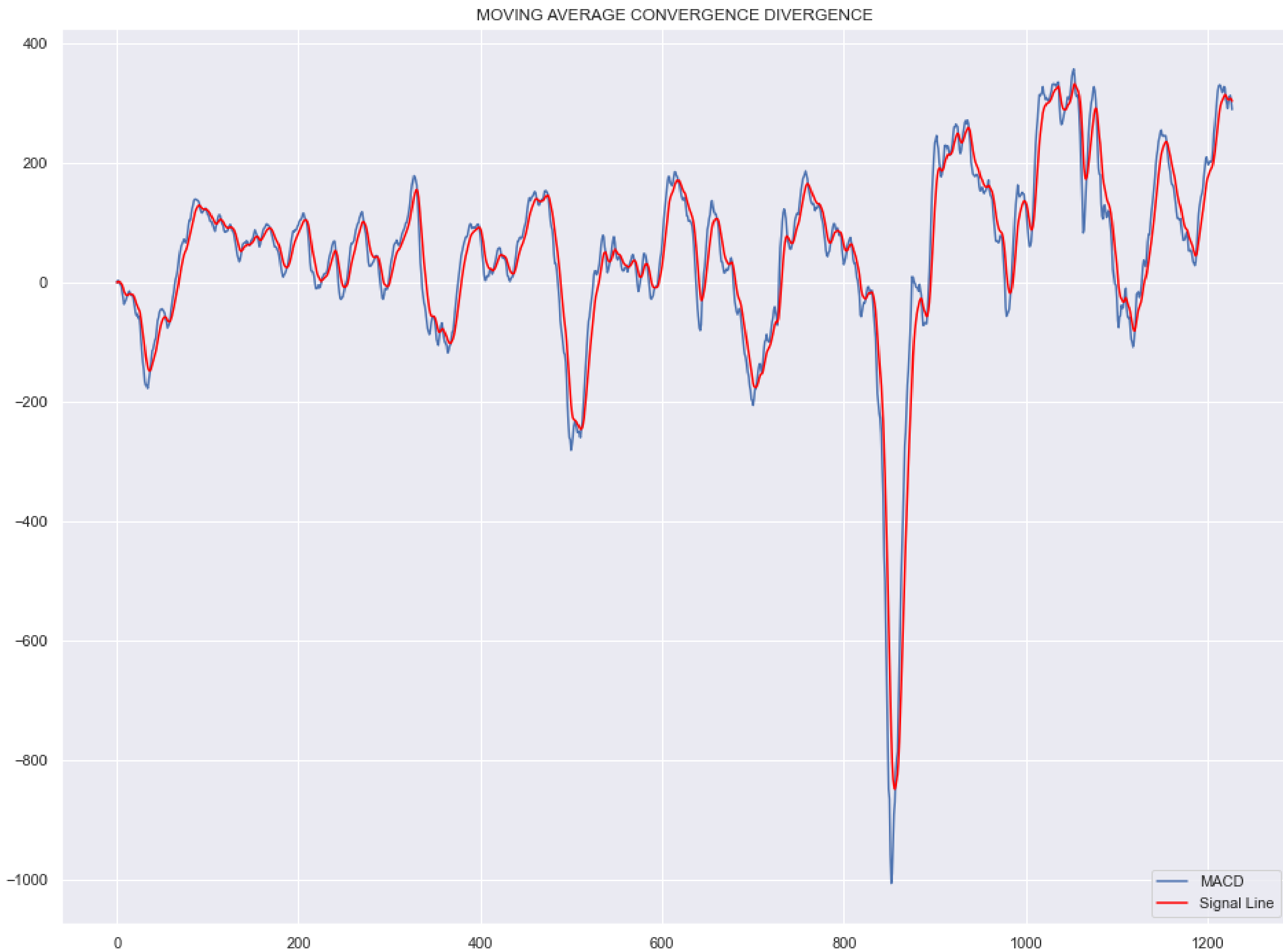
1. Moving averages is a calculation to analyze data points by calculating and creating a series of averages of different subsets from the entire data set. The simplest form of moving average is the simple moving average. It is the arithmetic mean of the given set of data points.
2. I plotted the Simple moving average and compared it with the closing prices.
3. Buying and selling opportunities can be predicted using the simple moving average computation.
4. When a stock's price trades below its average price, it indicates that traders are willing to sell the stock for a lower price. This indicates that traders are bearish on the stock market. As a result, one should consider selling options.
5. When a stock's price trades above its average price, it indicates that traders are willing to pay a greater price for the stock. This indicates that the traders believe the stock price will rise. As a result, one should consider purchasing opportunities.

EMA :

1. The exponential moving average is a type of weighted moving average that gives more weight to recent prices, giving lower importance to older market trends.
2. Buying and selling opportunities can be predicted using the simple moving average computation.
3. When the current market price exceeds the 10-day EMA, there are purchasing opportunities (go long).
4. Similarly for going short when the current market price is below the 10-day EMA.
5. In the zoom figure, we can clearly observe that indeed EMA responds faster to the change of trends and gets closer to them.

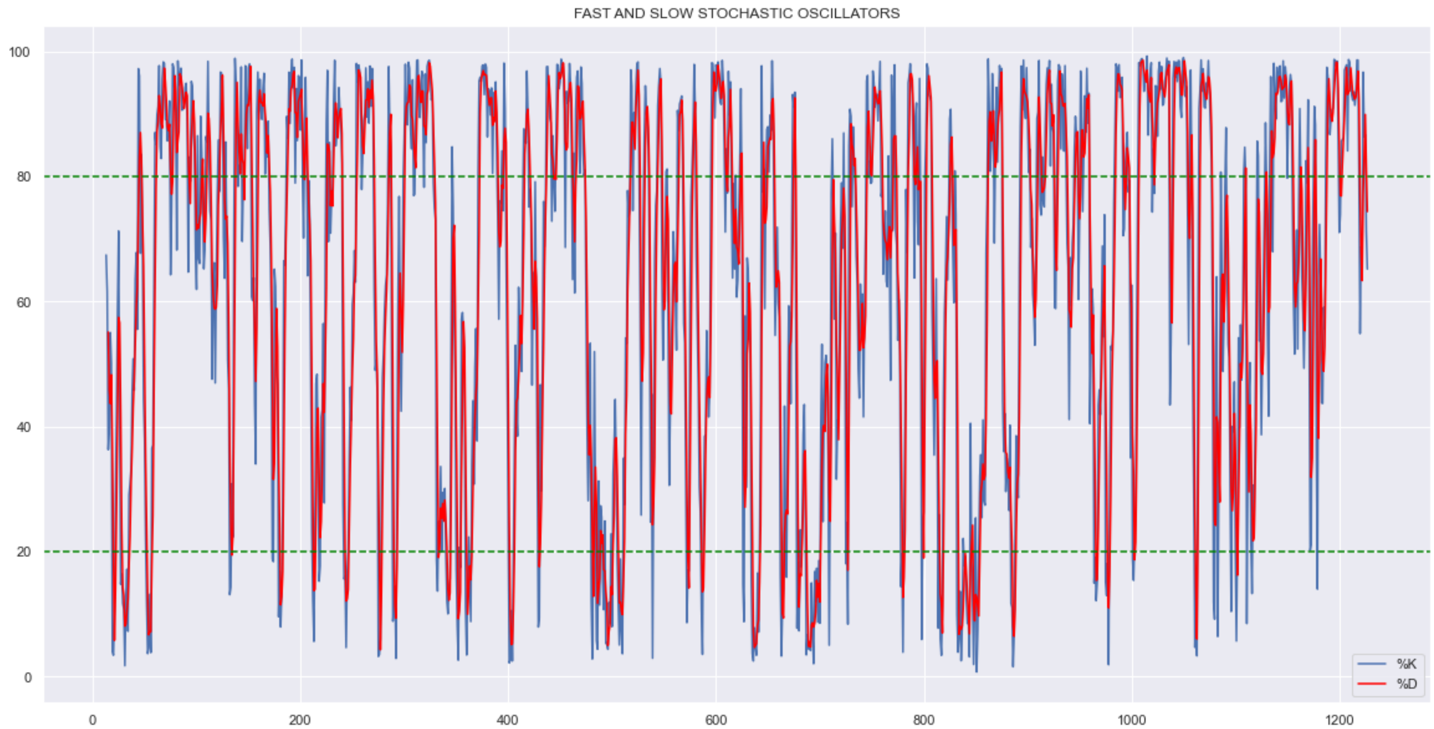


MACD :



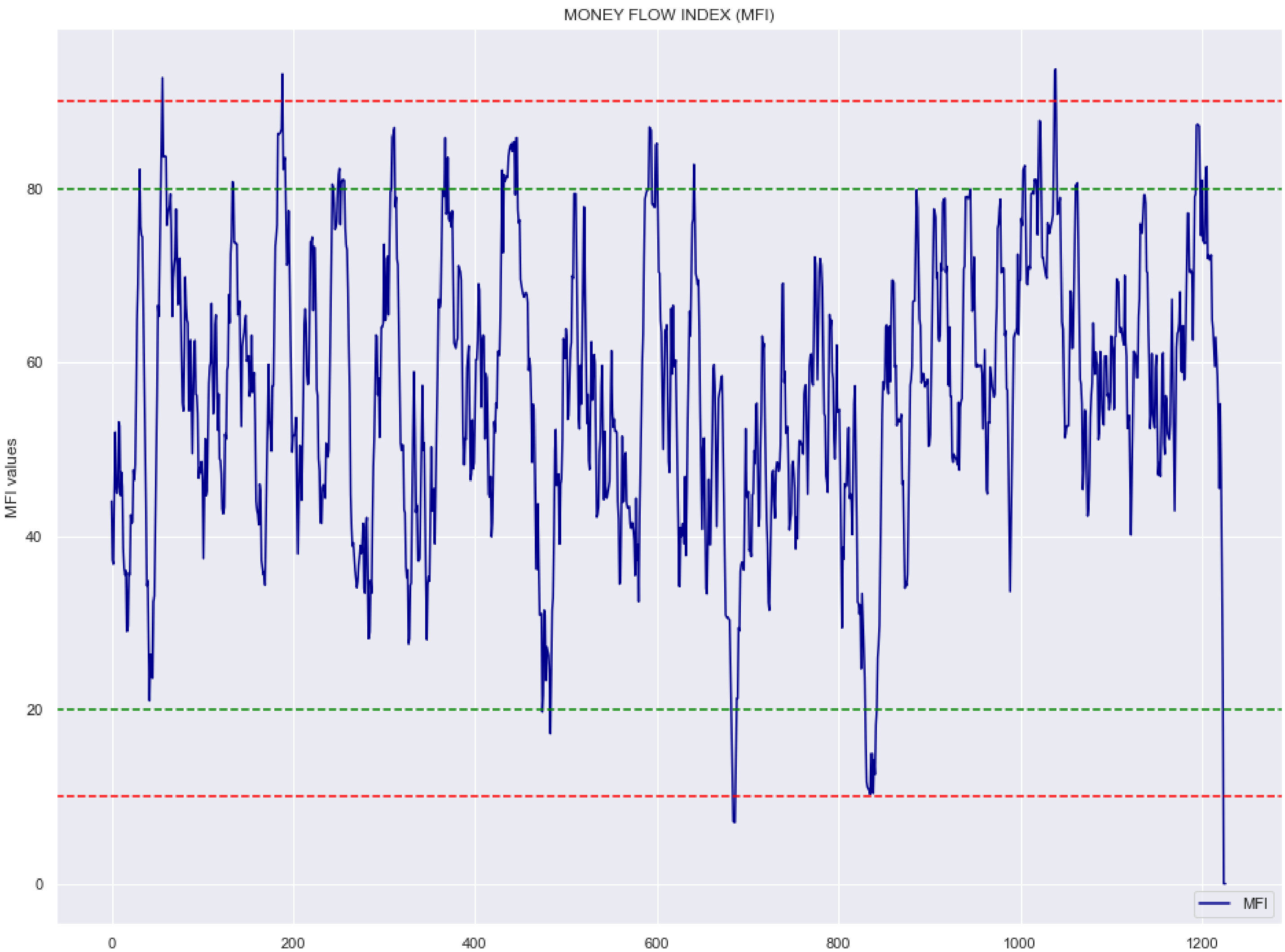
1. The moving average convergence divergence (MACD) is a trend-following momentum indicator that shows the relationship between two exponential moving averages. The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA.
2. The MACD is plotted alongside the signal line. We can see that MACD is above the signal line in some areas and below the signal line in others.
3. When the MACD crosses above its signal line after a minor decline in a longer-term advance, it is considered bullish confirmation, indicating that it is time to purchase.
4. It is a bearish indication when the MACD falls below the signal line, indicating that it may be time to sell.
5. When the MACD climbs above the signal line, the indicator offers a positive signal, indicating that the asset's price is expected to rise.

Stochastic Oscillators (%K and %D)



1. Stochastic Oscillator aims to detect trends in prices by comparing the closing price of the security with the highest price and lowest price achieved by the security during the last 14 days.
2. First a word of warning. Most advice from only using one indicator alone as a buy-sell signal. This also holds for the Stochastic Oscillator indicator. As the name suggests, it is only an indicator, not a predictor.
3. The indicator signals buy or sell when the two lines cross each other. If the %K is above the %D then it signals buy and when it crosses below, it signals sell.
4. Looking at the graph it makes a lot of signals (every time the two lines cross each other). This is a good reason to have other indicators to rely on.
5. An often misconception is that it should only be used when it is in the regions of 20% low or 80% high. But it is often that low and high can be for quite some time. Hence, selling if we reach the 80% high in this case, we would miss a great opportunity of a big gain.

Money Flow Index (MFI):



1. The Money Flow Index (MFI) is a technical indicator that uses both price and volume data to generate overbought and oversold indications.
2. Overbought is defined as a value above 80, and
3. oversold is defined as a reading below 20, though levels of 90 and 10 are frequently used as thresholds.

Task 2

In this task, we perform the Principal Component Analysis (PCA).

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

Step 1: I performed Standardization to standardize the range of the continuous initial variables so that each one of them contributes equally to the analysis.

More specifically, the reason why it is critical to perform standardization prior to PCA, is that the latter is quite sensitive regarding the variances of the initial variables. Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable.

$$z = \frac{value - mean}{standard\ deviation}$$

Once the standardization is done, all the variables will be transformed to the same scale.

Step 2: Next, I computed the covariance matrix to understand how the variables of the input data set are varying from the mean with respect to each other, or in other words, to see if there is any relationship between them. Because sometimes, variables are highly correlated in such a way that they contain redundant information. So, in order to identify these correlations, we compute the covariance matrix.

It's actually the sign of the covariance that matters :

->if positive then : the two variables increase or decrease together (correlated)

->if negative then : One increases when the other decreases (Inversely correlated)

The covariance matrix is not more than a table that summarizes the correlations between all the possible pairs of variables

Step 3: Next, I computed the eigenvalues and eigenvectors from the covariance matrix to identify the principal components.

Principal components are new variables that are constructed as linear combinations or mixtures of the initial variables. These combinations are done in such a way that the new variables (i.e., principal components) are uncorrelated and most of the information within the initial variables is squeezed or compressed into the first components. So, the idea is 10-dimensional data gives you 10 principal components, but PCA tries to put maximum possible information in the first component, then maximum remaining information in the second and so on.

Organizing information in principal components this way, will allow you to reduce dimensionality without losing much information, and this by discarding the components with low information and considering the remaining components as your new variables.

Geometrically speaking, principal components represent the directions of the data that explain a maximal amount of variance, that is to say, the lines that capture most information of the data. The relationship between variance and information here, is that, the larger the variance carried by a line, the larger the dispersion of the data points along it, and the larger the dispersion along a line, the more the information it has.

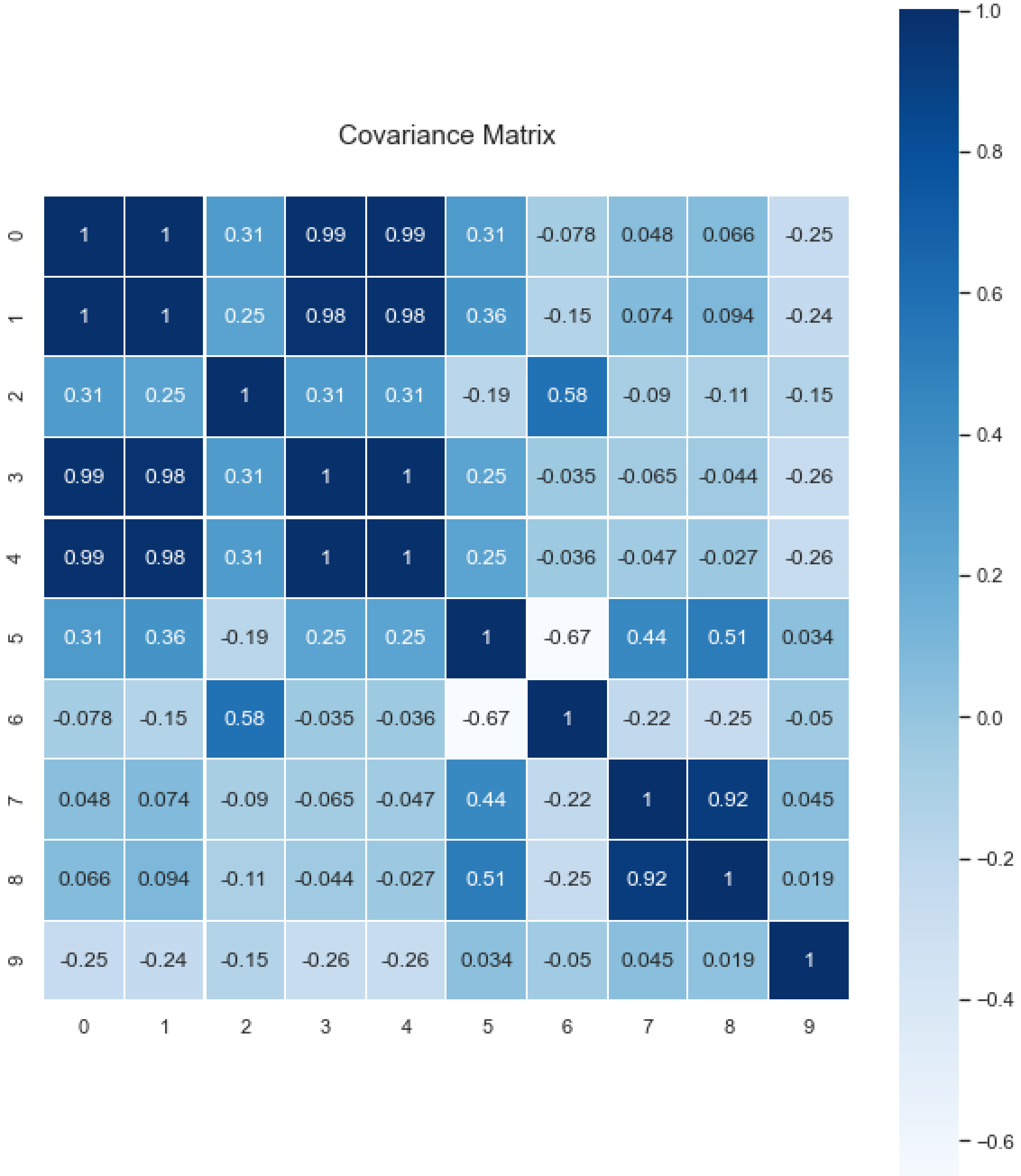
Principal components are constructed in such a manner that the first principal component accounts for the largest possible variance in the data set. This continues until a total of p principal components have been calculated, equal to the original number of variables.

Step 4: Computing the eigenvectors and ordering them by their eigenvalues in descending order, allow us to find the principal components in order of significance. In this step, I chose whether to keep all these components or discard those of lesser significance (of low eigenvalues), and form with the remaining ones a matrix of vectors that we call Feature vector.

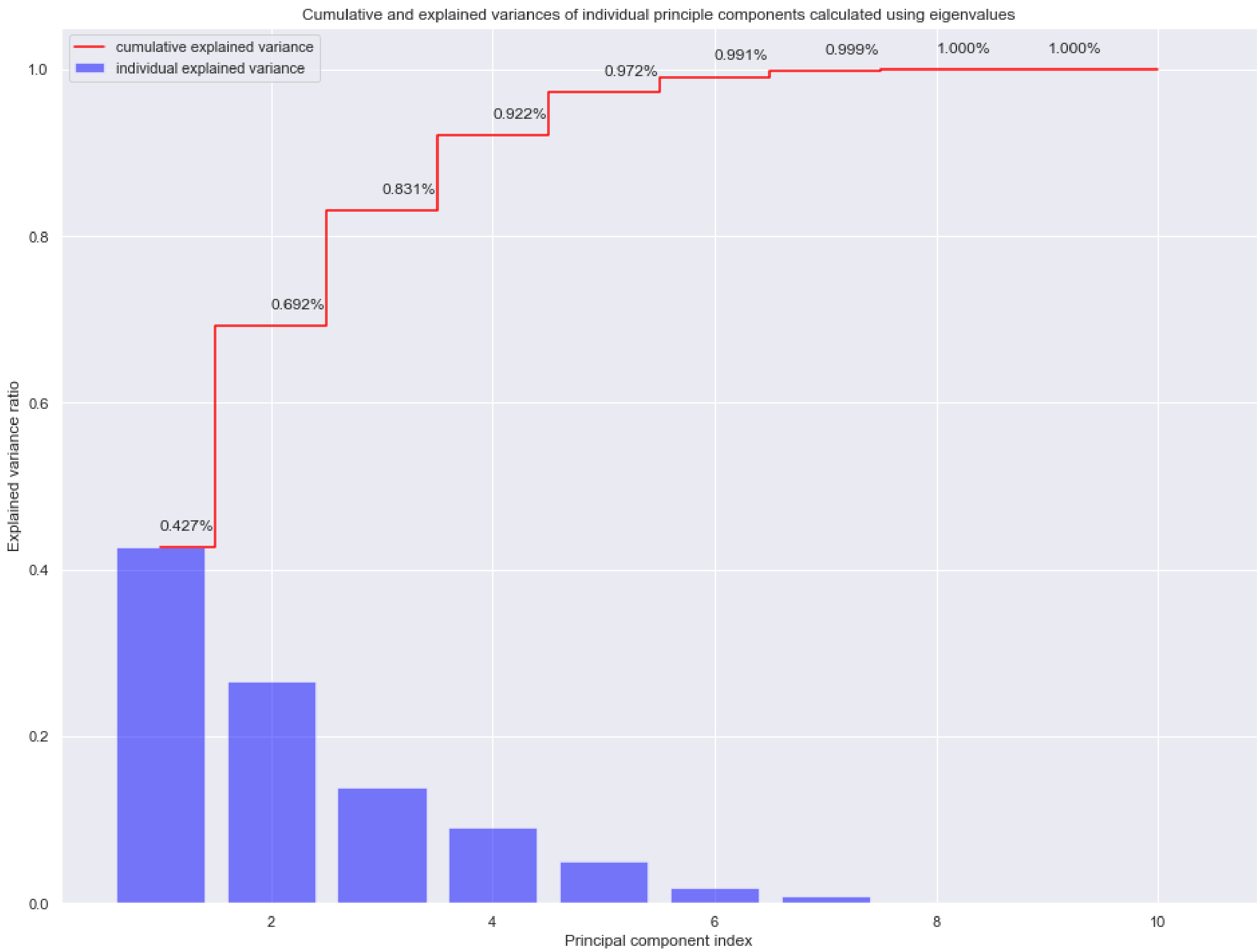
Step 5: I used the feature vector formed using the eigenvectors of the covariance matrix, to reorient the data from the original axes to the ones represented by the principal components. This can be done by multiplying the transpose of the original data set by the transpose of the feature vector.

Lastly, I also computed PCA using sklearn's inbuilt library and proved that it outputs the same result. Hence our principal component analysis is correct.

Covariance Matrix:



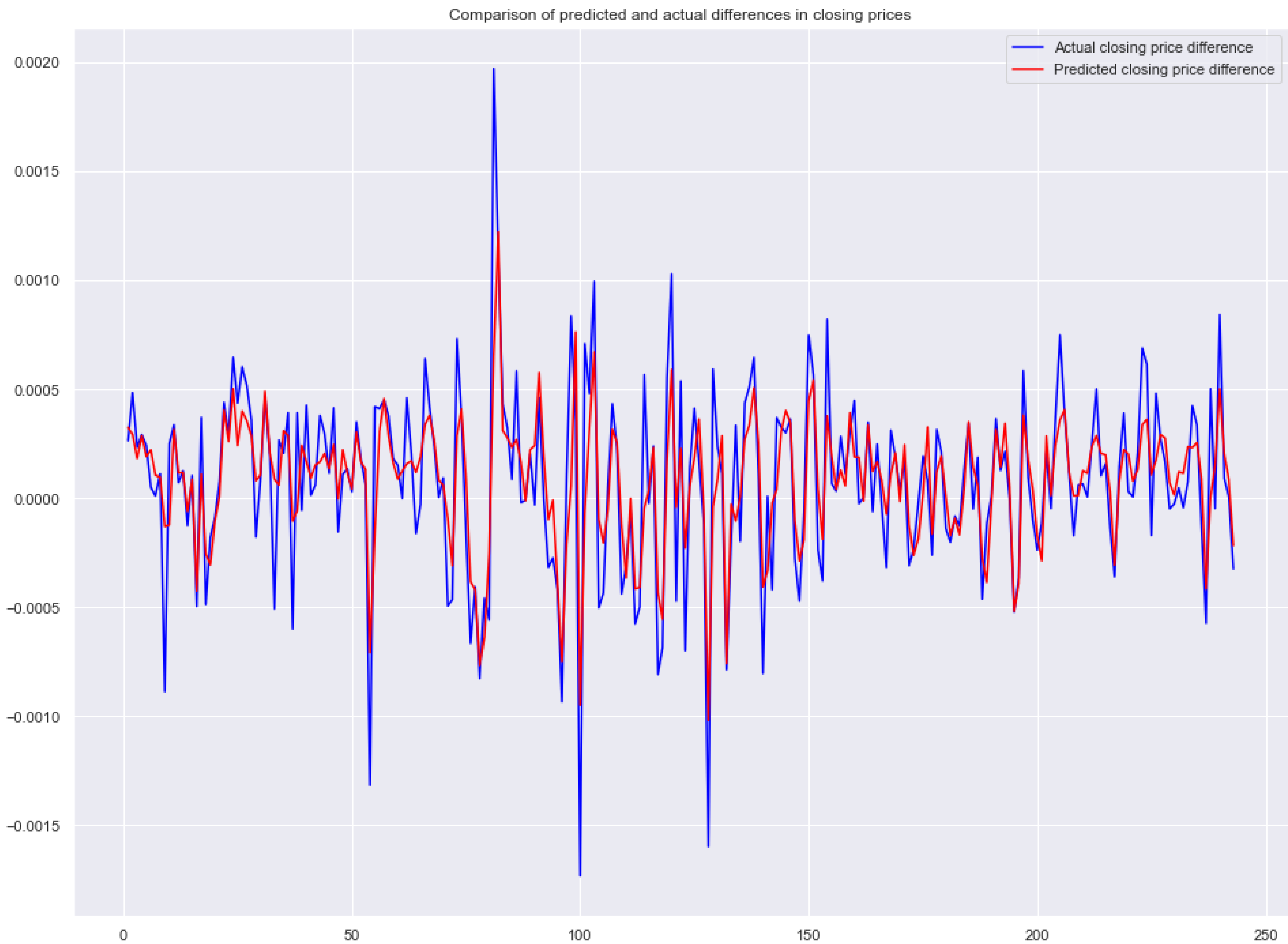
Explained Variances and Cumulative variances:



Task 3

Next we perform multivariate linear regression, with 95% variance (5 principal components) and 99% variance (6 principal components) with root mean squared error 0.02468 and 0.02464 respectively.

We plot the predicted price difference and actual price difference :


Comparison of predicted and actual differences in closing prices
Actual closing price difference
Predicted closing price difference
0.0020
0.0015
0.0010
0.0005
0.0000
−0.0005
−0.0010
−0.0015
0
50
100
150
200
250