Analysis of the U.S. Stock Market

by H. Quelle, C. Ryan, and T. Agarwal

Project Advisor: Dr. Melinda Lanius

The University of Arizona

Content

Outline for section 1

What is Principal Component Analysis?

- Principal Component Analysis (PCA) is a data reduction technique

What is Principal Component Analysis?

- Principal Component Analysis (PCA) is a data reduction technique
- Start with $A \in \mathbb{R}^{\mathrm{mxn}}$, with $m \gg n$

What is Principal Component Analysis?

- Principal Component Analysis (PCA) is a data reduction technique
- Start with $A \in \mathbb{R}^{\mathrm{mxn}}$, with $m \gg n$
- The data is reduced to a set of m eigenvectors

What is Principal Component Analysis?

- Principal Component Analysis (PCA) is a data reduction technique
- Start with $A \in \mathbb{R}^{mxn}$, with $m \gg n$
- The data is reduced to a set of m eigenvectors
- These vectors are called the principal components, and we can reduce the size of data using these

Outline for section 2

History

- The ideas of Principal Component Analysis are based in research done by Cauchy and Jacobi in the mid 19th century
- The first use in searching for items maximum variance was done by Harold Hotelling in 1933.
- Since then it has been used for just about anything that requires data reduction.
- It is used in research to reduce components, in genetics analysis, and in financial analysis.

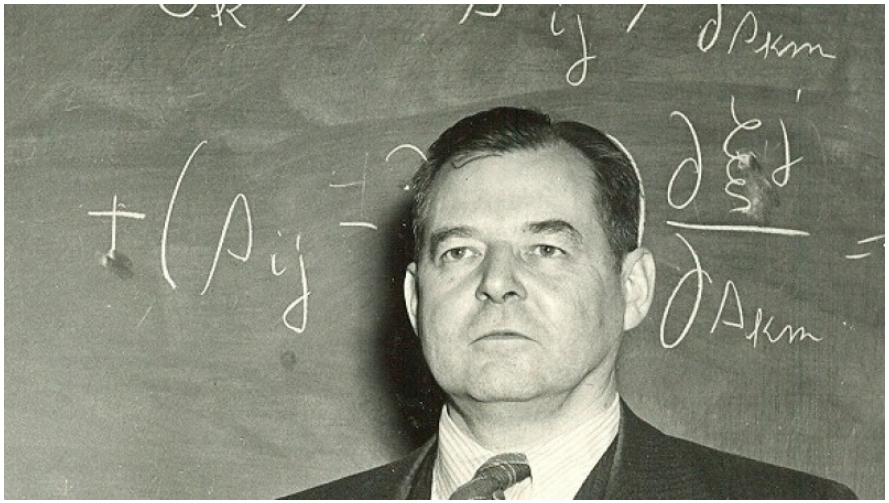


Figure: Irwin Collier (2018)

History

- The most recent use of PCA has been in facial recognition software
- For facial recognition there are eigenfaces instead of eigenvectors

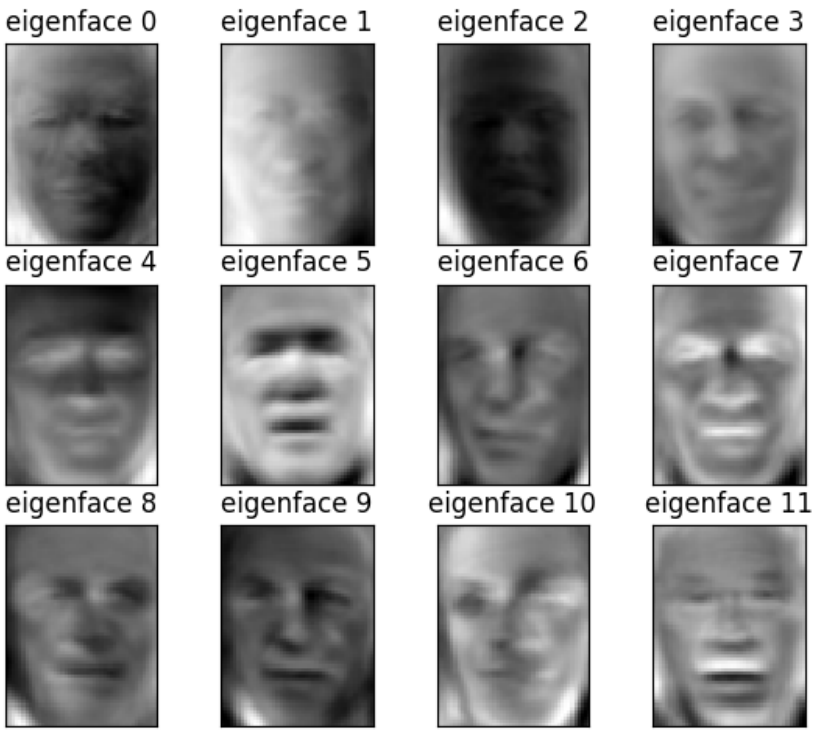


Figure: Eigenfaces

History

- Another example is PCA used for letter recognition.
- With this there are eigencharacters

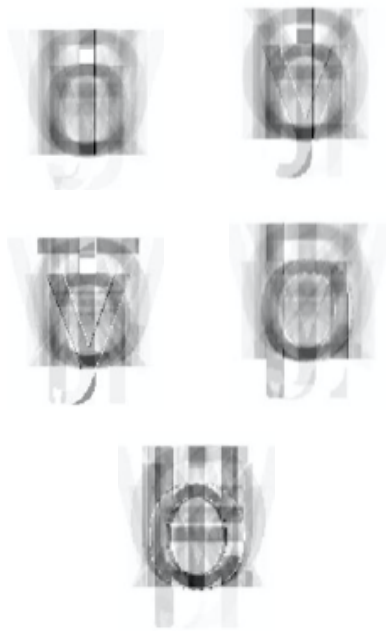


Figure: Eigencharacters

Outline for section 3

The Math

- Consider a data matrix

$$A = \begin{bmatrix} 2 & 6 \\ 7 & 9 \\ 4 & 2 \end{bmatrix} \tag{1}$$

The Math

- Consider a data matrix

$$A = \begin{bmatrix} 2 & 6 \\ 7 & 9 \\ 4 & 2 \end{bmatrix} \tag{1}$$

- Compute the average across the samples

$$\mu = \frac{1}{2} \begin{bmatrix} 2+6 \\ 7+9 \\ 4+2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix} \tag{2}$$

The Math

- Consider a data matrix

$$A = \begin{bmatrix} 2 & 6 \\ 7 & 9 \\ 4 & 2 \end{bmatrix} \tag{1}$$

- Compute the average across the samples

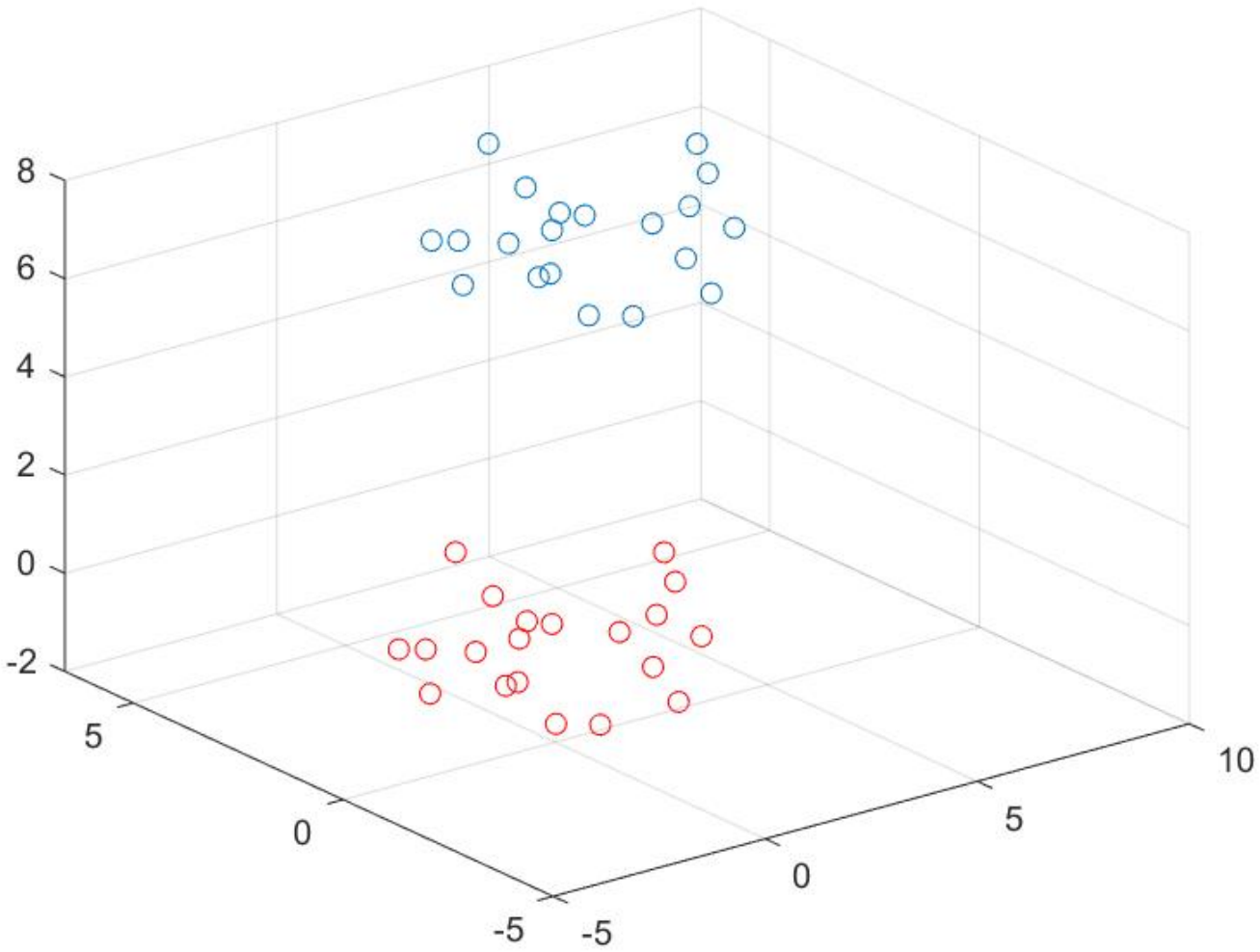
$$\mu = \frac{1}{2} \begin{bmatrix} 2+6 \\ 7+9 \\ 4+2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \\ 3 \end{bmatrix} \tag{2}$$

- Subtract the mean from each sample

$$\widehat{A} = \begin{bmatrix} -2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3}$$

The Math

What does this do?



The Math

- Recall the definitions of Variance and Covariance

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{ and } Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{4}$$

The Math

- Recall the definitions of Variance and Covariance

$$Var(X) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{ and } Cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{4}$$

- Calculate the Covariance matrix $C = \frac{1}{n-1}\widehat{A}\widehat{A}^T$

$$C = 1 \cdot \begin{bmatrix} -2 & 2 \\ -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -2 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 8 & 4 & -4 \\ 4 & 2 & -2 \\ -4 & -2 & 2 \end{bmatrix} \tag{5}$$

The Math

From C, we can find eigenvector and eigenvalue pairs $Cv = \lambda v$

$$\lambda_1 = 12 \text{ and } v_1 = \begin{bmatrix} -0.8165 \\ -0.4082 \\ 0.4082 \end{bmatrix} \tag{6}$$

The Math

From C, we can find eigenvector and eigenvalue pairs $Cv = \lambda v$

$$\lambda_1 = 12 \text{ and } v_1 = \begin{bmatrix} -0.8165 \\ -0.4082 \\ 0.4082 \end{bmatrix} \tag{6}$$

Projecting samples onto these eigenvectors creates a new, reduced version of the data

$$\begin{bmatrix} 2 \\ 7 \\ 4 \end{bmatrix} \approx 2.4495 v_1 + \mu = \begin{bmatrix} 2 \\ 7 \\ 4 \end{bmatrix} \tag{7}$$

Now only a few eigenvectors, the weights, and μ have to be kept to get the original data.

The Math

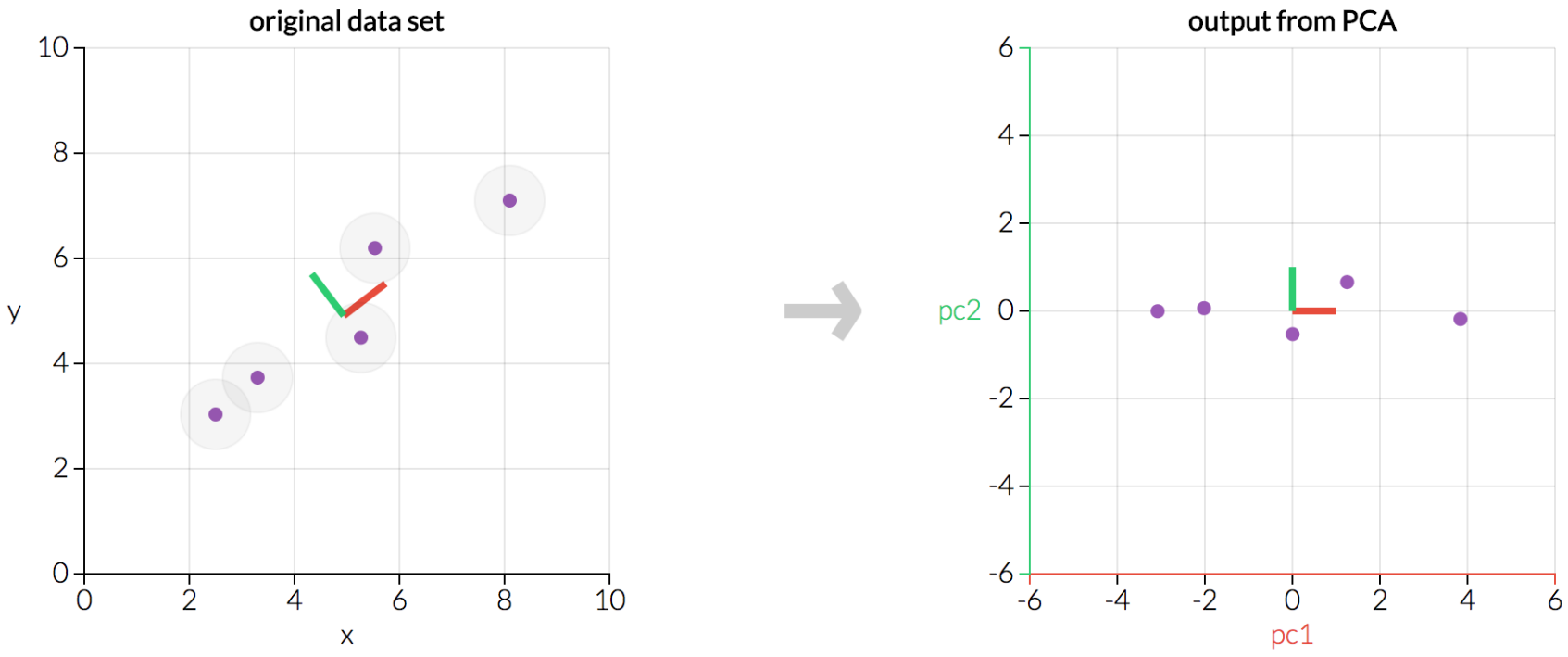


Figure: Matt Brems (2017), TowardsDataScience

The Math

What's special about the eigenvectors?

The Math

What's special about the eigenvectors?
The eigenvectors are only scaled by the matrix, not rotated.

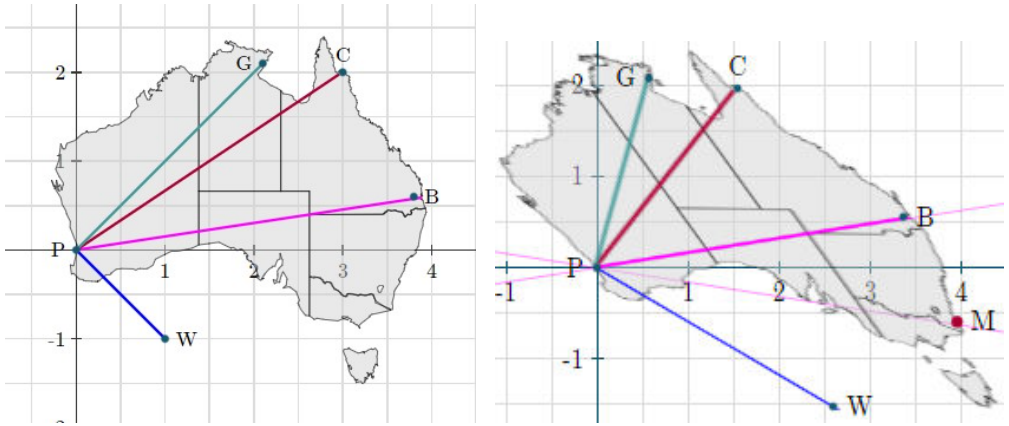


Figure: Murray Borne (Jan 2020), IntMath

The Math

What's special about the eigenvectors?
The eigenvectors are only scaled by the matrix, not rotated.

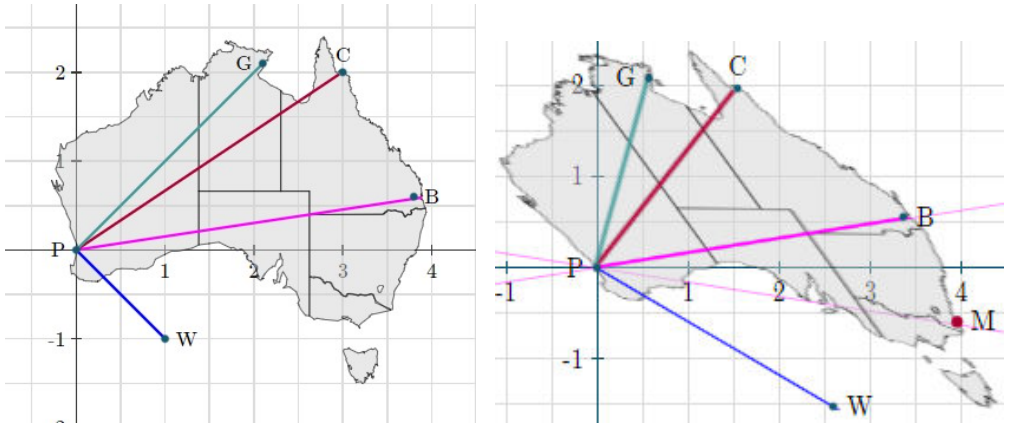


Figure: Murray Borne (Jan 2020), IntMath

Since the Covariance matrix only scales its eigenvectors, those eigenvectors must point in the direction of greatest variance in the data.

The Math

What's special about the eigenvectors?
The eigenvectors are only scaled by the matrix, not rotated.

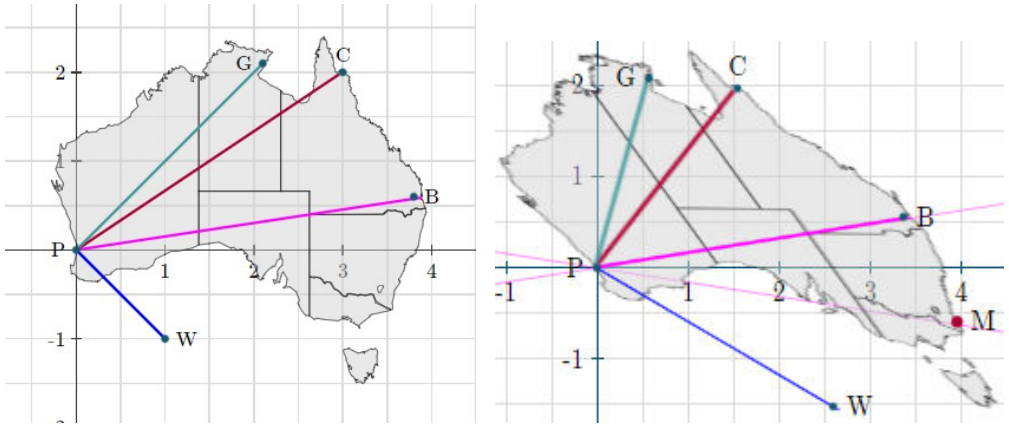


Figure: Murray Borne (Jan 2020), IntMath

Since the Covariance matrix only scales its eigenvectors, those eigenvectors must point in the direction of greatest variance in the data.
The eigenvector associated with the largest eigenvalue is the direction with the most variance in the data, the second eigenvector is the direction with the second most variance that is orthogonal to the first, and so on.

The Math

What's special about the weights?

The Math

What's special about the weights?
The magnitude of the weight tells us how much our original sample depends on the eigenvector v_i. If the magnitude is large, then much of the sample's variance is accounted for in eigenvector v_i. If the magnitude is small, then that v_i can be ignored when examining the sample.

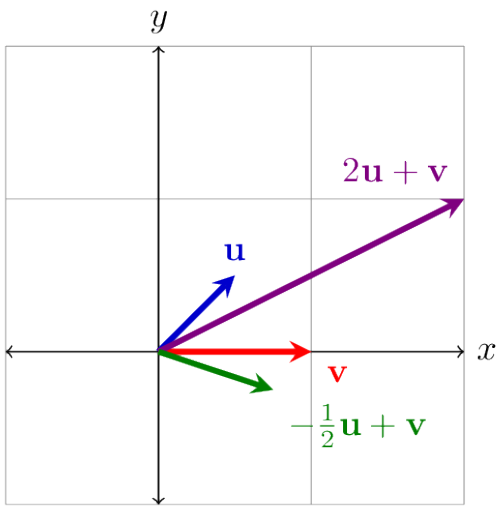


Figure 1: Vector combinations.

Figure: Jonathan Daugherty (2013)

Outline for section 4

Project Goals

1. We plan to reduce a very large set of stock data using PCA to something more predictable
2. Construct stocks using "eigenstocks"
3. Use the weights of the eigenstocks to determine the influence on the desired stock

Project Goals

To simplify matters, lets say there are two "eigenstocks," then looking at a stock from the original space, some linear combination of the eigenstocks will form the arbitrary stock

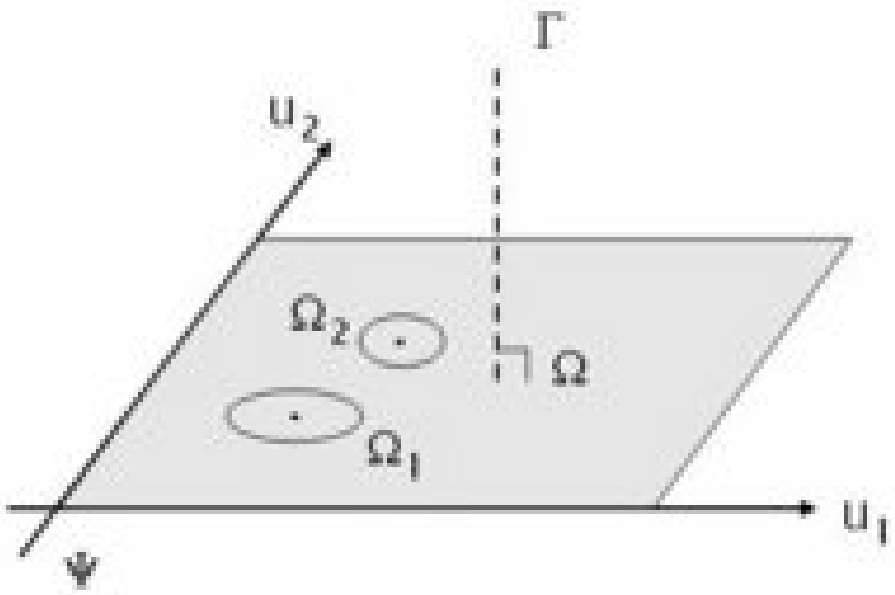


Figure: Sheng Zhang and Matthew Turk (2008), Scholarpedia, 3(9):4244.

Project Goals

In the end, ideally we would like to group stocks based on their similarities in weights associated with eigenstocks. Hopefully within these groups we can form a better understanding of how certain stocks are affected in similar ways with stocks that fall into similar or different categories of "business."

Outline for section 5

State of the Project

- Found a sizable data set with opening and closing values for almost every day since 1960
- The matlab can currently:
  - read data
  - create covariance matrix
  - reduce data to a set of eigenvectors
  - reduces original stocks into the space of the eigenvectors and finds the weights of the eigenvectors required for each stock.
- Still left to do:
  - Determine what the weight martix tells us

```
weight_matrix =

  Columns 1 through 15

    9.6126    0.1436    1.4059   10.2495   10.2550
   -1.6355   -0.5700   -2.7258   16.3045   -0.6183
   -0.1488    1.1945    0.5894   -0.5089   -0.4920
   -1.6767    6.4649   -3.9195    3.2229   -0.6230
   -0.9435   -0.6408    2.1433   13.5451    1.5465
```

Figure: Weight Matrix

References

- Collier, Irwin. "Columbia. Friedman's Lecture Notes to First Hotelling Lecture in Mathematical Economics, 1933." Economics in the Rear-View Mirror, 6 Oct. 2017, www.irwincollier.com/columbia-friedmans-lecture-notes-to-first-hotelling-lecture-in-mathematical-economics-1933/.
- Deshpande, Mohit. "Face Recognition with Eigenfaces." Python Machine Learning, 25 Sept. 2019, pythonmachinelearning.pro/face-recognition-with-eigenfaces/.
- Daughtery, Johnathan. "Linear Algebra, Part 3: Linear Combinations and Matrix Notation." Linear Algebra, Part 3: Linear Combinations and Matrix Notation - Derivative Works, 2013, jtdaugherty.github.io/posts/linear-algebra-3.html.
- Carol Anne Hargreaves, and Chandrika Kadirvel Mani, The Selection of Winning Stocks Using Principal Component Analysis, American Journal of Marketing Research, Vol. 1, No. 3, pp. 183-188, (2015);

References

- Muhammad Waqar, Hassan Dawood,Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, and Ping Guo, Prediction of Stock Market by Principal Component Analysis, Proceedings of the 2017 13th International Conference on Computational Intelligence and Security
- Strang, Gilbert. Linear Algebra and Learning from Data. Wellesley-Cambridge Press, 2019.