Guided Tour of Machine Learning in Finance

Week 3: Unsupervised Learning

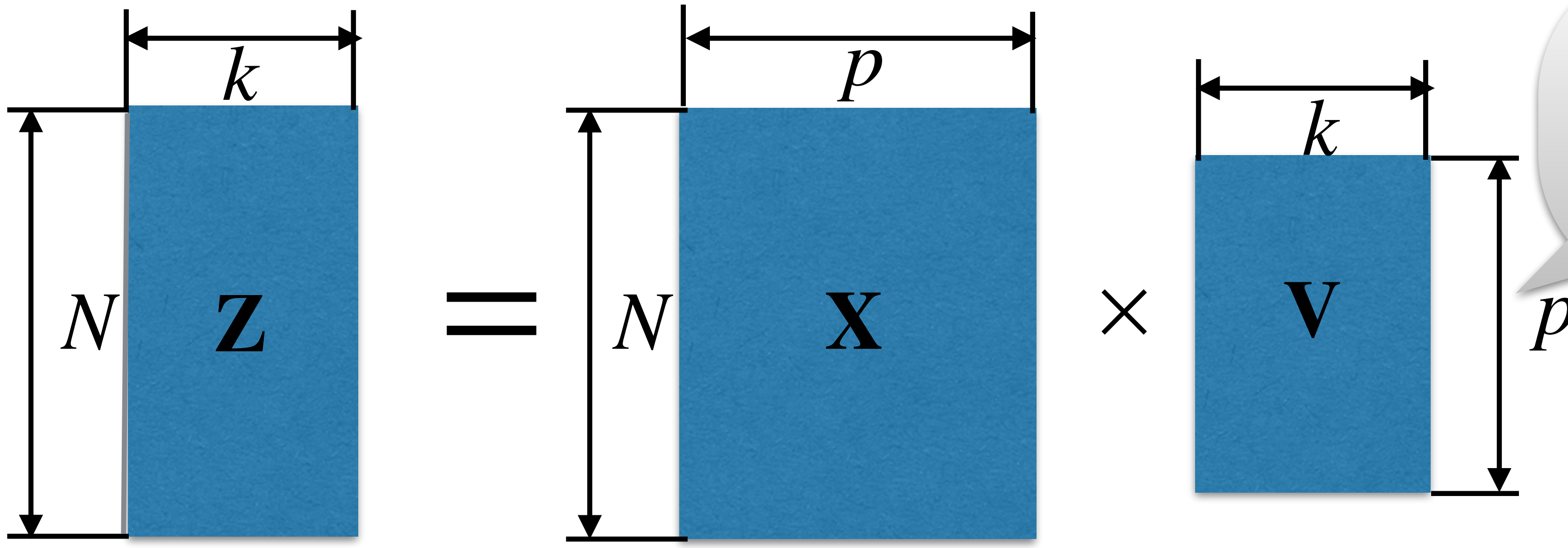
Dimension reduction with the PCA

Igor Halperin

NYU Tandon School of Engineering, 2017

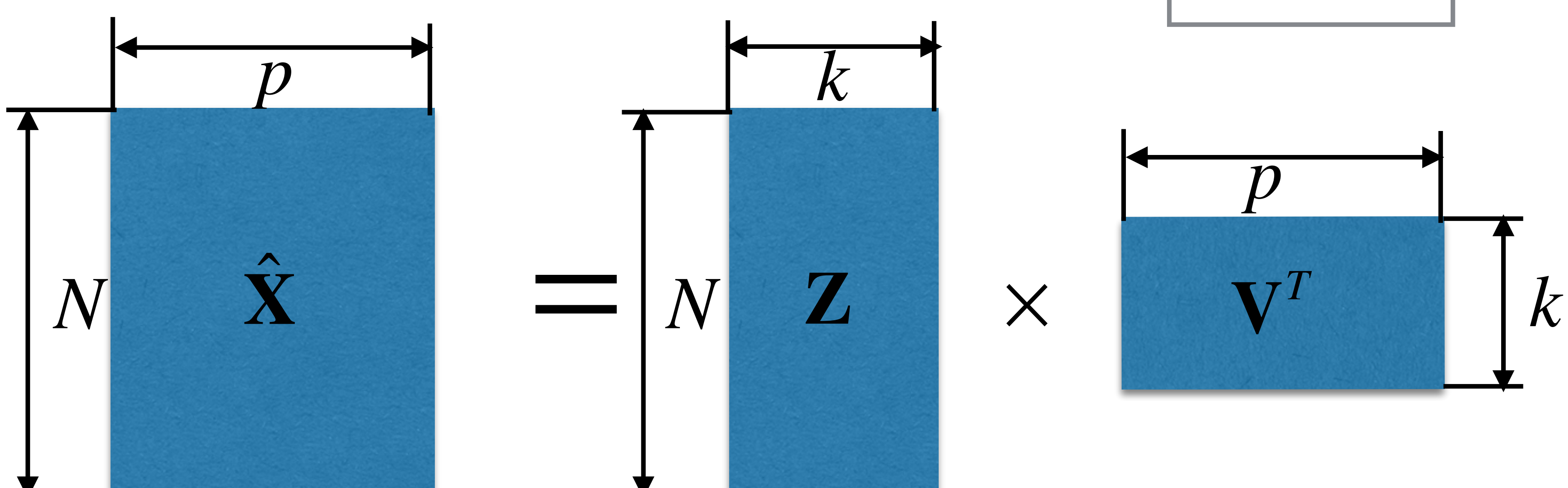
PCA as a dimension reduction method

We had a linear transform (linear encoder) of the data $\mathbf{Z} = \mathbf{X}\mathbf{V}$ parametrized by a $p \times k$ orthogonal matrix $\mathbf{V}$ with $\mathbf{V}\mathbf{V}^T = 1$



If $k = p, \mathbf{V} = \mathbf{U}$ this is just a rotation of the eigenvalue decomposition

A **decoded signal** (<u>inverse transform</u>) is obtained as $\boxed{\hat{\mathbf{X}} = \mathbf{Z}\mathbf{V}^T}$



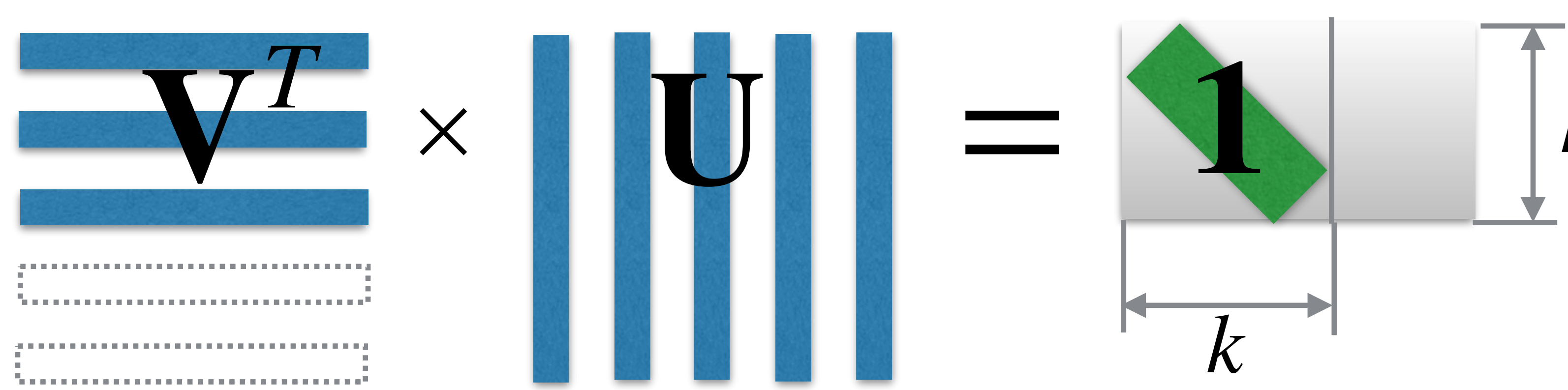
Dimension reduction with PCA

Dimension reduction: make the projection matrix $\mathbf{V}$ of the first $k \le p$ eigenvectors of $\mathbf{C}$

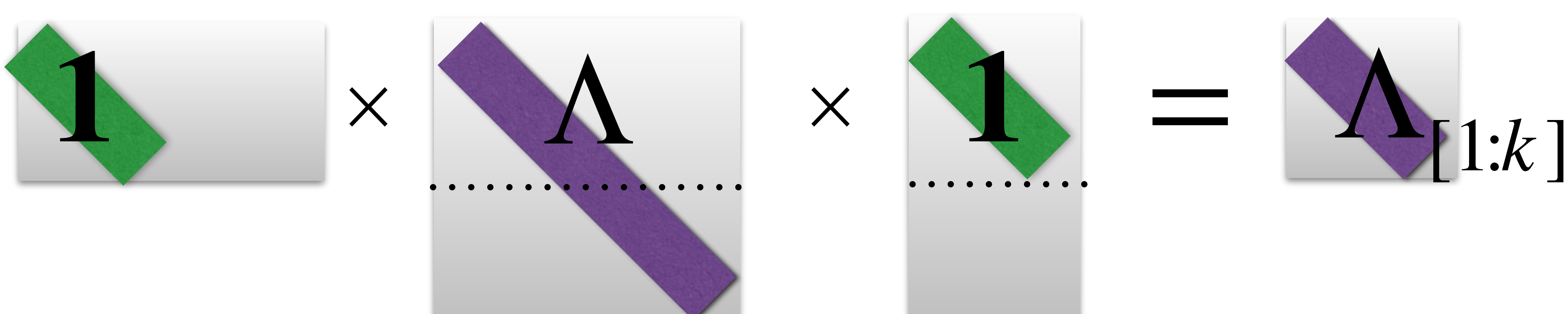
$$\boxed{\mathbf{V} = \mathbf{U}^{[1:k]}}$$

Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{XV}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{XV} = \mathbf{V}^T\mathbf{CV} = \mathbf{V}^T\mathbf{U}\Lambda\left(\mathbf{V}^T\mathbf{U}\right)^T \xrightarrow[\mathbf{V}=\mathbf{U}^{[1:k]}]{} \Lambda_{[1:k]}$$



$$\Leftarrow \mathbf{W_{ij}} = \mathbf{U_i^T U_j} = \delta_{ij},$$
$$i = 1,\ldots,k, \quad j = 1,\ldots,p$$



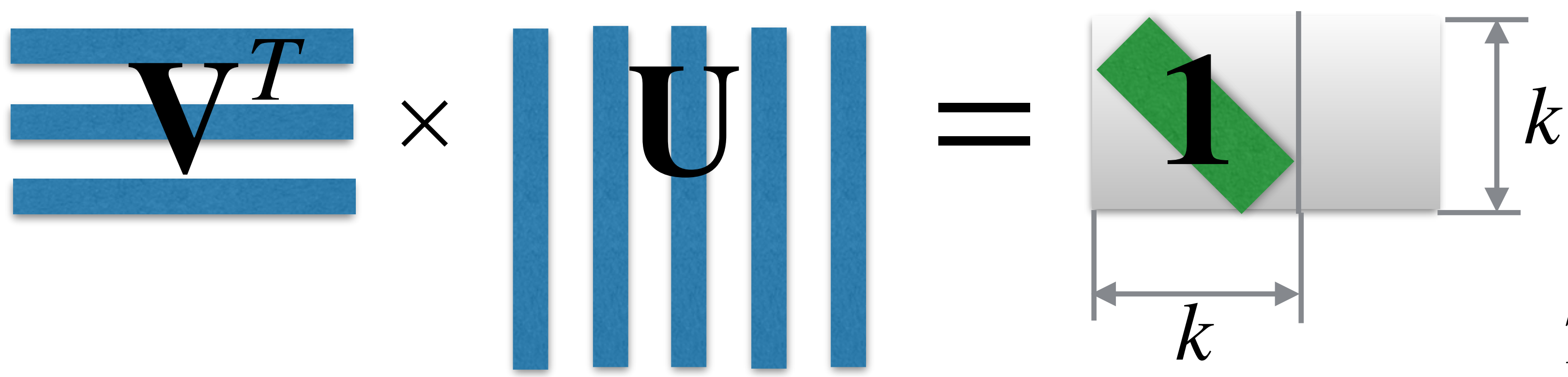
Dimension reduction with PCA

Dimension reduction: make the projection matrix $\mathbf{V}$ of the first $k \leq p$ eigenvectors of $\mathbf{C}$

$$\boxed{\mathbf{V} = \mathbf{U}^{[1:k]}}$$

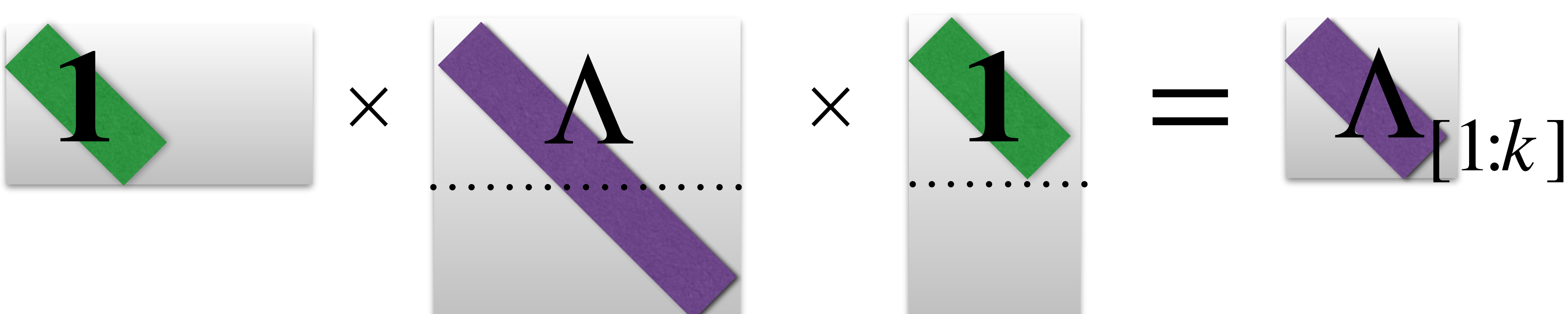
Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{XV}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{V}^T\mathbf{U}\boldsymbol{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T \xrightarrow[\mathbf{V}=\mathbf{U}^{[1:k]}]{} \boldsymbol{\Lambda}_{[1:k]}$$



Only a part of total variation of $\mathbf{X}$ is preserved:

$$TotVar[\mathbf{Z}] = \mathrm{Tr}\left[\boldsymbol{\Lambda}_{[1:k]}\right] = \sum_{i=1}^{k} \lambda_i$$



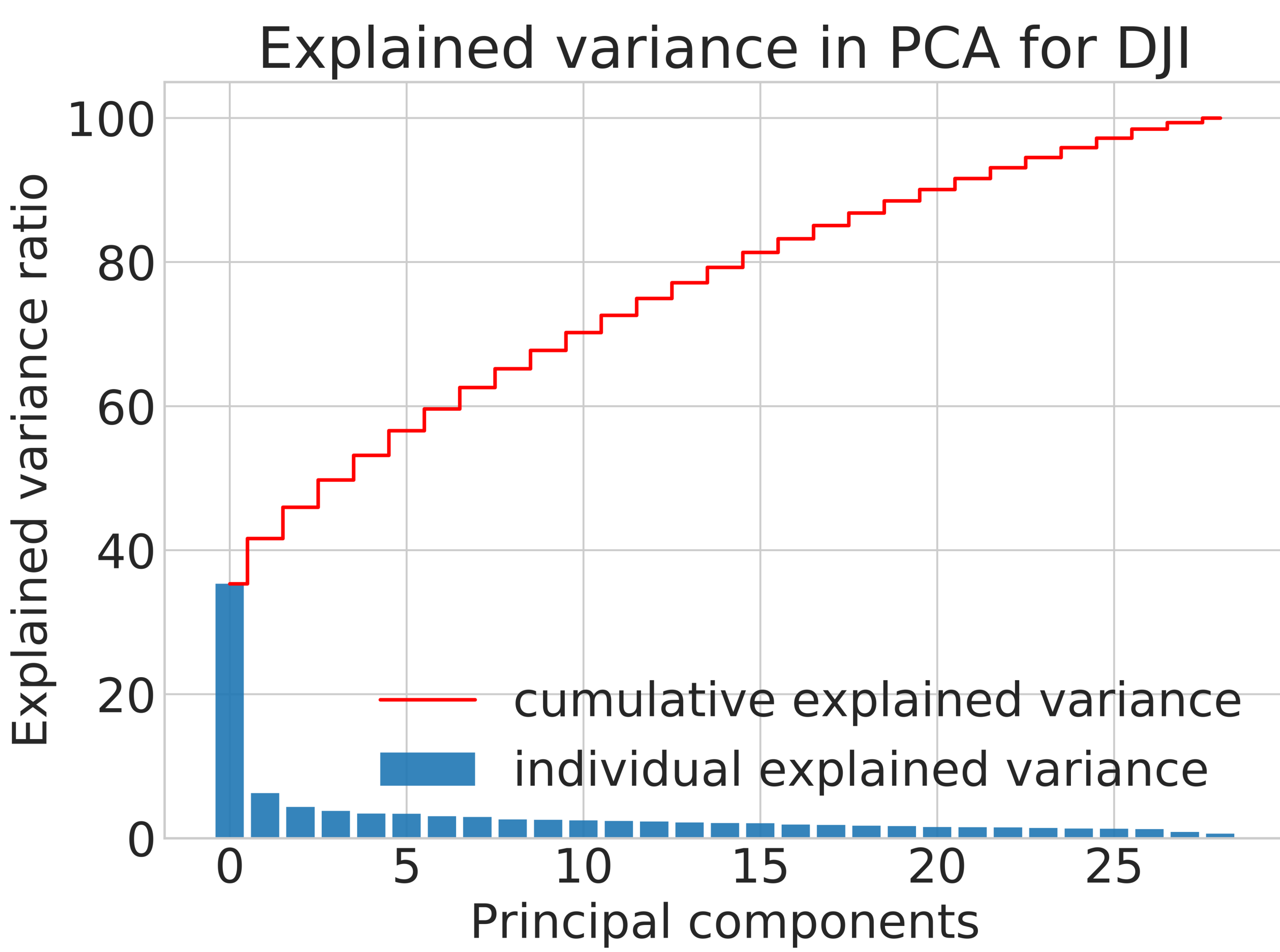
Variance explained by PCA

Dimension reduction: make the projection matrix $\mathbf{V}$ of the first $k \le p$ eigenvectors of $\mathbf{C}$

$$\boxed{\mathbf{V} = \mathbf{U}^{[1:k]}}$$

Use this to compute the covariance matrix of $\mathbf{Z} = \mathbf{X}\mathbf{V}$

$$Cov[\mathbf{Z}] = \frac{1}{N-1}\mathbf{Z}^T\mathbf{Z} = \frac{1}{N-1}\mathbf{V}^T\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}^T\mathbf{C}\mathbf{V} = \mathbf{V}^T\mathbf{U}\mathbf{\Lambda}\left(\mathbf{V}^T\mathbf{U}\right)^T \xrightarrow[\mathbf{V}=\mathbf{U}^{[1:k]}]{} \mathbf{\Lambda}_{[1:k]}$$

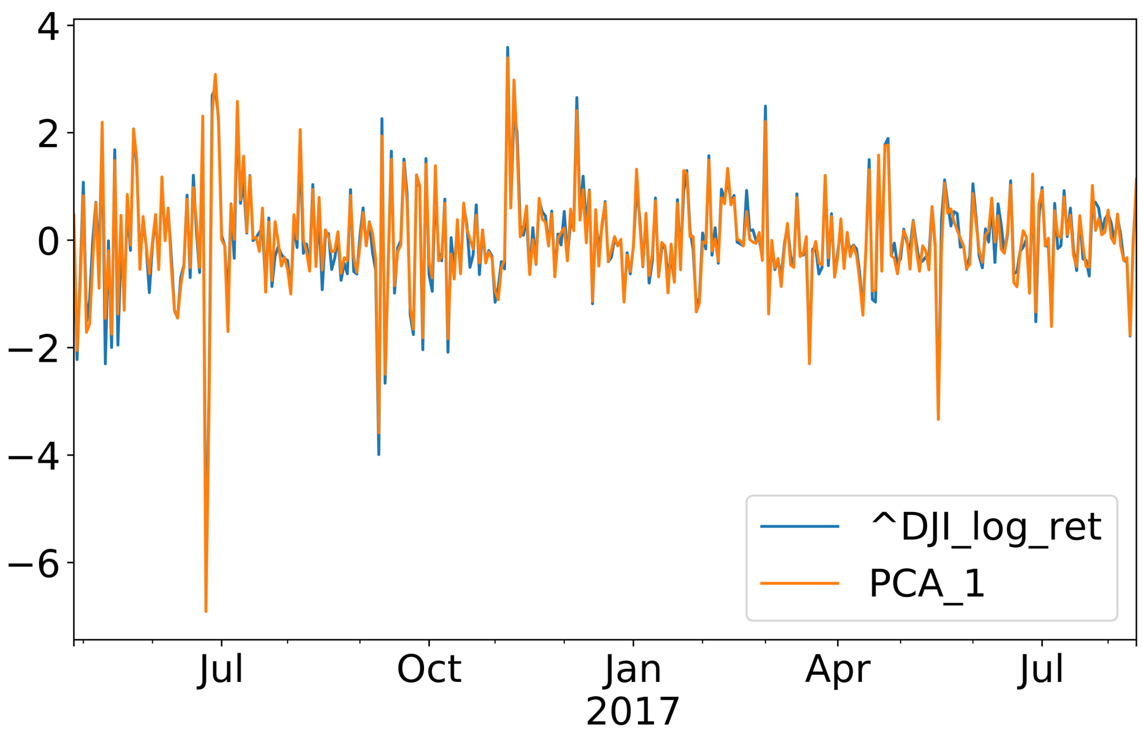


Only a part of total variation of $\mathbf{X}$ is preserved:

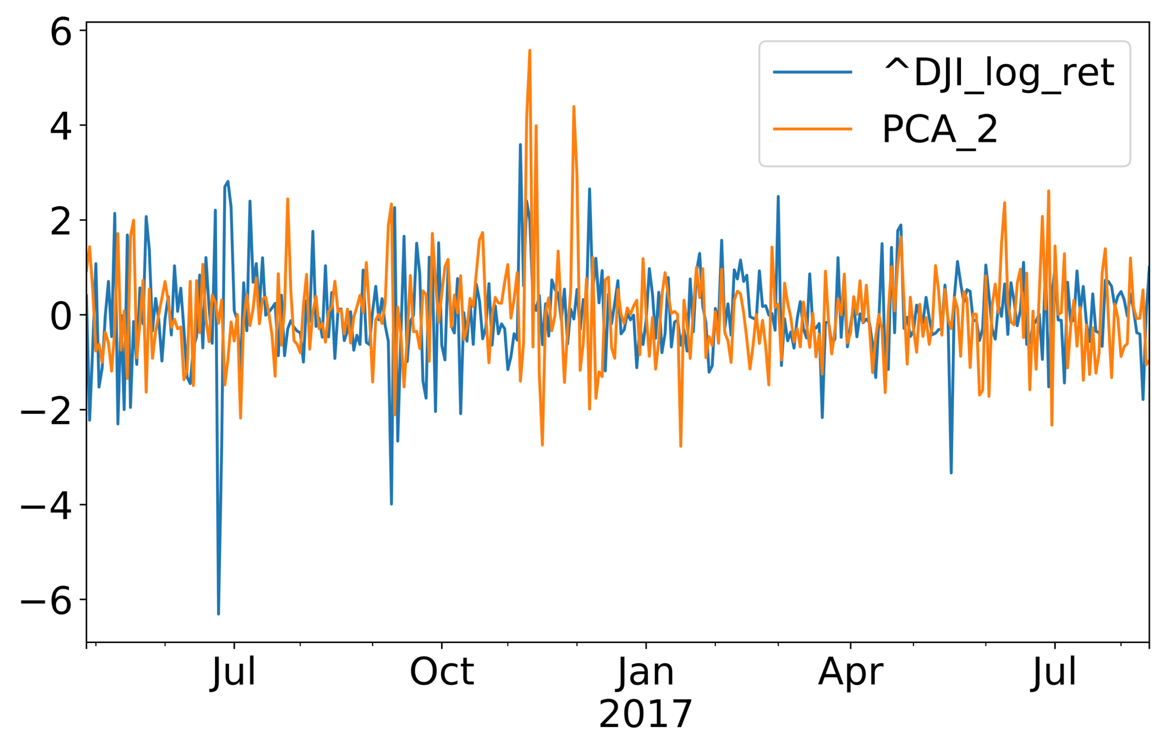
$$TotVar[\mathbf{Z}] = \mathrm{Tr}\left[\mathbf{\Lambda}_{[1:k]}\right] = \sum_{i=1}^{k} \lambda_i$$

PCA projections vs the market

Compare the actual returns on the DJI index with returns of the two leading eigen-portfolios



PC 1



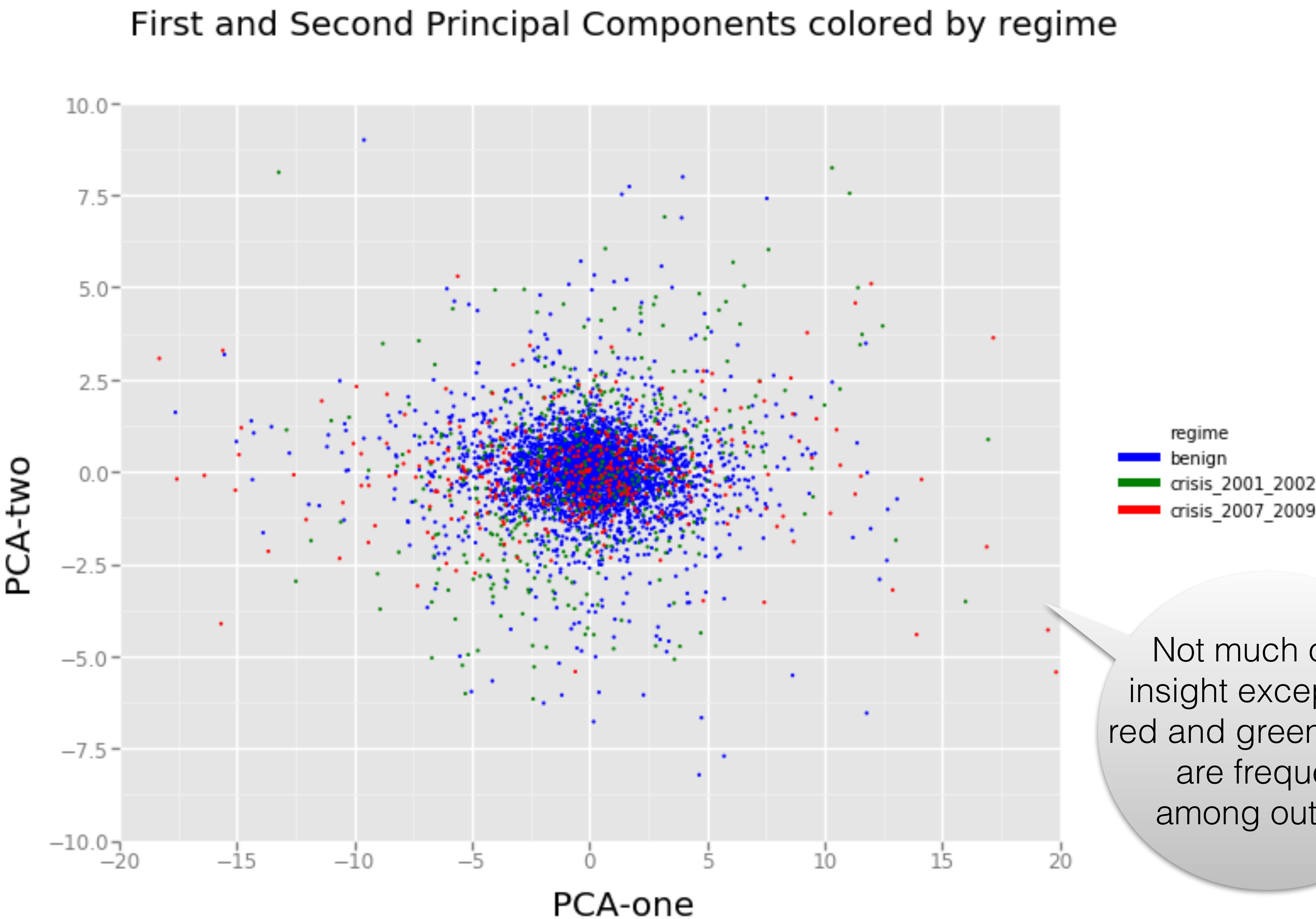
PC 2

Correlated with the market

Uncorrelated with the market

PCA projections: 2D visualization

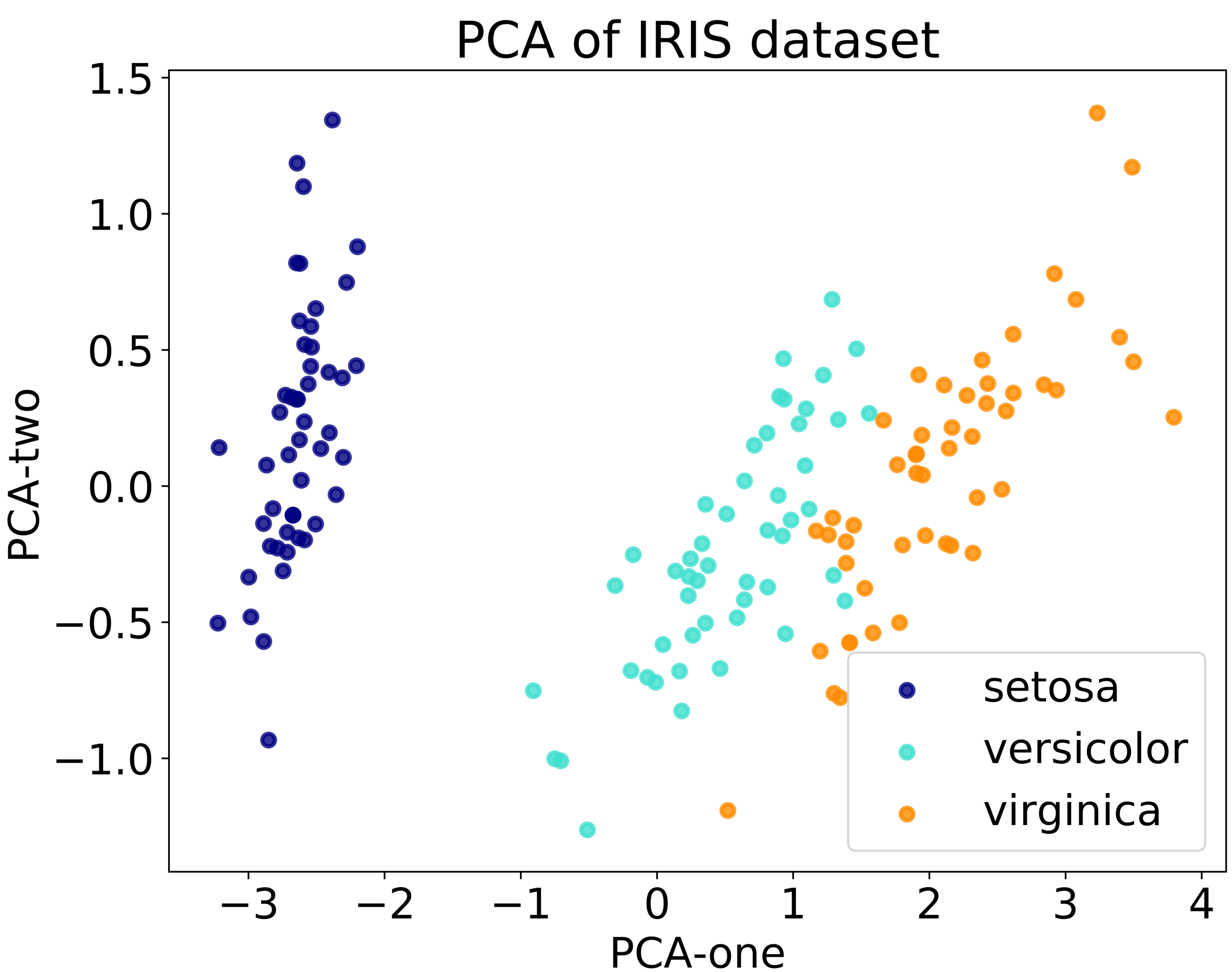
Compare the actual returns on the DJI index with returns of the two leading eigen-portfolios



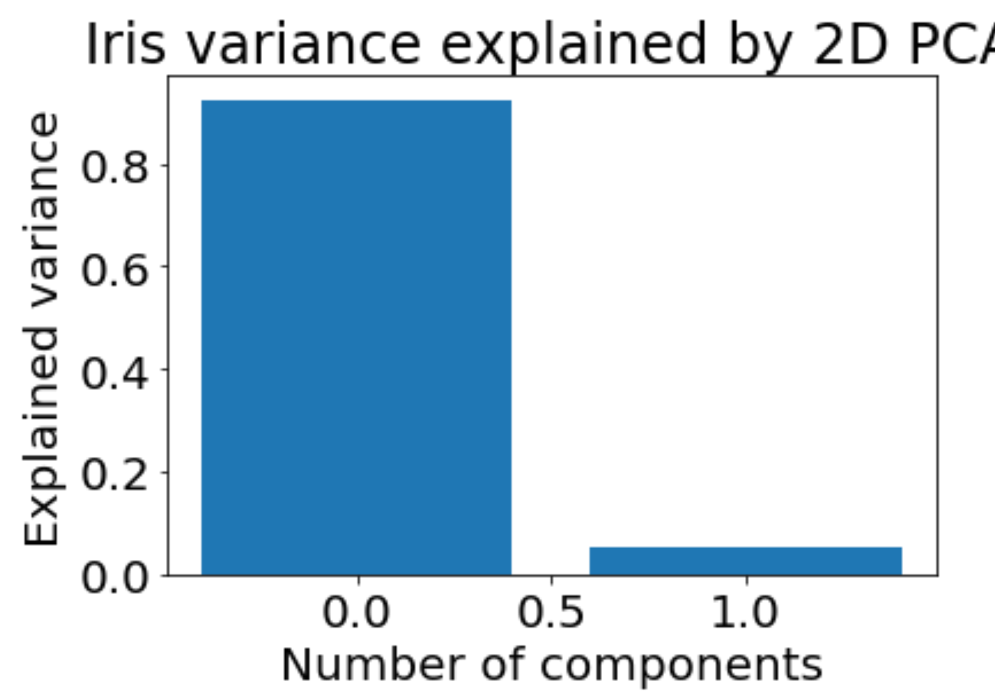
Not much of an insight except that red and green points are frequent among outliers

PCA 2D visualization for the Iris dataset

Compare with the 2D PCA projection for the Iris dataset: 3 types of Iris, 4 features: Sepal Length, Sepal Width, Petal Length and Petal Width, 150 examples, see http://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html.



Different Iris types are well separated by a 2D projection



Quizz: Why does the 2D PCA perform worse for stocks than for Iris?

1. Because stock prices are produced by human interactions, while variability of the Iris species is produced by nature
2. The Iris data is stationary, and is more homogeneous/has less outliers than the stock data
3. Smaller dimension of the feature space
4. Stronger correlations between features
5. The labels are real for Iris (they are Iris' types), but "noisy" for the "regime" label in the stock analysis

Multiple choice answers:

1. Answer 1
2. Answer 3
3. Answers 2 and 4
4. Answer 2 and 5
5. Answers 3, 4, 5
6. Answers 2, 3, 4, 5
7. All of the above

Correct answer: 6

2D PCA for stocks vs Iris

Why does the 2D PCA perform much worse for stocks than for Iris?

- Smaller dimension of the feature space
- Stronger correlations between the features (the first PC explains 92% of total variance!)
- The Iris data is stationary, and is more homogeneous/has less outliers than the stock data
- The labels are real for Iris (they are Iris' types), but "noisy" for the "regime" label in the stock analysis

Financial applications of the PCA

Numerous financial applications of the PCA:

- Systematic trading: construction of market-neutral portfolios with low volatility, extraction of trading signals
- Systemic risk detection methods (e.g. M. Kritzman et. al. "Principal Components as a Measure of Systemic Risk", 2010)
- Risk management: sector risk exposure management of trading book or banking book portfolios, de-noising of empirical correlation matrices
- 2D PCA for data visualization
- PCA for yield curve modeling (fixed income, commodities)