Artificial Intelligence in Finance

Unsupervised learning - part B

Štefan Lyócsa

Department of Finance, Faculty of Economics and Administration

December 4, 2024

Outline for Section 1

The problem

Challenges working with data:

- Dimensionality reduction:
  - **feature reduction** (curse of dimensionality).
  - **feature creation**.
- **Imperfect collinearity** - problem of correlated predictors.
- **Signal** extraction $\rightarrow$ remove the noise.

Useful methods to help out:

- Averaging of standardized variables.
- **P**rincipal **C**omponent **A**nalysis (PCA).
- Graph (Network) theory to identify **complex structures**.
- (Dynamic) factor models,...

Outline for Section 2

Averaging standardized variables

Example

Monthly BTC/USD:

- We plot **Risk** and **Volume** over-time.
- Seem to behave similarly - can we **extract** what is similar in the behavior of the two series (underlying component)?

In order not to have a single variable excessive influence on the analysis, **we standardize each variable**.

- **Z-Score** standardization (already defined before).
- **Rank** standardization (outlier robust approach).

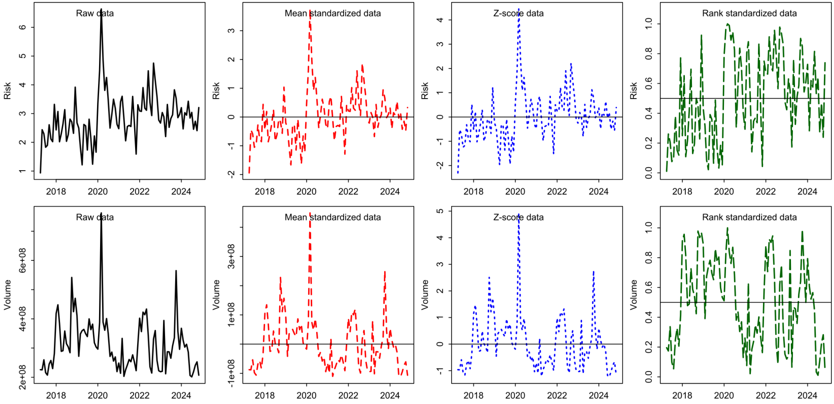
Rank standardization

Let $X_{t,k}$ be k^{th} variable at month $t = 1, 2, ..., T$. Variables are **transformed** to $\boldsymbol{Z_t} = (Z_{1,t}, Z_{2,t}, ...)^\top$, where given k^{it} indicator and $X_{[r],k}$ ordered $X_{t,k}$ we have:

$$Z_{t,k} = \begin{cases} \frac{r}{T} & X_{[r],i} \leq X_{t,k} < X_{[r+1],k} \\ 1 & X_{[T],k} \leq X_{t,k} \end{cases} \tag{1}$$

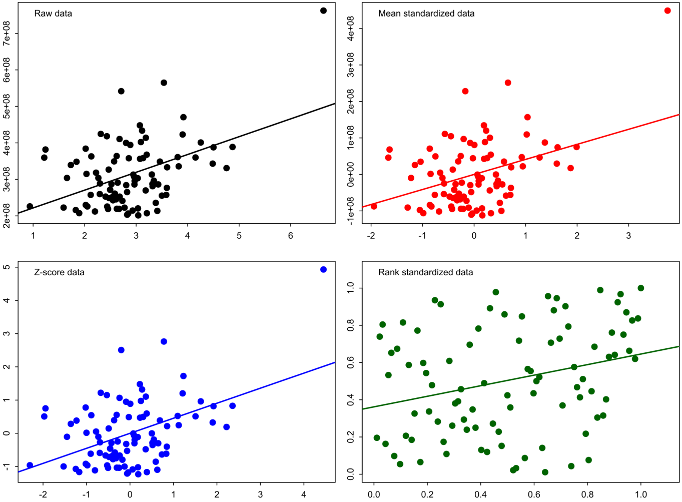
Averaging of standardized variables

Let's take a look at the **time-series dynamics**:

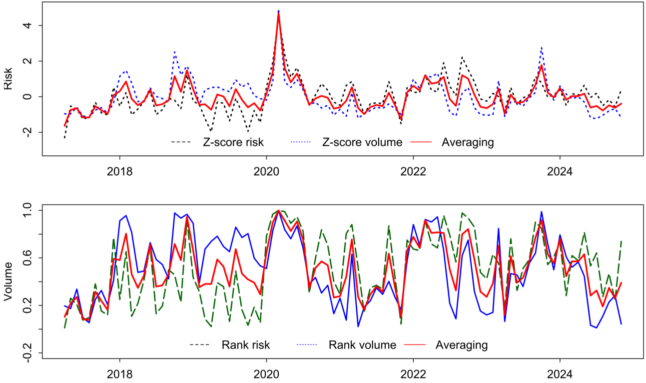


Averaging of standardized variables

Let's take a look at the scatter-plot:



Averaging standardized variables



- Z-Score: $\rho(\text{Risk}, \text{Vol.}) = 0.45$; $\rho(\text{Risk}, \text{Ave.}) = \rho(\text{Vol.}, \text{Ave.}) = 0.85$.
- Rank: $\rho(\text{Risk}, \text{Vol.}) = 0.28$; $\rho(\text{Risk}, \text{Ave.}) = \rho(\text{Vol.}, \text{Ave.}) = 0.80$.

Outline for Section 3

Principal component analysis

Intuition

Averaging is **simple**, but there is a price for simplicity:

- All variables receive the same weight. Is that a reasonable assumption? Well if it works...
- What if there is more **then just one** underlying factor?
  - Many reasons why a firm goes bankrupt $\rightarrow$ no single underlying reason.
  - Risk, Volume, Extreme returns of stock prices might co-move because of multiple factors at play, e.g. market-, global-level uncertainty.

Principal component analysis

Principal component analysis is a process of transforming the ($i = 1, 2, ..., N, k = 1, 2, ..., K$) $N \times K$, Z-score standardized matrix $\boldsymbol{X}$ of features to $\boldsymbol{Z}$ ($N \times K$) that contains K uncorrelated columns and includes same information as the original matrix $\boldsymbol{X}$.

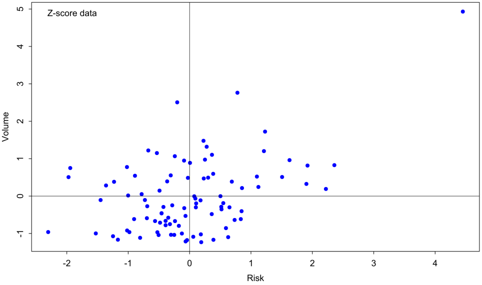
Matrix $\boldsymbol{Z}$ might be useful:

- If variables (features) have common **unobserved** factors, one of the columns of the $\boldsymbol{Z}$ matrix describes most of the movement (variation) in the data.
- One can decide to use $k \ll K$ (e.g. 1 or 2) such informative columns $\rightarrow$ **reducing the dimensionality** of the feature space.

Principal component analysis

Intuition

Assume zero mean variables. What is a **characteristic pattern** in these data?



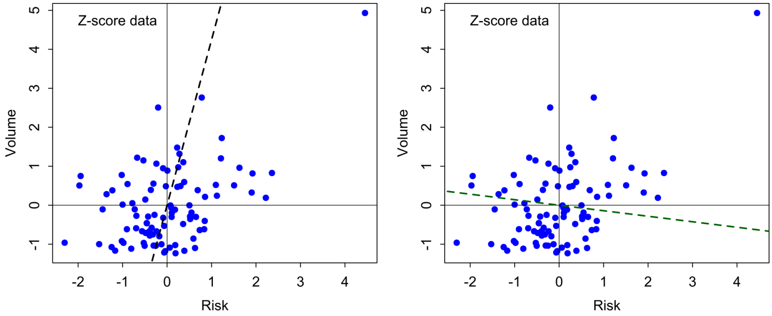
1. Find a **line** such, that the **projection** of each observation on the line (from origin to the point on line) will result in a highest variation; sum of squares of the projected points on that line.
2. Find a line **orthogonal** to the first line.

Principal component analysis

Intuition

The projection:

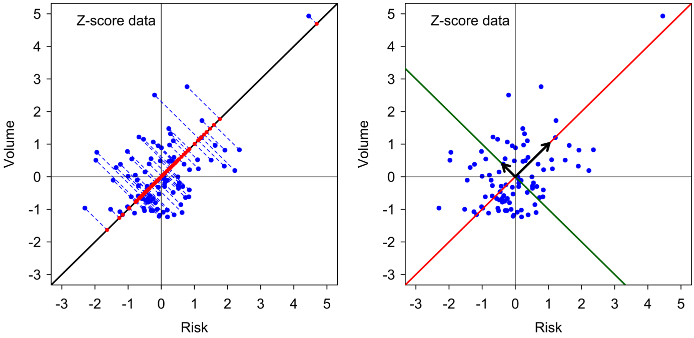
line that maximizes the variance of projected points (minimizes orthogonal distance to the line) $\rightarrow$ but why?



Principal component analysis

Intuition

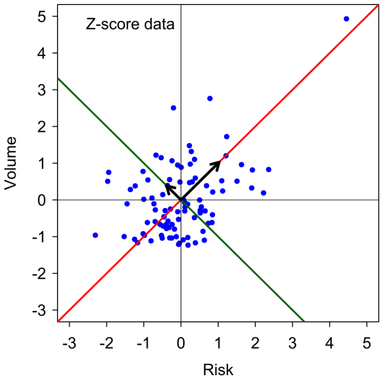
Lines in the figure characterize the co-movement of data. **Eigenvectors** is a vector in the feature space that represent the direction along which max. variance in data is represented. **Eigenvalues** represent the magnitude of the variance which is explained in a specific direction.

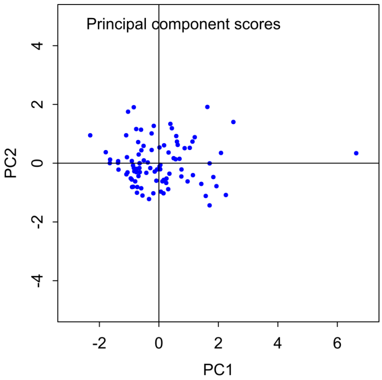


Principal component analysis

In the PCA we are searching for (unit) **vectors** that create lines with these properties $\rightarrow$ eigenvectors. The directions of these eigenvectors are **principal components**.

Note that we can view these principal components as **new axes**.

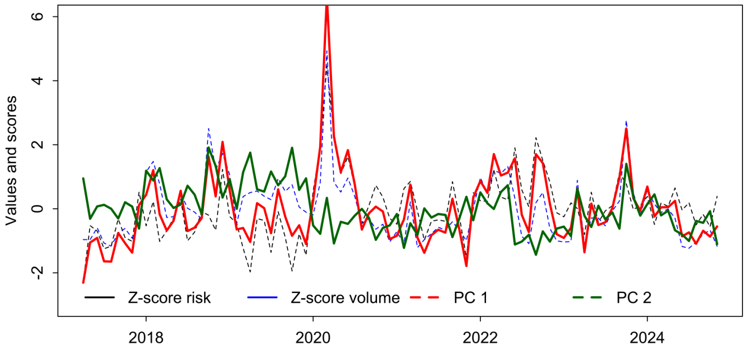




Principal component analysis

Intuition

Finally, given these lines, we can retrieve the **artificial** data - the unobserved factors.



Algorithm in PCA

Common algorithm to find the eigenvectors and eigenvalues is singular value decomposition (prcomp() in R) or covariance matrix estimation (princomp() in R).

1. Standardize data (e.g. remove mean or go for Z-Score). In the following we assume Z-score standardization.
2. Estimate **covariance matrix**.
3. Find **eigenvalues** (determined by the magnitude of variation, i.e. the one maximized while searching for the vector) and **eigenvectors** of the covariance matrix.
4. Analyze results:
   - Feature vector and loadings (weights).
   - Explained variance.
   - How many common factors to select?

Estimation of eigenvectors and eigenvalues

Let $\boldsymbol{A}$ be a (square and symmetric) $K \times K$ covariance matrix of the standardized variables, i.e. $a_{i,i}$ are variances and $a_{i,j}, i \neq j$ are covariances between variables. The **eigenvalue** is given as λ and **eigenvector** ($K \times 1$) as $\boldsymbol{v}$ and are defined as:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \tag{2}$$

Let $\boldsymbol{I}$ be an a $K \times K$ identity matrix with diagonal elements equal to 1 and 0 otherwise.

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{I}\boldsymbol{v} \tag{3}$$

Estimation of eigenvectors and eigenvalues

$$(\boldsymbol{A} - \lambda \boldsymbol{I})v = 0 \tag{4}$$

We want non-trivial solution to eigenvector v, i.e. **should be a non-zero vector**. That would be a solution if we could isolate vector v by simplifying the equation with $(\boldsymbol{A} - \lambda \boldsymbol{I})^{-1}$. In order to avoid that, we should not able to the the simplification, i.e. the matrix $(\boldsymbol{A} - \lambda \boldsymbol{I})^{-1}$ **should not** exists, which means that:

$$\det|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \tag{5}$$

Finding the solution leads to **eigenvalues**. Substituting each eigenvalue one-by-one to Eq. (2) leads to **eigenvectors**. Recall, that eigenvectors show the direction of the relationship in the data, while eigenvalues the magnitude.

Feature vector and loadings (weights)

- The elements of the eigenvector v ($K \times 1$) can be understood as **weights** (or contributions) of each of the features ($k = 1, 2, ..., K$) to the artificially created feature variable (sometimes referred to as principal component).
- **If data were Z-Score standardized** (or have similar variance) we can interpret these elements as weights, i.e. how much weight contributes the given variable to the new feature, relative to the rest of the variables.
- The values allow us to interpret the new feature vector.

Feature vector and loadings (weights)

Let $\boldsymbol{X}$ be a $N \times K$ matrix of Z-Score standardized features and let's stack the column vectors υ (for each component) into a $K \times K$ matrix $\boldsymbol{\Upsilon}$. The artificial variables can be extracted as:

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Upsilon} \tag{6}$$

The $(N \times K)$ matrix $\boldsymbol{Z}$ is the transformed matrix of the original dataset $(\boldsymbol{X}) \rightarrow$ the goal of the principal component analysis.

Feature vector and loadings (weights)

Continuing the example above, the two **eigenvectors** are:

- $\boldsymbol{v}_1 = (0.707, 0.707)^\mathsf{T}$
- $\boldsymbol{v}_2 = (-0.707, 0.707)^\mathsf{T}$

The relative magnitude of the relationship is given by **eigenvalues** that are 1.473 and 0.527. The two principal component scores found as:

$$p_{i,1} = x_{i,1} \times 0.707 + x_{i,2} \times 0.707$$
$$p_{i,2} = x_{i,1} \times (-0.707) + x_{i,2} \times 0.707$$

If variables have similar variance or were Z-score standardized, the elements of the eigenvector can be interpreted as **loadings** - weights.

Explained variance

- Note, that the variance of Z-score standardize features stored in matrix $\boldsymbol{X}$ as columns, is 1. Thus the overall variance is K (i.e. equal to number of columns/features).
- Let denote the variance of a column (principal component score) of matrix $\boldsymbol{Z}$ as $D[\boldsymbol{Z}_j], j = 1, 2, ..., K$. For Z-score standardized features, the sum of variances $\sum_{j=1}^{K} D[\boldsymbol{Z}_j] = K$ as well.
- **Importance** of a principal score variable can be defined by the **amount of variance it explains**:

$$EV_j = \frac{D[\boldsymbol{Z}_j]}{\sum_{j=1}^{K} D[\boldsymbol{Z}_j]} \tag{7}$$

  for Z-score standardized variables:

$$EV_j = \frac{D[\boldsymbol{Z}_j]}{\sum_{j=1}^{K} D[\boldsymbol{Z}_j]} = K^{-1} \times D[\boldsymbol{Z}_j] \tag{8}$$

How many common factors to select?

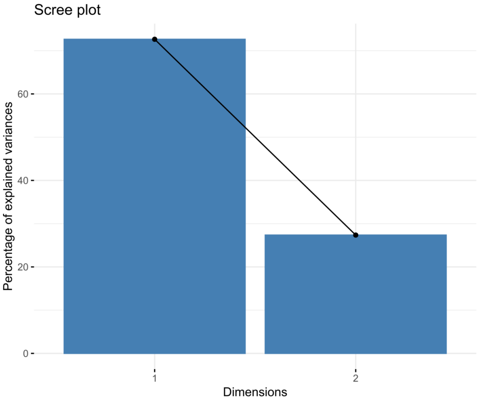
Couple of rules-of-thumbs:

- Select first n scores that explain **at least** 80% **of the total variance**.
- **Guttman - Kaiser criterion** to use scores that have eigenvalues greater as 1 [3, 5, 4] (or more general, the explained variance is above the average).
- Use a scree-plot, where on x-axis are $j = 1, 2, ..., K$ principal components and on y-axis variance $D[\boldsymbol{Z}_j]$, but values are **ordered from highest** variance **to lowest**. Select only factors on the steep part of the resulting line (before the kink).

How many common factors to select?

From the previous example:

- Variance explained is 72.60% and 27.40%.



Outline for Section 4

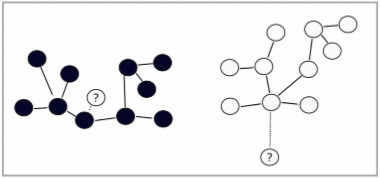
Intuition

The idea is to identify **hidden complex structures**. The setup:

- We have data $X_{i,k}, i = 1, 2, ..., N, k = 1, 2, ..., K$, where i is an observation and k is a given feature.
- Similarly to cluster analysis, we select some **distance measure** and construct a $N \times N$, similarity matrix, $\boldsymbol{D}$, between observations (e.g. distance matrix, correlation matrix, ...).

Intuition

Given the similarity matrix and some **rules** we create a graph; which we hope uncovers hidden relationships (e.g. clusters of risky assets, profitable clients,....). A graph $G = (V, W)$ is a mathematical object where V is a set of vertices (nodes) and W is a set of (weighted edges).



Vertices tend to be given, how about edges?

Intuition

Example

What type of an asset is Bitcoin? Let $P_{t,i}$ be the closing price of asset $i = 1, 2, ...,$ at time $t = 1, 2, ...,T$. Daily returns are defined as:

$$R_{t,i} = 100 \times \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}} \tag{9}$$

The elements of the $N \times N$ correlation matrix $c_{i,j} \in \boldsymbol{C}$ correspond to correlations $\rho_{i,j} \in [-1, 1]$ between daily returns of assets i and j.

Intuition

Example

In our case of $N = 27$ assets:

$$\begin{bmatrix} 1 & 0.14 & \dots & 0.31 & 0.29 \\ 0.14 & 1 & \dots & 0.28 & 0.15 \\ \dots & \dots & \dots & \dots & \dots \\ 0.31 & 0.28 & \dots & 1 & 0.43 \\ 0.29 & 0.15 & \dots & 0.43 & 1 \end{bmatrix}$$

We have $N \times (N - 1) \times 2^{-1} = 351$ correlations $\rightarrow$ a **complex system of relationships**. How to select *relevant* correlations?

- Minimum spanning tree, e.g. [7].
- Asset graphs, e.g. [8].
- Granger causality graphs, e.g. [6, 2].

Intuition

Example

Correlations are not distances (...), we use the transformation from Mategna [7] to arrive at $\boldsymbol{D}$:

$$d_{i,j} = \sqrt{2(1 - \rho_{i,j})} \in [0, 2] \tag{10}$$

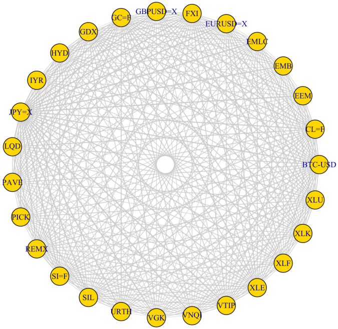
Lower valued of $d_{i,j} \rightarrow$ similar objects. The matrix from above now looks as:

$$\begin{bmatrix} 0 & 1.31 & \dots & 1.17 & 1.19 \\ 1.31 & 0 & \dots & 1.20 & 1.30 \\ \dots & \dots & \dots & \dots & \dots \\ 1.17 & 1.20 & \dots & 0 & 1.07 \\ 1.19 & 1.30 & \dots & 1.07 & 0 \end{bmatrix}$$

Complete graph

Example

Now we can create a graph, vertices are financial assets, edges are distances:

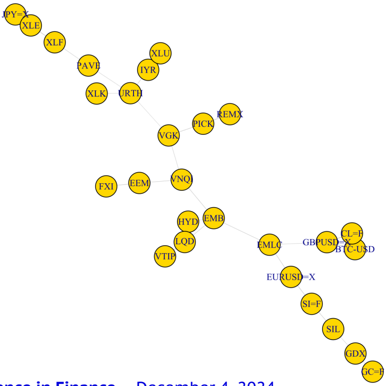


This *mess* is called a **complete** graph and is not very useful.

Minimum spanning tree

Example

A **sub-graph** of the complete graph, that retains only $N-1$ edges which connect all the vertices together, with the minimum possible total edge weight and no cycles:



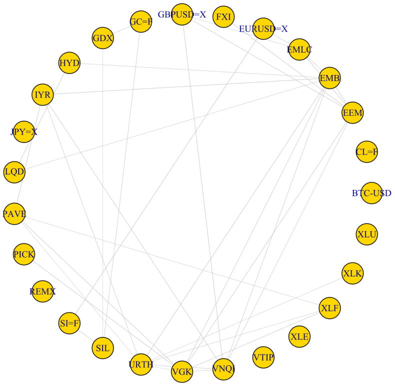
Minimum spanning tree

How do we create such graphs?

- Prim's algorithm [9].
- Kruskal's algoritm [9].
- Boruvka's algoritm [1]: Borůvka, O. (1926). O jistém problému minimálním. Práce Mor. Přírodověd. Spol. V Brně III 3: 37−58.

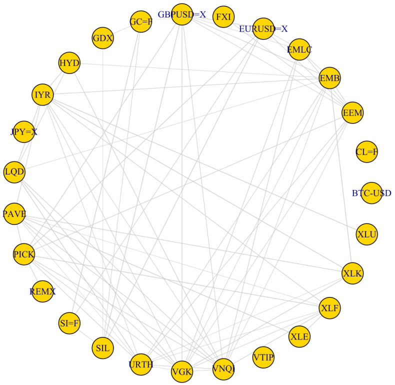
Asset graph

MST has one topological property - the resulting tree (graph) must be connected. This requirement does not make much sense in economics/finance [10]. A much simpler alternative is to retain r% lowest distances. For example 10%:



Asset graph

MST has one topological property - the resulting tree (graph) must be connected. This requirement does not make much sense in economics/finance [10]. A much simpler alternative is to retain $r\%$ lowest distances. For example 30%:



Notes on the use of networks

Networks allow us to extract some useful information, e.g.:

- vertex degree (local centrality measure),
- betweeness (global centrality measures),
- clusters of observations,
- ....

These can be used as features.

[1] Otakar Borvka. “O jistém problému minimálnım”. In: (1926).

[2] Oleg Deev and Štefan Lyócsa. “Connectedness of financial institutions in Europe: a network approach across quantiles”. In: *Physica A: Statistical Mechanics and its Applications* 550 (2020), p. 124035.

[3] Louis Guttman. “Some necessary conditions for common-factor analysis”. In: *Psychometrika* 19.2 (1954), pp. 149–161.

[4] Henry F Kaiser. “A note on Guttman’s lower bound for the number of common factors.”. In: *British Journal of Statistical Psychology* (1961).

[5] Henry F Kaiser. “The application of electronic computers to factor analysis”. In: *Educational and psychological measurement* 20.1 (1960), pp. 141–151.

[6] Štefan Lyócsa, Tomáš Vrost, and Eduard Baumöhl. "Return spillovers around the globe: A network approach". In: *Economic Modelling* 77 (2019), pp. 133–146.

[7] Rosario N Mantegna and H Eugene Stanley. *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

[8] Jukka-Pekka Onnela et al. "Asset trees and asset graphs in financial markets". In: *Physica Scripta* 2003.T106 (2003), p. 48.

[9] Robert Clay Prim. "Shortest connection networks and some generalizations". In: *The Bell System Technical Journal* 36.6 (1957), pp. 1389–1401.

[10] Tomáš Vrost, Štefan Lyócsa, and Eduard Baumöhl. "Granger causality stock market networks: Temporal proximity and preferential attachment". In: *Physica A: Statistical Mechanics and its Applications* 427 (2015), pp. 262–276.

MUNI
ECON

Artificial Intelligence in Finance

Unsupervised learning - part B

Štefan Lyócsa

Department of Finance, Faculty of Economics and Administration

December 4, 2024

MASARYK
UNIVERSITY