

Week 8 Examination Assignment

K0021N, Luleå University of Technology

Rickard Garvare

2025-04-27

Regarding the Assignment

This document outlines the **individual** examination assignment for this week. We strongly suggest that you work through the literature, the videos, and the training assignments recommended for this week before progressing to the examination assignment.

Prepare and document your solution

You need to prepare your solutions before the examination session. At the examination session, you will need to have your imported data files and prepared solutions easily available so that you will be able to perform the tasks you are given at the session.

There are some options for you to prepare, document, and save your solutions. You can do it in a standard R script. If using Quarto, executable code can be combined with your own notes. You can also create MS Word, pdf, or html documents with your solutions. Whatever choice you make it is important that you can run your prepared solutions fast and easy at the examination sessions. Please make it easy for the reader to understand what you have done, and it should also be easy to distinguish program code from other types of text.

Examination Assignment using PCA and K-Means Clustering

In this assignment, you will work with the `w8_exam_data.csv` dataset.

Scenario

You are a central part of the strategic analytics team at a multinational company that is transitioning from selling stand-alone industrial equipment to (as your CEO describes it) “offering outcome-based solutions through platform ecosystems and AI-driven services”.

The company faces intensified competition as traditional product-centric players, digital-native companies, and AI-powered platforms now have started to disrupt the market. To sustain a competitive

edge, your company aims to maximize value creation, value capture, and value delivery through digital servitization, focusing on predictive maintenance, performance optimization, and integrated service contracts.

Your team manager now wants you to analyze available market and customer data to:

- Segment customers based on usage patterns, technological readiness, and preferences for service models.
- Provide actionable insights on how to tailor outcome-based offerings and platform-driven strategies.

About the provided dataset

You will work with a simulated dataset intended to mimic real-world data on 200 companies that are current or potential clients. Each company has been scored on several key dimensions:

- **x1, Annual_Equipment_Spend:** Total yearly spend on industrial equipment.
- **x2, AI_Readiness:** Level of AI infrastructure and competence.
- **x3, Platform_Adoption:** Extent of integration with digital platforms.
- **x4, Service_Contract_Preference:** Preference for service contracts over purchasing products.
- **x5, Predictive_Maintenance_Adoption:** Usage level of predictive maintenance solutions.
- **x6, Outcome_Orientation:** Focus on business outcomes (uptime, efficiency) vs. product ownership.
- **x7, Digital_Maturity:** Maturity of digital processes and IT systems.
- **x8, Data_Share_Willingness:** Willingness to share operational data with suppliers.
- **x9, Customization_Demand:** Need for highly customized solutions.
- **x10, Innovation_Partner_Score:** Openness to joint development of new services.

As seen when examining the dataset it contains of 13 features named **x1** ... **x13**. We do not know the scales of the characteristics mentioned above. Neither do we know what the extra three features may represent in terms of company characteristics.

List of tasks

Prepare by importing your data. If necessary, clean up names of variables/features as you see fit. Make sure the variables have the correct classes to fit the algorithms that you will use. Standardize if needed.

Kindly note that numbers in this dataset have a decimal comma instead of a decimal point. To overcome this problem, you might use the following code in RStudio when importing the data:

```
w8_exam_dataset <- read.csv("w8_exam_data.csv", dec = ",", sep = ";")
```

Run a Principal Component Analysis (PCA) to reduce dimensionality and identify underlying patterns in the data. Decide how many principal components to retain in your model (combining what you can learn from a scree plot, from cumulative variance explained, and from the cutoff value for eigenvalues). Analyze variable loadings and scores. If possible, interpret what each retained principal component may represent in business terms. Visualize using graphs of your choosing.

Perform a cluster analysis using the K-Means algorithm to segment the clients into meaningful groups. Use suitable methods to decide on an appropriate number of clusters. Visualize using graphs of your choosing. Make an effort to make graphs visually appealing. Make a short description of each cluster's characteristics (e.g., "digitally mature companies with high outcome orientation").

Provide a summary of your conclusions. Include a recommendation regarding which cluster you identify as the most promising one. Critically discuss advantages and limitations of using PCA and K-Means in this scenario.

Regarding additional assignments at the examination session

If you have solved the above tasks, documented your solutions, and are able to run the code fast in the classroom you should be well-prepared to solve the additional assignment(s) that you will get at the examination session. In preparation for the examination assignments it is also important that you have worked through the training assignments. If you have done all this, do not worry!