Technical University of Denmark

02452 Machine Learning
DTU Compute
Fall 2025

DTU

Machine Learning Project 1

Data: Feature extraction, and visualization

September 21, 2025

Group 94

Name	Student number
Vincent Van Schependom	s251739
Diego Armando Mijares Ledezma	s251777
Albert Joe Jensen	s204601

Table 1: Group members.

Task	Vincent	Diego	Albert
Section 1	30%	30%	40%
Section 2	40%	30%	30%
Section 3	30%	40%	30%
Section 4	30%	30%	40%
LaTeX	90%	5%	5%

Table 2: Contributions & responsabilities table.

Introduction

The objective of this report is to apply the methods that were discussed during the first section of the course *Machine Learning* [1] to a chosen dataset. The aim is to get a basic understanding of the data prior to the further analysis (project report 2).

The particular dataset that is being investigated is the *Glass Identification* dataset from 1987 by B. German [2]. Table 1 lists our full names and student numbers, while Table 2 shows an overview of the contribution of each team member.

Contents

1 The *Glass Identification* dataset **2**

2 A close look at the different attributes **2**

3 Descriptive analysis of the dataset **2**
3.1 Extreme values and outliers . . . 2
3.2 Distribution of the attributes . . . 2
3.3 Correlation between attributes . . . 2

4 Principal Component Analysis **3**
4.1 Dimension reduction . . . 3
4.2 Principal directions . . . 3
4.3 Projected data . . . 3

1 The *Glass Identification* dataset

[TODO: The introduction to the data set.]

2 A close look at the different attributes

Attribute	Description	Type of variable
`ID`	Observation ID (excluded from analysis)	Numeric (discrete)
`RI`	Refractive Index	Continuous
`Na`	Sodium oxide (Na_2O)	Continuous
`Mg`	Magnesium oxide (MgO)	Continuous
`Al`	Aluminum oxide (Al_2O_3)	Continuous
`Si`	Silicon oxide (SiO_2)	Continuous
`K`	Potassium oxide (K_2O)	Continuous
`Ca`	Calcium oxide (CaO)	Continuous
`Ba`	Barium oxide (BaO)	Continuous
`Fe`	Iron oxide (Fe_2O_3)	Continuous
`type`	Type of glass	Nominal

Table 3: [TODO: Fill in.]

	Abbreviation in dataset	Description
1	`BW-FP`	Building Window, Float Processed
2	`BW-NFP`	Building Window, Non Float Processed
3	`VW-FP`	Vehicle Window, Float Processed
4	`VW-NFP`	Vehicle Window, Non Float Processed
5	`containers`	Containers
6	`tableware`	Tableware (e.g. ...)
7	`headlamps`	Headlamps (e.g. ...)

Table 4: [TODO: Fill in.]

[TODO: Detailed explaination of the attributes of the data.]

3 Descriptive analysis of the dataset

[TODO: ...]

3.1 Extreme values and outliers

[TODO: ...]

3.2 Distribution of the attributes

[TODO: ...]

3.3 Correlation between attributes

[TODO: ...]

4 Principal Component Analysis

[TODO: Explain why we standardise.]

4.1 Dimension reduction

We aim to reduce the 9-dimensional dataset into an M-dimensional one (with $M < 9$).

[TODO: How many PC's do we keep? I think $M = 5$.]

4.2 Principal directions

The *principal directions* of the (first M) principal components are the rotations, corresponding to each principal component $\text{PC}_i = \boldsymbol{v_i}$ in the transform-matrix $\boldsymbol{V}_M$. This matrix is used when computing the projected coordinates $\boldsymbol{B} = \boldsymbol{V}_M \boldsymbol{X}$ of the original data $\boldsymbol{X}$ onto the subspace spanned by the first M principal components.

Variable	$\mathbf{PC}_1$	...	$\mathbf{PC}_M = \mathbf{PC}_{5??}$
RI	0	...	0
Na	0	...	0
Mg	0	...	0
Al	0	...	0
Si	0	...	0
K	0	...	0
Ca	0	...	0
Ba	0	...	0
Fe	0	...	0

Table 5: The principal directions (a.k.a. the *loadings*) of the first M principal components $\text{PC}_i = \boldsymbol{v_i}$ in the rotation matrix $\boldsymbol{V}_M$. [TODO: Describe what these directions mean in terms of the original attributes.]

4.3 Projected data

[TODO: Visualisations of the projected data]

Summary

[TODO: A short summary of what we discussed in the whole paper.]

References

[1] Tue Herlau, Mikkel N. Schmidt, and Morten Mørup. *Introduction to Machine Learning and Data Mining*. Technical University of Denmark (DTU), Lyngby, Denmark, 2023. Lecture notes, Fall 2023, version 1.0. This document may not be redistributed. All rights belong to the authors and DTU.

[2] B. German. Glass Identification. UCI Machine Learning Repository, 1987. DOI: https://doi.org/10.24432/C5WW2P.