

Homework

Principal Components Analysis

Computational Linear Algebra for Large Scale Problems

Politecnico di Torino

A.Y. 2024/2025

Contents

1	How the Homework Must be Prepared and Uploaded	2
2	How to Prepare the Report	2
3	The Homework	2
3.1	Dataset Description	2
3.2	The Files	3
3.3	Exercises	3
3.3.1	Simulation of a Real-World Problem	4
A	Survey's Questions	5
A.1	Music Preferences	5
A.2	Movie Preferences	5
A.3	Hobbies & Interests	6
A.4	Phobias	7
A.5	Health Habits	7
A.6	Personality Traits, Views on Life, and Opinions	7
A.7	Spending Habits	9
A.8	Demographics	10

1 How the Homework Must be Prepared and Uploaded

Here, all the necessary information to prepare the homework for the exam:

- The homework can be done alone or in group. Groups with three or more people are not allowed;
- The assignment must consist in **exactly** two files:
 1. The Ipython notebook with the exercises. It must be named as

HWpca_Studentsurname1_Studentsurname2.ipynb

or *HWpca_Studentsurname1.ipynb* for single students;

2. The PDF obtained by exporting/printing the Ipython notebook (same name of the notebook).

Very important: this file is the official and “printed” version of your report! Then, be sure that the plots are well represented and all the comments refer to visible codes/pictures/tables etc.

- The compressed file must be uploaded within the date indicated by professor Berrone;
- **Attention:** the only python modules and packages allowed are the ones used in the laboratories; i.e., the ones specified in the requirements file on the web page of the course. The teacher must be able to run your code using the environment created for the laboratories!

2 How to Prepare the Report

The report of this homework is a document where the student (or the group of students) writes and comments the addressed problem, the performed procedures, and the results of the exercises adding the codes. For this reason, the report must be written on the given jupyter-notebook, since it allows to “merge” the textual descriptions and argumentations with python codes.

In particular, an almost-empty notebook is given to the students. Inside this notebook there are the instructions for the homework.

Together with the notebook, also the dataset is given.

3 The Homework

Here, we introduce the dataset used for this homework and we list the exercises and the analyses to be performed.

In particular, all the files required for this homework (included this document) are available in the folder *DellaSanta/HWpca* of the web page of the course.

3.1 Dataset Description

The dataset used in this homework is the *Young-People-Survey* (YPS) dataset; the dataset origin dates back to 2013, when were asked to the students of the Statistics class at the *Faculty of Social and Economic Sciences of the University of Bratislava* to invite their friends to participate in a survey. The questions in the survey concern eight different categories of questions:

1. Music preferences (19 items)
2. Movie preferences (12 items)
3. Hobbies & interests (32 items)

4. Phobias (10 items)
5. Health habits (3 items)
6. Personality traits, views on life, & opinions (57 items)
7. Spending habits (7 items)
8. Demographics (10 items)

Observation 3.1 (Type of data). Excluding the “Demographics” group, almost all the answers consist of a rating between 1 and 5. The non-rating answers are instead characterized by categorical values; however, these values can be considered ordered...

Moreover, we can identify three types of questions that group “similar” categories:

1. *Entertainment questions*: music preferences, movie preferences, hobby & interests (63 items);
2. *Personality questions*: phobias, health habits, personality traits (etc.), spending habits (77 items);
3. *Demographics questions*: demographics (10 items).

For a detailed description of the survey’s questions, see the Appendix [A](#).

3.2 The Files

Inside the homework’s folder, excluding the notebook, the students can find two files:

- *responses_hw.csv*: the data file with all the responses of the people to the survey (each row is a person, each column is a question);
- *columns_hw.csv*: a file that show the original question in the survey and the column name in *responses_hw.csv*.

Some characteristics to be known concerning the dataset are;

- The data file *responses_hw.csv* consists of 1010 rows and 150 columns (139 integer and 11 categorical).
- The data contain missing values.

3.3 Exercises

Given the YPS dataset, the students have to use the Principal Component Analysis (PCA) to reduce the dimensionality of the problem and, then, have to identify meaningful clusters of people (if existing) using the k -Means algorithm.

Remark 3.1 (Simulating the typical situation in a company). *With this homework we want to simulate (for example) the situation where a company wants to identify meaningful profiles of their customers. Typically, all the information gathered with a customer survey must be summarized with few and easily-comprehensible concepts to help the managers with future decisions. Then, in these situations, very often the number m of chosen Principal Components (PCs) is very low, giving more importance to the dimensionality reduction than the preservation of the information.*

3.3.1 Simulation of a Real-World Problem

Considering Remark 3.1, let us assume that the YPS dataset is the result of a survey given to the customers of the company you are working for. Then, your manager asks you to:

1. Make the results of the survey easier to be interpreted. In particular, you must summarize the information in the dataset with at most $m = 5$ features but (if possible) with at least 33% of preserved information;
2. Identify customers profiles according to their entertainment preferences and/or their personality. Specifically, identify a minimum of 3 up to 10 customer profiles.

A Survey's Questions

A.1 Music Preferences

1. I enjoy listening to music.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
2. I prefer.: Slow paced music 1-2-3-4-5 Fast paced music (integer)
3. Dance, Disco, Funk: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
4. Folk music: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
5. Country: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
6. Classical: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
7. Musicals: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
8. Pop: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
9. Rock: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
10. Metal, Hard rock: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
11. Punk: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
12. Hip hop, Rap: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
13. Reggae, Ska: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
14. Swing, Jazz: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
15. Rock n Roll: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
16. Alternative music: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
17. Latin: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
18. Techno, Trance: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
19. Opera: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)

A.2 Movie Preferences

1. I really enjoy watching movies.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
2. Horror movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
3. Thriller movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
4. Comedies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
5. Romantic movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
6. Sci-fi movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
7. War movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
8. Tales: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
9. Cartoons: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
10. Documentaries: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
11. Western movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)
12. Action movies: Don't enjoy at all 1-2-3-4-5 Enjoy very much (integer)

A.3 Hobbies & Interests

1. History: Not interested 1-2-3-4-5 Very interested (integer)
2. Psychology: Not interested 1-2-3-4-5 Very interested (integer)
3. Politics: Not interested 1-2-3-4-5 Very interested (integer)
4. Mathematics: Not interested 1-2-3-4-5 Very interested (integer)
5. Physics: Not interested 1-2-3-4-5 Very interested (integer)
6. Internet: Not interested 1-2-3-4-5 Very interested (integer)
7. PC Software, Hardware: Not interested 1-2-3-4-5 Very interested (integer)
8. Economy, Management: Not interested 1-2-3-4-5 Very interested (integer)
9. Biology: Not interested 1-2-3-4-5 Very interested (integer)
10. Chemistry: Not interested 1-2-3-4-5 Very interested (integer)
11. Poetry reading: Not interested 1-2-3-4-5 Very interested (integer)
12. Geography: Not interested 1-2-3-4-5 Very interested (integer)
13. Foreign languages: Not interested 1-2-3-4-5 Very interested (integer)
14. Medicine: Not interested 1-2-3-4-5 Very interested (integer)
15. Law: Not interested 1-2-3-4-5 Very interested (integer)
16. Cars: Not interested 1-2-3-4-5 Very interested (integer)
17. Art: Not interested 1-2-3-4-5 Very interested (integer)
18. Religion: Not interested 1-2-3-4-5 Very interested (integer)
19. Outdoor activities: Not interested 1-2-3-4-5 Very interested (integer)
20. Dancing: Not interested 1-2-3-4-5 Very interested (integer)
21. Playing musical instruments: Not interested 1-2-3-4-5 Very interested (integer)
22. Poetry writing: Not interested 1-2-3-4-5 Very interested (integer)
23. Sport and leisure activities: Not interested 1-2-3-4-5 Very interested (integer)
24. Sport at competitive level: Not interested 1-2-3-4-5 Very interested (integer)
25. Gardening: Not interested 1-2-3-4-5 Very interested (integer)
26. Celebrity lifestyle: Not interested 1-2-3-4-5 Very interested (integer)
27. Shopping: Not interested 1-2-3-4-5 Very interested (integer)
28. Science and technology: Not interested 1-2-3-4-5 Very interested (integer)
29. Theatre: Not interested 1-2-3-4-5 Very interested (integer)
30. Socializing: Not interested 1-2-3-4-5 Very interested (integer)
31. Adrenaline sports: Not interested 1-2-3-4-5 Very interested (integer)
32. Pets: Not interested 1-2-3-4-5 Very interested (integer)

A.4 Phobias

1. Flying: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
2. Thunder, lightning: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
3. Darkness: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
4. Heights: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
5. Spiders: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
6. Snakes: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
7. Rats, mice: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
8. Ageing: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
9. Dangerous dogs: Not afraid at all 1-2-3-4-5 Very afraid of (integer)
10. Public speaking: Not afraid at all 1-2-3-4-5 Very afraid of (integer)

A.5 Health Habits

1. Smoking habits: Never smoked - Tried smoking - Former smoker - Current smoker (categorical)
2. Drinking: Never - Social drinker - Drink a lot (categorical)
3. I live a very healthy lifestyle.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

A.6 Personality Traits, Views on Life, and Opinions

1. I take notice of what goes on around me.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
2. I try to do tasks as soon as possible and not leave them until last minute.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
3. I always make a list so I don't forget anything.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
4. I often study or work even in my spare time.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
5. I look at things from all different angles before I go ahead.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
6. I believe that bad people will suffer one day and good people will be rewarded.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
7. I am reliable at work and always complete all tasks given to me.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
8. I always keep my promises.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
9. I can fall for someone very quickly and then completely lose interest.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
10. I would rather have lots of friends than lots of money.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
11. I always try to be the funniest one.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

12. I can be two faced sometimes.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
13. I damaged things in the past when angry.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
14. I take my time to make decisions.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
15. I always try to vote in elections.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
16. I often think about and regret the decisions I make.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
17. I can tell if people listen to me or not when I talk to them.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
18. I am a hypochondriac.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
19. I am emphatetic person.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
20. I eat because I have to. I don't enjoy food and eat as fast as I can.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
21. I try to give as much as I can to other people at Christmas.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
22. I don't like seeing animals suffering.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
23. I look after things I have borrowed from others.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
24. I feel lonely in life.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
25. I used to cheat at school.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
26. I worry about my health.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
27. I wish I could change the past because of the things I have done.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
28. I believe in God.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
29. I always have good dreams.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
30. I always give to charity.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
31. I have lots of friends.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
32. Timekeeping.: I am often early. - I am always on time. - I am often running late. (categorical)
33. Do you lie to others?: Never. - Only to avoid hurting someone. - Sometimes. - Everytime it suits me. (categorical)
34. I am very patient.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
35. I can quickly adapt to a new environment.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
36. My moods change quickly.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
37. I am well mannered and I look after my appearance.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
38. I enjoy meeting new people.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

39. I always let other people know about my achievements.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
40. I think carefully before answering any important letters.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
41. I enjoy company of children.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
42. I am not afraid to give my opinion if I feel strongly about something.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
43. I can get angry very easily.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
44. I always make sure I connect with the right people.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
45. I have to be well prepared before public speaking.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
46. I will find a fault in myself if people don't like me.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
47. I cry when I feel down or things don't go the right way.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
48. I am 100
49. I am always full of life and energy.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
50. I prefer big dangerous dogs to smaller, calmer dogs.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
51. I believe all my personality traits are positive.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
52. If I find something the doesn't belong to me I will hand it in.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
53. I find it very difficult to get up in the morning.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
54. I have many different hobbies and interests.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
55. I always listen to my parents' advice.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
56. I enjoy taking part in surveys.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
57. How much time do you spend online?: No time at all - Less than an hour a day - Few hours a day - Most of the day (categorical)

A.7 Spending Habits

1. I save all the money I can.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
2. I enjoy going to large shopping centres.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
3. I prefer branded clothing to non branded.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
4. I spend a lot of money on partying and socializing.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

5. I spend a lot of money on my appearance.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
6. I spend a lot of money on gadgets.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)
7. I will happily pay more money for good, quality or healthy food.: Strongly disagree 1-2-3-4-5 Strongly agree (integer)

A.8 Demographics

1. Age: (integer)
2. Height: (integer)
3. Weight: (integer)
4. How many siblings do you have?: (integer)
5. Gender: Female - Male (categorical)
6. I am: Left handed - Right handed (categorical)
7. Highest education achieved: Currently a Primary school pupil - Primary school - Secondary school - College/Bachelor degree (categorical)
8. I am the only child: No - Yes (categorical)
9. I spent most of my childhood in a: City - village (categorical)
10. I lived most of my childhood in a: house/bungalow - block of flats (categorical)