HÖGSKOLAN
I HALMSTAD

Project "Applied linear algebra" – 1,5hp
(Course MA8017 "Engineering Mathematics", 2024)

Examiner – Elena Haller

Introduction **What is it all about?**
You have passed Linear Algebra course in your early education. Now it is time to refresh knowledge and a nice way to do this is to apply Linear Algebra methods to problems arising in other fields of science/engineering (learning-by-doing).
So, this project covers basic concepts of linear algebra and how they can be used.

Purpose **Why do we do that?**
The major aim of the project is to prepare you for your future master research by imitating all steps of self-conducted problem analysis in its "mini-version".

Structure **What do I have to do?**
There are 2 problems to solve and for each of them you are supposed

- To present theoretical foundations of method you are using (corresponds to **background** section of your MS thesis)
- To apply them in Matlab/Python programming environment (corresponds to **method** section of your MS thesis)
- To present the results of your analysis in reader-friendly way (corresponds to **results** section of your MS thesis)
- To elaborate on limitations of your solution (corresponds to **discussion** section of your MS thesis)
-

Technical specifications **What is Elena going to grade?**
After completing the two problems (see below) you write ***1(one) pdf document*** that later you submit at bb.
The suggested structure of the document is as follows:

- Title page (hh-template must be used)
- Table of content
- Problem 1
  - Problem statement
  - Background (technical description of math-methods you want to use)
  - Method (pseudocode with comments on mathematics and references to the background)
  - Results (what you got from your script – graphs, coefficients etc. with text-descriptions)

    - Discussion (what are the limitations of the method you used, what are possible generalizations/improvements)
    - Appendix (full version of the code)
- Problem 2
    - Same subsections
- References

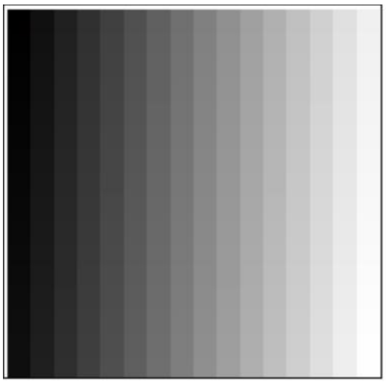
Formal info **When and who is going to do this?**

- This assignment is supposed to be done ***in groups of not more than 3 persons*** (not 4, not 5…)
  Please, fill in your groups in the following document:
  Group list (to click)

- ***One bb-submission per group***
- The deadline is ***December 30***
- Oral examination is based on your submitted reports and scheduled on **January 9&10**

Problem 1 “Image compression”

Images displayed on a computer screen are actually a collection of dots of color (pixels). This implies that any image is stored as a collection of matrices. You can change those matrices to make the computer display a different picture: changing how dark or bright something is or flipping an image upside down. Each image matrix contains intensity coefficients for one color. So, only one matrix is required for black-and-white (grayscale) images.



0	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240
1	17	33	49	65	81	97	113	129	145	161	177	193	209	225	241
2	18	34	50	66	82	98	114	130	146	162	178	194	210	226	242
3	19	35	51	67	83	99	115	131	147	163	179	195	211	227	243
4	20	36	52	68	84	100	116	132	148	164	180	196	212	228	244
5	21	37	53	69	85	101	117	133	149	165	181	197	213	229	245
6	22	38	54	70	86	102	118	134	150	166	182	198	214	230	246
7	23	39	55	71	87	103	119	135	151	167	183	199	215	231	247
8	24	40	56	72	88	104	120	136	152	168	184	200	216	232	248
9	25	41	57	73	89	105	121	137	153	169	185	201	217	233	249
10	26	42	58	74	90	106	122	138	154	170	186	202	218	234	250
11	27	43	59	75	91	107	123	139	155	171	187	203	219	235	251
12	28	44	60	76	92	108	124	140	156	172	188	204	220	236	252
13	29	45	61	77	93	109	125	141	157	173	189	205	221	237	253
14	30	46	62	78	94	110	126	142	158	174	190	206	222	238	254
15	31	47	63	79	95	111	127	143	159	175	191	207	223	239	255

Color images have 3 matrices – one each of RGB (red-green-blue) components.

Storing digital images requires large amounts of computer memory. Therefore, we always want to reduce the memory storage without losing too much information from the image, i.e. trying to preserve quality. Image compression is also required to communicate/stream images faster.

To do this, methods of linear algebra are used. There are several techniques to compress images, one of them is Principal Component Analysis (PCA) that can be performed using Singular Value Decomposition (SVD).

Questions for Problem 1:

1) [Background] Define PCA and SVD concepts in linear algebra and provide the mathematical relation between SVD and PCA; i.e. write the formula how they are related to each other.
2) [Problem formulation] Given the provided dataset of faces, apply PCA on them and find the principal components. Use the few components associated with large eigenvalues, and reconstruct faces using only these few components as a compression technique. How many do you need to not distinguish the visual difference?

P.S.

It is worth mentioning that in this case we need to save a set of eigen vectors (one set for all images). We do it once and we use it for compressing & reconstructing each image. As mentioned, it is needed only once and therefore the storage for that set is not counted.

For some guidelines, use the following source

https://www.geeksforgeeks.org/ml-face-recognition-using-eigenfaces-pca-algorithm/

Problem 2 ”Housing price estimation”
The market price *(y)* of any house depends on many factors $(x_1,x_2,\ldots,x_N)$ including location, view, surrounding etc. As more adequate price is set as smoother (and fairer) the selling process goes. Therefore, the goal is to predict/set the price of the house in accordance with those parameters. To do that the linear regression algorithm can be used.
Questions for Problem 2:

1) [Background] Define the least squares method for overdetermined systems and its connection to linear regression models.
2) [Problem formulation] Given the provided dataset program linear regression algorithm to determine the linear formula of price calculations/estimations/predictions:

$$y = a_1x_1 + a_2x_2 + \cdots + a_Nx_N$$

The dataset for this problem contains

- 506 instances
- 13 attributes (Obs! Both - numerical and categorical) for each instance:
    - CRIM per capita crime rate by town
    - ZN proportion of residential land zoned for lots over 25,000 sq.ft.
    - INDUS proportion of non-retail business acres per town
    - CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
    - NOX nitric oxides concentration (parts per 10 million)
    - RM average number of rooms per dwelling
    - AGE proportion of owner-occupied units built prior to 1940
    - DIS weighted distances to five employment centres
    - RAD index of accessibility to radial highways
    - TAX full-value property-tax rate per $10,000
    - PTRATIO pupil-teacher ratio by town
    - LSTAT % lower status of the population
    - Target (MEDV) Median value of owner-occupied homes in $1000’s