SANS-MIRI Homework 1: short programming exercises

Jorge Garcia Vidal, Jose M. Barcelo Ordinas and Pau Ferrer Cid

October 23, 2023

The course SANS (Statistical Analysis of Networks and Systems) belongs to the Master MIRI (Master of Innovation and Research in Computer Science) of the Faculty of Computer Science of Barcelona. The course is an introduction to some mathematical foundations used in data science. The course content includes an introduction to probability, linear algebra, and estimation.

This document reports Homework 1, which consists of solving short exercises related to basic concepts of algebra and probability. The objective is that for each exercise, the student writes some code (in Python), and obtains the results. Python libraries can be used when necessary.

We expect you to produce a PDF report setting out your main findings for each exercise. These can be lists, tables, graphs, etc. Students are expected not just to give the numerical result, but to explain what results they expect, justify the results and relate their conclusions to the theory. In each exercise, we mention the sections of the lecture notes related to the exercise.

Exercise 1 (LN Alg: section 1.7 and 1.8) *Let us assume the undetermined system and overdetermined systems represented by matrices* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *respectively:*

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 1 & 4 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 1 & 5 \\ 1 & 4 & 7 \\ 5 & 2 & 6 \\ 3 & 4 & 2 \end{bmatrix}$$

a) specify the row space, column space, null space and left space of matrices A *and* B*,*

b) obtain the right pseudoinverse for A *and the left pseudoinverse for matrix* B*,*

c) obtain the solution of the system $Ax=b$ *with* $b^\top=(2, 1)$*, and of* $Bx=c$ *with* $c^\top=(1, 2, 1, 5, 4)$*.*

Exercise 2 (LN Alg: section 2) *Let us assume the following matrices:*

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 1 & 2 & 1 & 3 \\ 3 & 1 & 0 & 2 \\ 2 & 3 & 4 & 5 \end{bmatrix} \qquad B = \begin{bmatrix} 5 & 0 & 2 & -1 \\ 0 & 2 & 0 & -1 \\ 2 & 0 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

$$C = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \qquad D = \begin{bmatrix} 4 & 3 \\ 3 & 1 \end{bmatrix}$$

a) *specify possitive definitness of each matrix,*

b) *investigate whether the matrices produced by the eigendecomposition are orthogonal matrices,*

c) *specify the projection of vector* $b^\top$*=(1, 2, 1, 5, 5) onto the subspace generated by matrix* E *and vector* $b^\top$*=(1, 2, 1, 5) onto the subspace generated by* F *(use these matrices only in this subsection),*

$$E = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 1 & 2 & 1 & 3 \\ 3 & 1 & 0 & 2 \\ 2 & 3 & 4 & 5 \\ 2 & 1 & 3 & 2 \end{bmatrix} \qquad F = \begin{bmatrix} 5 & 0 & 2 & -1 \\ 0 & 2 & 0 & -1 \\ 2 & 0 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

d) *take vectors* $b_1^\top$*=(1, 0),* $b_2^\top$*=(0, 1),* $b_3^\top$*=(-1, 0) and* $b_4^\top$*=(0, -1) and plot how these vectors are transformed using matrix* D*, i.e., d=D**b_i*. Repeat the process showing step by step the transformation using an eigendecomposition of* D*, i.e., show how the vector* b *is transformed by the eigenvector and eigenvalues matrices in the unit circle.*

Exercise 3 (LN Alg: section 3) *Solve the following short questions:*

a) *take matrices* A *and* B *of exercise 2 and obtain the trace(A), trace(B), trace(A+B), trace(AB), trace(BA),*

b) *obtain the Hadamard product of matrices* A *and* B*. e.e.,* $A \odot B$*,*

c) *Write in matricial form the following quadratic function:* $f(x)= x_1^2 + 2\, x_2^2 - 4\, x_3^2 + 2\, x_1 x_2 - 2\, x_1 x_3 + 2\, x_1 + 2\, x_2 - x_3 + 3$*. Check whether the matrix* $\boldsymbol{P}$ *of the quadratic form, i.e.,* $\boldsymbol{x}^\top \boldsymbol{P} \boldsymbol{x}$*, is positive definite, and thus this quadratic form is a convex function.*

Exercise 4 (LN Alg: section 4) *Let us assume matrices A and B:*

$$A = \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 1 & 4 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 1 & 5 \\ 1 & 4 & 7 \\ 5 & 2 & 6 \\ 3 & 4 & 2 \end{bmatrix}$$

a) *Obtain the SVD of matrices A and B, and check orthogonality of eigenvectors,*

b) *write economy, compact and truncated SVD representations,*

c) *obtain the rank-1 and rank-2 best matrix approximations,*

d) *write the l-1, l-2, l-∞ norms, and the spectral, Frobenious and nuclear norms,*

e) *obtain the condition number of the matrices*

Exercise 5 (LN Prob: section 3) *Prof the following combinatorial identities*

a) $\binom{n-1}{k-1} + \binom{n-1}{k} = \binom{n}{k}$.

b) $\sum_{i=r}^{n} \binom{i}{r} = \binom{n+1}{r+1}, \quad n > r$.

c) $\sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r}$. *(Hint: use $(1+x)^m(1+x)^n = (1+x)^{m+n}$ and equate coefficients).*

Exercise 6 (LN Prob: section 4) *We have a box with three red balls and three white balls. One friend of us randomly picks one of the balls of the box, but our friend does not tell us whether the ball is red or white. After this, the box has only 5 balls.*

a) *Now, we randomly pick 1 ball from the box. The ball is red. Which is the probability that the first removed ball was also red?*

b) *After this, we pick an additional ball from the box. Which is the probability that this new ball is red?*

Exercise 7 (LN Prob: section 6.1, 7 and 13) *Generate $M = 10000$ samples of a random variable X following the distributions listed below. In all cases, plot the graphic of the samples $X(m)$ for $m = 1...100$, and the histogram of the Cumulative Distribution Function (CDF) with the obtained M samples. For discrete random variables plot also the histogram of the Probability Mass Function (PMF). For the continuous random variables, plot the histogram of the Density Function using adequate bins. Estimate the expected value and variance and compare with the theoretical values.*

a) *Bernoulli with* $p = p\{X = 1\} = 3/4$,

b) *Binomial with* $n = 10$ *and* $p = 3/4$,

c) *Geometric with* $p = 3/4$ *(*X *taking values in* $\{0, 1, ...\}$*)*,

d) *Poisson with* $\lambda = 1$,

e) *Uniform in* $[0, 1]$,

f) *Exponential with* $\lambda = 1$,

g) *Univariate Gaussian with* $\mu = 1$, $\sigma^2 = 1$

h) *Univariate Gaussian with* $\mu = 1$, $\sigma^2 = 5$

i) *Beta with* $\alpha = 1$, $\beta = 5$

j) *Sum of independent samples following distributions in f) and h).*

Exercise 8 (LN Prob: section 6.1, 7, 9 and 13) *For a random variable* X *following the distributions listed below, generate* $M = 1000$ *independent samples of the random variable* $X^* = \frac{\sum_{k=1}^{N} X(k)}{N}$, *with* $N = 100$. *For the all cases, plot* $X^*(m)$ *for* $m = 1...M$ *and the histogram of the CDF of* X^*. *Estimate the average and variance of* X^*. *Relate the obtained results with the Law of Large Numbers (LLN).*

a) *Bernoulli, with* $p = p\{X = 1\} = 3/4$,

b) *Uniform in* $[0, 1]$,

c) *Exponential with* $\lambda = 1$, *and*

d) *Gaussian with* $\mu = 1$, $\sigma^2 = 1$.