# Advanced Matrix Algebra and Applications - Python Module

18-22 September 2019

## Problem Sheet - 2

1. An institution asks the audience of each of its courses to rate the course on a scale of 1 to 5. The following matrix $A$ consists of the ratings of 10 students for 4 courses.

$$A = \begin{bmatrix} 5 & & 2 & \\ & 4 & & 1 \\ & 2 & 5 & \\ & & 4 & 4 \\ & 3 & & 1 \\ 5 & & 1 & \\ 4 & & & 3 \\ 2 & & 4 & \\ 2 & & & 5 \\ 2 & & 4 & \end{bmatrix}$$

The entry $a_{ij}$ is the rating given by the $i^{th}$ student for the $j^{th}$ course. If a student hasn't taken a particular course, then the corresponding entry is left blank. Let us address the $i^{th}$ student as $S_i$. The 4 courses, in the order of appearance in the matrix, are Linear Algebra, Probability and Statistics, Compiler Design, and Computer Architecture. The average ratings for each course is 3.

Let $B$ be the matrix with the blank entries in $A$ being replaced by the average rating of the corresponding course.

Q1) Compute the 'best' low rank approximation of $B$. (retain 2 singular values)

Q2) Compute the spectral norm(2-norm) and the nuclear norm of $B$

  (i) using the numpy.linalg.norm method in Python.

  (ii) using SVD of $B$.

Q3) Give an orthonormal basis for the subspace of $\mathbb{R}^4$ spanned by the rows of $B$.

Q4) Give an orthonormal basis for the subspace of $\mathbb{R}^{10}$ spanned by the columns of $B$. Find the coordinates of the columns of $B$ with respect to this basis.

Q5) Find the eigenvalues and the corresponding eigenvectors of the matrices $B^TB$ and $BB^T$ without computing these matrices explicitly.

Q6) Complete the matrix $A$ as follows:
Let $L$ be the 'best'low rank approximation of $B$. If $a_{ij}$ is missing, then $a_{ij} := \frac{1}{4} \sum_{n \in N} l_{nj}$ where $N = \{\text{index of 4 nearest neighbours of } i^{th} \text{ row vector of } L \text{ w.r.t cosine distance}\}$.

Q7) Based on the completed matrix $A$, which course would you recommend to $S_5$ that he/she hasn't taken before?

2. Attendance percentages and marks scored by students in a course are recorded. Let $(x_i, y_i)$ be the ordered pair of the attendance percentage and the marks of the $i^{th}$ student. The data is as follows:

$\{(68, 72), (75, 75)], (87, 87), (74, 71), (61, 70), (95, 91), (74, 79), (67, 69), (81, 82), (83, 84)\}$

Suppose that the relationship between the two variables is only nearly linear, and the approximate linear relationship can be expressed as:

$$y = \beta_0 + \beta_1 x$$

Then the above set of equations have an equivalent matrix form as:

$$Y = X\beta$$

where $X = \begin{bmatrix} 1 & 68 \\ 1 & 75 \\ 1 & 87 \\ 1 & 74 \\ 1 & 61 \\ 1 & 95 \\ 1 & 74 \\ 1 & 67 \\ 1 & 81 \\ 1 & 83 \end{bmatrix}$, $Y = \begin{bmatrix} 72 \\ 75 \\ 87 \\ 71 \\ 70 \\ 91 \\ 79 \\ 69 \\ 82 \\ 84 \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta 1 \end{bmatrix}$.

Q1) Does $Y$ belong to the column span of $X$? Why(not)?

Q1) Compute the pseudoinverse of $X$.

Q3) Compute the 'best' choice of $\beta$ in the least squares sense.

Q4) Find the projection of $Y$ on the column space of $X$.

Q5) What is the minimum number of marks a student would get in the course, based on this model?

3. Suppose that for another course, the relationship between the two variables in Question no. 2 is nearly quadratic. The new data set is: $\{(68, 88), (75, 38)], (87, 73), (74, 64), (61, 48), (95, 62), (74, 62), (67, 72), (81, 48), (83, 39)\}$ Assume that the approximate quadratic relationship can be expressed as

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

The equivalent matrix form is:

$$Y = X\beta$$

where $X = \begin{bmatrix} 1 & 68 & 4624 \\ 1 & 75 & 5625 \\ 1 & 87 & 7569 \\ 1 & 74 & 5476 \\ 1 & 61 & 3721 \\ 1 & 95 & 9025 \\ 1 & 74 & 5476 \\ 1 & 67 & 4489 \\ 1 & 81 & 6561 \\ 1 & 83 & 6889 \end{bmatrix}$, $Y = \begin{bmatrix} 88 \\ 38 \\ 73 \\ 64 \\ 48 \\ 62 \\ 62 \\ 72 \\ 48 \\ 39 \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_0 \\ \beta 1 \\ \beta_2 \end{bmatrix}$.

Q1) Find the rank and condition number of $X$ and $X^T X$..

Q2) Compute the 'best' choice of $\beta$ in the least squares sense.

Q3) Suppose $Y = [84, 40, 63, 61, 50, 67, 65, 63, 44, 40]^T$. Compute the 'best' choce of $\beta$ for the new $Y$. Compute the norm of the difference between the two estimates of $\beta$.

Q4) Compute the largest eigenvalue and the corresponding eigenvector of $X^T X$.

4. The link graph for a certain collection of webpages is encoded in the following matrix:

$$A = \begin{bmatrix} 0.05 & 0.1 & 0.2 & 0.5 & 0.19 \\ 0.15 & 0.02 & 0.1 & 0.2 & 0.1 \\ 0.3 & 0.28 & 0.01 & 0.1 & 0.2 \\ 0.4 & 0.4 & 0.09 & 0.1 & 0.5 \\ 0.1 & 0.2 & 0.6 & 0.1 & 0.01 \end{bmatrix}$$

where $a_{ij}$ is the proportion of outlinks from page $i$ to page $j$.

Q1) Compute the largest eigenvalue and eigenvector of $A$ by Power method.

Q2) Does $A$ have an Eigenvalue Decomposition(Spectral Decomposition)? Why(not)?

Q3) Compute all the eigenvalues and the corresponding eigenvectors of $A$ and $A^T$.

Q4) The dominant eigenvector,$r$, of $A$ is known as the pagerank vector associated with the collection. $r_i$ denotes the long-run chances of a click on a random link leading to the page $i$. Rank the webpages based on the pagerank vector.

*****