SUPPORTIVE

22UCU303: MATHEMATICAL FOUNDATIONS FOR DATA ANALYSIS

CREDITS:4 MARKS: 40 + 60

SYLLABUS

Unit I

Linear Algebra

Vectors and Matrices - Addition and Multiplication – Bases - Linear independence and dependence of vectors – Orthogonal matrices - Rank of matrix - Finding Eigen values and eigen vectors of a matrix - Problems based on Rank – Nullity theorem.

Unit II Distances and Nearest Neighbors

Metrics - Lp Distances and their Relatives - Lp Distances - Mahalanobis Distance - Cosine and Angular Distance - KL Divergence - Distances for Sets and Strings - Jaccard Distance - Modeling Text with Distances - Bag-of-Words Vectors - k-Grams

Unit III

Optimization for Data Science

Unconstrained Multivariate Optimization - Gradient Descent Learning Rule - Multivariate optimization with equality constraint – Multi variate optimization with in equality constraint

Unit IV Cross validation

Cross validation – Multiple linear Regression modelling building and selection – Classification – Performance measures

Unit V

Principal Component Analysis

Data Matrices - Projections - SSE Goal - Singular Value Decomposition - Best Rank-k Approximation Principal Component Analysis - Computation of PCA Components - Reduction of Two dimension data set to one dimension – Drawing Graph for PCA

Text Book:

Text Book

1. **MATHEMATICAL FOUNDATIONS FOR DATA ANALYSIS**, JEFF M. PHILLIPS, 2018
   https://www.cs.utah.edu/~jeffp/M4D/M4D-v0.4.pdf

Reference Books:

1. **Introduction to Statistics and Data Analysis**, Third Edition, Roxy Peck, Chris Olsen, Jay Devore,

   https://www.spps.org/cms/lib/MN01910242/Centricity/Domain/859/Statistics%20Textbook.pdf

2. Introduction To Linear Algebra – Fifth Edition , By Gilbert Strang, 2016

E – Resources

https://youtu.be/MLaJbA82nzk

Unit I

1.1 Linear Algebra

1.1.1 Vectors and Matrices

For the context of data analysis, the critical part of linear algebra deals with vectors and matrices of real numbers

In this context, a vector $v = (v_1, v_2, \ldots, v_d)$ is equivalent to a point in R^d. By default a vector will be a column of d numbers (where d is context specific)

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ . \\ v_n \end{bmatrix}$$

but in some cases we will assume the vector is a row $v^T = [v_1 \quad v_2 \quad . \quad . \quad . \quad v_n]$.

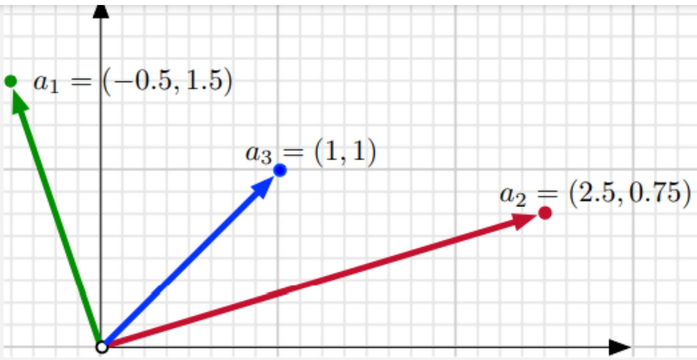
An $n \times d$ matrix A is then an ordered set of n row vectors $a_1, a_2, \ldots a_n$

$$A = [a_1; \ a_2; \ . \ . \ . \ a_n] = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} & . \ . \ . & A_{1,d} \\ A_{2,1} & A_{2,2} & . \ . \ . & A_{2,d} \\ A_{n,1} & A_{n,2} & . \ . \ . & A_{n,d} \end{bmatrix}$$ where the vector $a_i = [A_{i,1} \quad A_{i,2} \quad . \quad . \quad . \quad A_{i,d}]$ and $A_{i,j}$ is the element of the ith row and jth column. We can

We can write $A \in R^{n \times d}$ when it is defined on the reals.

1.1.2 Geometry of Vectors and numbers

Let A be an $n\times d$ matrix. Let $a_1, a_2, \ldots a_n \in R^d$. The vector is an arrow from the origin O = {0,0,0......0} to that point.



This picture with $n = 3$ points in $d = 2$ dimensions is equivalent to the 3×2 matrix representation

$$A = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -0.5 & 1.5 \\ 2.5 & 0.75 \\ 1 & 1 \end{bmatrix}.$$

1.1.3 Transpose of a Matrix.

Let A = [a_1; a_2; . . . a_n] = $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$ = $\begin{bmatrix} A_{1,1} & A_{1,2} & \cdot\cdot & \cdot & A_{1,d} \\ A_{2,1} & A_{2,2} & \cdot\cdot & \cdot & A_{2,d} \\ A_{n,1} & A_{n,2} & \cdot\cdot & \cdot & A_{n,d} \end{bmatrix}$ be a given matrix. Then A^T is defined as $A^T = [a_1 \quad a_2 \quad \cdot \quad \cdot \quad \cdot \quad a_n]$ = $\begin{bmatrix} A_{1,1} & A_{2,1} & \cdot\cdot & \cdot & A_{n,1} \\ A_{1,2} & A_{2,2} & \cdot\cdot & \cdot & A_{n,2} \\ A_{1,d} & A_{2,d} & \cdot\cdot & \cdot & A_{n,d} \end{bmatrix}$

1.1.4 Matrix - vector notation of Linear Equations

Consider the linear system of equations $3x_1 - 7x_2 + 2x_3 = -2$, $-1x_1 + 2x_2 - 5x_3 = 6$. Here number of equations is n = 2 and dimension is the number of variables x_1, x_2, x_3 d = 3. Matrix - vector notation of this system is Ax = b where $b = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$

A = $\begin{bmatrix} 3 & -7 & 2 \\ -1 & 2 & -5 \end{bmatrix}$ x = $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.