Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining & MA633 Data Mining Lab

Assignment-1

Date: 25-09-2020

1. Give examples of (a) vector space, (b) finite dimensional vector space (c) infinite dimensional vector space.

2. Consider the following data $\{x_i, y_i\}$: (i) $\{(1,-1),(2,11),(4,19)\}$ (ii) $x_1 = (2,3,2)^T, y_1 = 7, x_2 = (4,-5,5)^T, y_2 = 28, x_3 = (-3,7,-2)^T, y_3 = -19$
   (a) Write $f(x_i), i = 1,2,\ldots N$.
   (b) Express the data in the form $Xw = y$.
   (c) Find a set of vectors that spans the range of X.
   (d) Find the dimension of the range of X.
   (e) Find the hyperplane that generates the data and plot it for (i) by applying direct method.
   (f) Report the values of the parameters.
   (g) Find the projection of vector y onto the range space of X for (i).

3. Consider the following matrix:
$$X = \begin{bmatrix} 3 & 4 \\ 6 & 2 \end{bmatrix}$$
   (a) Find the rank of X. On the basis of rank of X, discuss the existence of the inverse of X.
   (b) Find the condition number of X.

4. Consider the following data:
$$w_0 + 2w_1 - w2 = 1$$
$$w_0 - w_1 + 2w_2 = 1$$
$$w_0 + 3w_1 c + cw_2 = 1$$

(a) Express $y = [1,1,1]^T$ as a linear combination of vectors. Are those vectors linearly independent? Is that possible to find the unknown parameters of the equation? Justify your answer.

(b) Find the following: no of attributes, no of data points. Write each x_i and corresponding y_i.

(c) Find the hyperplane that generates the data using least square regression, by applying the direct method.

5. Consider the following set of points (x_i, y_i): $\{(-1,0),(2,2),(4,3)\}$. Is that possible to find the hyperplane that generates the data? Justify your answer.

6. Give an example of monotonically increasing and decreasing function.

7. Find the projection of x onto $V = \text{span}\{u,v,w\}$ where $x = (-1,1,1,2)^T, u = (0,2,-2,1)^T, v = (1,3,0,1)^T, w = (1,1,1,1)^T$.

8. Explain the term: subspace of a vector space, span of a set of vectors. Check whether:

   (a) Range space of matrix transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ a subspace of $\mathbb{R}^m$.

   (b) All vectors $(v_1, v_2, v_3)^T$ in $\mathbb{R}^3$ with $v_1 - v_2 + 2v_3 = 0$ a subspace of $\mathbb{R}^3$.

   (c) All vectors $(v_1, v_2)^T$ in $\mathbb{R}^2$ with $v_1 \geq v_2$ a subspace of $\mathbb{R}^2$.

   (d) All $v = (v_1, v_2, v_3)^T \in \mathbb{R}^3$ with positive v_1, v_2, v_3.

   (e) All $v = (v_1, v_2, v_3)^T \in \mathbb{R}^3$ with positive $v_1 - v_2 + v_3 = k$ where k is a constant.

9. Consider $Xw = y$, where X is a 7×5 matrix.

   (a) Is X onto? Justify your answer.

   (b) What is the maximum dimension the range space of X can have? Justify your answer.

   (c) If it is possible to find the parameter w, what should be the rank of X? In this case, is X one-to-one? Justify your answer.

10. Find $\arg\min f(x)$ and $\min f(x)$where

   (a) $f(x) = x^3 + x^2 + 3x, x \in \mathbb{R}$

   (b) $f(x) = sinx + cosx, x \in \mathbb{R}$

   (c) $f(x) = \exp(||x||^2), x \in \mathbb{R}^2$

11. Find

   (a) ∇f at the point $x = (-1,1)^T$ where $f(x) = sin(cos(||x||^2))$

   (b) $\nabla_w f(w)$ where $f(w) = \exp(-||Xw - y||^2)$

12. Consider the following regression problem: $\{(x_i, y_i), i = 1,2,3,4\}$ where $x_1 = (-1,1)^T$, $x_2 = (0,2)^T$, $x_3 = (1,1)^T$, $x_4 = (2,1)^T$ and $y_1 = -1$, $y_2 = 1$, $y_3 = 2$, $y_4 = 0$. Let $Xw = y$ be the corresponding matrix equation. If it is possible to find a hyperplane that generates the data then

   (a) What should be the domain and range of X? Is X one one and onto?

   (b) What is the dimension of range of X?

   (c) Write a basis of the range of X.

   (d) Write the expression to find the projection of y onto the range of X.

13. Verify whether the following relations are functions. If so, check whether they are one-one/onto and write the domain, co-domain and range.

   (a) $f : \mathbb{R} \to \mathbb{R}$, where $f(x) = x^2$

   (b) $f : \mathbb{R} \to \mathbb{R}$, where $f(x) = (x+1)^2$

   (c) $f : [0,1] \to \mathbb{R}$, where $f(x) = x^2$

   (d) $f : [0,1] \to \mathbb{R}$, where $f(x) = (x+3)^2$

   (e) $f : \mathbb{R} \to \mathbb{R}$, where $f(x) = \sqrt{x}$

   (f) $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x) = \langle x, a \rangle, a \in \mathbb{R}^2$

   (g) $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, where $f(x,y) = \langle x, y \rangle$

   (h) $f : \mathbb{R}^3 \to \mathbb{R}$, where $f(x) = ||x||^2$

   (i) $f : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$, where $f(x,y) = \exp(||x-y||^2)$

   (j) $f : \mathbb{R}^2 \times \mathbb{R}^2 \to M(2,R)$, where $f(x,y) = xy^T$, where $M(2,R)$ is the set of all 2×2 real matrices.

14. Find the distance between $x_1^T = (1,1,-2)$ and $x_2^T = (7,-3,2)$ using Euclidean distance formula, norm and innerproduct expressions. (write all the relevant steps).

15. Write short notes on iterative techniques and gradient optimization approach.

Notes

- All the files related with the assignment should be saved in a single folder and send to sumitra@iist.ac.in.
- Last date of submission: 02-10-2020.
- **As far as assignments are concerned, students are expected to observe academic honesty and integrity. Though the students can collaborate and discuss, copying directly other students' assignment or allowing your own assignment to be copied constitute academic dishonesty and is highly discouraged.**