Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-1

Date: 13-08-2018

1. Describe the following concepts:
   (a) Vector space; span of a set of vectors; linear combination of vectors; linearly independent and dependent vectors; dimension of a vector space; basis of a vector space.
   (b) Give examples of (a) vector space, (b) finite dimensional vector space (c) infinite dimensional vector space.

2. Consider the following data: $\{(-1,0),(0,2),(1,4),(2,5)\}$
   (a) Express the data in the form $Xw = y$
   (b) Find a set of vectors that spans the range of X.
   (c) Find the dimension of the range of X.
   (d) Find the hyperplane that generates the data and plot it.
   (e) Report the values of the parameters.

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

4. 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 in $\mathbb{R}^3$ with $v_1 - v_2 + 2v_3 = 0$ a subspace of $\mathbb{R}^3$.
   (c) All vectors in $\mathbb{R}^2$ with $v_1 \geq v_2$ a subspace of $\mathbb{R}^2$. Here $[v_1, v_2, v_3]$ are components of a vector.

5. Analyze Data 1 and Data 2 using linear regression
   (a) Plot $J(w)$ against iteration for Data 1 & Data 2.
   (b) Plot $J(w)$ against w obtained in each iteration for Data 1.

(c) Plot the hyperplane that generates the data.

(d) Report the parameters of the model.

6. Verify whether the following relations are functions. If so, check whether they are one-one/onto.

   (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 $f(x) = \sqrt{x}$

7. Find the distance between $x_1^T = (0,-2)$ and $x_2^T = (-5,7)$ using euclidean distance formula, norm and innerproduct expressions. (write all the relevant steps).

8. Analyze Data 3 using linear regression

   (a) Find the parameters using direct method and iterative methods.

   (b) Compare the time taken and space requirements of two approaches.

   (c) Report the values of the hyperparameters and parameters of the model.

   (d) Analyze the performance of the model.

9. Analyze Data 4 using linear regression

   (a) Report the values of the parameters of the model.

   (b) Analyze the performance of the model

10. Analyze Wine quality data set using linear regression (download data from UCI web repository)

    (a) Specify the preprocessing methods applied on the data.

    (b) Apply k fold cross validation and hold out method.

    (c) Assess the performance of the model.

    (d) Report the values of the hyperparameters and the parameters of the model.

    (e) Apply batch as well as online optimization algorithms and compare their performance.

11. Write short notes on Gradient Descent.

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: 23-08-2018.
- **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.**