Applied Machine Learning **12-02-2025**

Lecture 1-4: Machine Learning Algorithms

Lecture: 1-4 *Student:*

READ THE FOLLOWING CAREFULLY:

Deadline for Assignment Submission:

11:59 PM, 01 March 2025 (strict deadline—no late submissions will be accepted).

- Assignments must be submitted via the **Taxila eLearn portal** using the provided submission link.
- Use a **Jupyter Notebook** for your solutions:
  - **For theoretical questions:** Solve them in a handwritten note and upload **clear images** of your solutions into the Jupyter Notebook.
  - **For coding/implementation tasks:** Write and execute your code directly in the notebook.
  - Ensure that **all images are properly displayed** in the Jupyter Notebook before submission.
- **Each answer must include the corresponding question number.**
- File naming format: `rollno_firstname_lastname_assignmentno.ipynb`

Failure to follow the guidelines may result in penalties.

-3.1 Assignments

-3.1.1 Programming Questions

Consider the Iris dataset. The dataset is available here: `https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html`.

1. Write a small paragraph describing the Iris dataset. 2 Marks
2. Identify the features/ attributes in Iris dataset? 4 Marks
3. Identify the total number of classes in Iris dataset? 3 Marks
4. In a table, summarize the total data instances of each class (Remember table and figure should have self contained appropriate captions.) 3 Marks
5. Split the Iris dataset randomly into training (80%) and testing (20%) (you can use sklearn train-test split - randomseed= 42) 2 Marks.
6. In a table, provide the number of data instances used for training and testing for each class. 2 Marks

7. Using the train data (obtained after splitting the total data into training and testing), perform three fold crossvalidation to find the best value of k in k Nearest Neighbour classifier (the k value can range from 1 to 25, and use euclidean norm to compute the distance). (You can use the k-fold crossvalidation package provided in sklearn for hyperparameter tuning - `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html`). 5 Marks

8. Plot the average macro f1-score obtained using three fold crossvalidation with respect to the different values of k considered in three fold crossvalidation. 3 Marks

9. Identify the best value of k for which you get the peak performance in three fold crossvalidation. 2 Marks

10. Using the best value of k, evaluate the performance of the k nearest neighbour classifier on the testdata (Remember testing should be done only once!). 2 Marks

11. Report the test accuracy, precision, recall, f1-score and macro f1-score. 4 Marks

-3.1.2 Vector Space

12. Define the following (Refer to chapter 3 of the book: Introduction to Linear Algebra (Fifth Edition) by Prof. Gilbert Strang) :

    - Vector Space. 1 Mark
    - Column Space of a Matrix A. 1 Mark
    - Row Space of a Matrix A. 1 Mark
    - Right Null Space of a Matrix A. 1 Mark
    - Left Null Space of a Matrix A. 1 Mark
    - Dimension of a Vector Space. 1 Mark
    - Basis set of a Vector Space. 1 Mark
    - Rank of a Matrix A. 1 Mark
    - $L2$ norm of a vector x. 1 Mark

    Fill in the blanks:

13. $Ax = b$ has a solution when b lies in ______ space of A. 1 Mark

14. Two nonzero vectors are orthogonal when their ______ is ______. 2 Marks

15. Two nonzero vectors are orthonormal when their dot product is ______ and the L2 norm of two vectors are ______ respectively. 2 Marks

16. Consider matrices A of size $m \times n$and $B = [A \quad A]$ of size $m \times 2n$ (repeated A twice). A and B has same ______ space and ______ space. 2 Marks

17. Are the following statements True or False? Justify or give examples to support your reasoning.

    - Orthogonality of two nonzero vectors implies linear independence. 2 Marks
    - Linear independence of two vectors implies orthogonality. 2 Marks
    - Dimension of row space and column space of an $m \times n$ matrix A are same. 2 Marks
    - Row rank and Column rank of an $m \times n$ matrix A are same. 2 Marks

- If two $m \times n$ matrices A and B have the same row space, column space, right null space and left null space, then $A = B$. 2 Marks

18. For the given matrix A, find the basis set for column space and row space. Also geometrically depict the basis set that spans the column space. 5 Marks

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

-3.1.3 Programming Question

19. Create a random 5×4 matrix A with rank 2 and a 5×1 vector b such that $Ax = b$ has infinite solution. Write the python code and also generate infinite solutions using loop. 5 Marks

20. Create a 3×4 matrix with rank 3, check whether right null space and left null space exist. Comment. Write a python code to verify. 2 Marks

21. Is it possible to create a no solution case for the above question. Justify if Yes or No. 1 Mark

22. Write a python code for generating ten b vectors such that $Ax = b$ has no solution. The matrix A is given below. 5 Marks

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 5 & 8 & 11 & 14 \\ 3 & 5 & 7 & 9 \end{bmatrix}$$

-3.1.4 Linear Regression using Least Squares

23. Mathematically derive the matrix formulation for linear regression. 2 Marks

24. Does the following system of linear equations $Ax = b$ has a solution? If it does not have a solution can you find an approximate solution using the following: 1 Marks

    - Method of least squares (you can use python for this) and justify why the system of linear equations does not have a solution. 2 Marks

    The system of linear equations $Ax = b$ is as follows:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

25. For the data (data.txt) attached in the email find the following using python: 2 Marks

    - Find a line that best fit the data with minimum error (sum of squares). [Don't use inbuilt code in python].
    - Find a second degree, third degree and fourth degree polynomial that fits the data respectively. Also find the error in each case and note down your inference. ([Don't use inbuilt code in python]. Refer the slides for help). 2 Marks