**Department: Computer Science and Engineering**
**Course Name: Machine Learning** **Code: CS 564**
**Full Marks-60** **Time: 2 hours**

**Make reasonable assumptions as and whenever necessary. Answer the questions to any sequence but answers to all the parts of any question should appear together. Marks will be deducted if this is not followed properly.**

1. You are given a data set on cancer detection. You've built a classification model and achieved an accuracy of 96%. Why shouldn't you be happy with your model performance? What can you do about it? 4 marks
2. What do you understand by Bias Variance trade off? What is leave-one-out cross-validation? Mention a situation where leave-one-out cross-validation is more preferred over k-fold cross-validation. 3+3+2= 8 marks
3. What is regularization? What is the error function used in logistic regression. Mathematically derive the update equations of logistic regression. 2+ 3+ 4=9 marks
4. Write short notes on the following:
   a) Cluster validity index
   b) Filter vs Wrapper based feature selection strategies
   c) Bidirectional feature selection
   d) LDA

   4 X 3= 12 marks
5. (a) What are the characteristics of noise points in DBSCAN?
(b) Assume you apply DBSCAN to the same dataset, but the examples in the dataset are sorted differently. Will DBSCAN always return the same clustering for different orderings of the same dataset? Give reasons for your answer
(c) K-Means does not explicitly use a fitness function. What are the characteristics of the solutions that K-Means finds --- which fitness function does it implicitly minimize?
(d) In general, K-means is limited to find clusters having complex shapes. What could be done to enable K-means to find clusters in arbitrary shapes (e.g. consider a post processing method)? 2+ 4+ 2+3=11 marks
6. Consider the following data set:

| | X1 | X2 |
|---|---|---|
| A | 1 | 1 |
| B | 2 | 1 |
| C | 2 | 2 |
| D | 6 | 4 |
| E | 6 | 5 |
| F | 7 | 5 |
| G | 10 | 3 |
| H | 10 | 2 |

Apply single linkage agglomerative algorithm. Show the distance matrix, in each step of the algorithm. Also show the dendrogram.

6+2 = 8 marks

7. Apply PCA on the following data:

| X1 | X2 |
|---|---|
| 2 | 1 |
| 3 | 4 |
| 5 | 0 |
| 7 | 6 |
| 9 | 2 |

(a) Draw the original objects with respect to descriptors X1 and X2 (b) plot the new data along the principle axes.

2+6= 8 marks