



**Jahangirnagar University**  
**Department of Computer Science and Engineering**  
4<sup>th</sup> Year 2<sup>nd</sup> Semester B.Sc. (Hons.) Final Examination -2020

Course Title: Pattern Recognition and Machine Learning  
Time: 3 Hours

Course No: CSE - 451  
Full Marks: 60

[Answer each of the following questions. Each question carries equal marks. Figures in the right margin indicate marks.]

1.

**Section-I (CO1)**

(Answer *all of them*)

- a) *State* the differences between supervised and unsupervised learning method. 2
- b) *List* out the issues that may arise in Machine Learning. 2
- c) *Mention* the procedure for dealing with missing data. 2
- d) *Tabulate* the differences between binary-class and multi-class. 2
- e) *Define*: i) Pattern Recognition      ii) Machine Learning 2
- f) *Define*: i) Support Vector Machine (SVM)      ii) Kernel Machine 2

2.

**Section-II (CO2)**

(Answer *any 3 out of 4*)

- a) *Explain* the learning process with some example. Illustrate the differences among classification, regression and clustering learning processes. 4
- b) In normal messages, the set of words {dear, friend, lunch, money} comes as of {8, 5, 3, 1} times. and in spam messages, the set of words {dear, friend, lunch, money} comes as of {2, 1, 0, 4} times. If a message "lunch money money money money" comes, *outline* the problems that may face by naïve Bayes algorithm. 4
- c) *Illustrate* the splitting rules used in decision tree algorithm mentioning the concept of pure data set and impure data set. 4
- d) *Give example* of overfitting in decision tree algorithm and overfitting avoidance rules. 4

3.

**Section-III (CO3)**

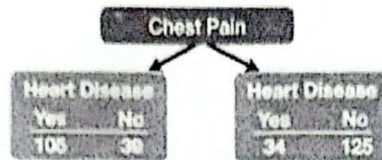
(Answer *any 3 out of 4*)

- a) In normal messages, the set of words {dear, friend, lunch, money} comes as of {8, 5, 3, 1} times. and in spam messages, the set of words {dear, friend, lunch, money} comes as of {2, 1, 0, 4}

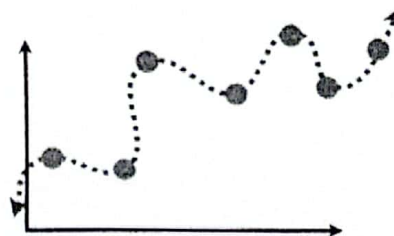
we cannot  
much  
es of

times. When a message comes up with a couple of word "dear friend", demonstrate how naïve Bayes algorithm classify them correctly.

- b) **Determine** the Gini impurity for chest pain for the decision sub-tree below. 4



- c) **Interpret** the functions of squiggly line in linear regression. 4



- d) **Examine** whether a squiggly line is overfit or not. 4

#### Section-IV (CO4)

(Answer any 3 out of 4)

4. a) **Deduce** a cross validation technique with the help of logistic regression algorithm, K-nearest neighbor algorithm and support vector machine. 4
- b) **Devise** some techniques by which we can select an optimum  $K$  value of the  $K$ -nearest neighbor algorithm for data classification. 4
- c) **Analyze** the functions of a covariance matrix in Principle Component Analysis (PCA). 4
- d) For the following data set, **calculate** the data points for a linear regression line, where  $x$  is independent variable and  $y$  is dependent variable. 4

$x$	$y$
1	3
2	4
3	2
4	4
5	5

### Section-V (CO5)

(Answer *any 2 out of 3*)

5. a) *Justify* that the 10-fold cross validation technique is better than that of 2/3-fold cross validation techniques. 6
- b) Take a look at the figure below. 6



*Choose* an appropriate Support Vector Machine (SVM) that can classify the above data points.

- c) Given a confusion matrix with multiple classes C1, C2, ...C6. 6

Class	C1	C2	C3	C4	C5	C6
C1	62	10	7	0	0	1
C2	10	60	6	2	0	2
C3	5	6	46	11	2	10
C4	0	2	8	65	0	5
C5	0	1	11	10	57	1
C6	1	3	1	1	10	64

*Evaluate* the performance of that model in terms of FPR, FNR, F1 Score, Precision, Sensitivity, Specificity and Accuracy of the model including details calculation and formulas.



**JAHANGIRNAGAR UNIVERSITY**  
Department of Computer Science and Engineering  
4<sup>th</sup> Year 2<sup>nd</sup> Semester B.Sc. (Honors) Final Examination 2021

Course Code: CSE-451  
Time: 3 hrs

Title: *Pattern Recognition and Machine Learning*  
Full Marks: 60

*Answer each of the following questions.*

1.

Section – I (CO1)  
(Answer *all* of the following questions.)

- (a) Define (*any two*): i) Pattern ii) Machine learning iii) Activation Function 2  
iv) Confusion matrix
- (b) What are the key benefits of machine intelligence over human intelligence? 2
- (c) State the differences between *clustering* and *classification*. 2
- (d) What is a class imbalanced dataset in machine learning? 2
- (e) Define (*any two*): i) Deep Learning ii) Ensemble method 2  
iii) Decision tree iv) HOG feature
- (f) Tabulate the differences between *logistic regression* and *linear regression*. 2

2.

Section – II (CO2)  
(Answer *any three* out of *four*.)

- (a) How can you identify a High Variance (Low Bias) model? How can you fix it? 4
- (b) Summarize the steps of PCA (*Principal Component Analysis*) algorithm. 4
- (c) Explain the concept of '*Kernel trick*' in machine learning with an example. 4
- (d) Demonstrate the process of finding the best fit line using the least square method 3+1=4  
given the following data:  
 $(x_1, y_1) = (1, 1), (x_2, y_2) = (2, 4), (x_3, y_3) = (3, 4)$ ,  
and also find the predicted value for  $x=4$ .

3.

Section – III (CO3)  
(Answer *any three* out of *four*.)

- (a) Determine the *eigenvalues* and *eigenvectors* of the following matrix: 4  
$$A = \begin{pmatrix} -6 & 3 \\ 4 & 5 \end{pmatrix}$$
- (b) Construct a plan to utilize logistic regression as a classifier beside regression. 4
- (c) Draw and interpret the *McCulloch-Putts model* of a neuron. 4
- (d) Assume, you want to cluster 7 observations into 3 clusters using K-Means clustering algorithm. After first iteration clusters, C1, C2, C3 has following observations: 4  
C1: {(2,2), (4,4), (6,6)} C2: {(0,4), (4,0)} C3: {(5,5), (9,9)}  
What will be the cluster centroids if you want to proceed for second iteration?



Section – IV (CO4)  
(Answer *any three* out of *four*.)

12. (a) Distinguish between *conventional machine learning* and *deep learning* algorithm. 4
- (b) Draw and analyze the *CNN (Convolutional Neural Network) architecture*. 4
- (c) Discuss the difference and tradeoff between two factors of bias and variance. How these factor play a role in machine learning models. Explain briefly. 4
- (d) List the steps of ADABOOST algorithm. 4

5.

Section – V (CO5)  
(Answer *any two* out of *three*.)

- (a) Explain the process of constructing a decision tree using ID3 algorithm (use your own dataset). 6
- (b) Given that  $k = 3$ , use the K-Means algorithm to cluster the following 6 records in Table 1 into 3 clusters. Suppose that the initial seeds (centers of each cluster) are E02, E04 and E06. Run the k-means algorithm for 1 epoch only and show the following: 6

EmployeeID	YearService	Income (K)
E01	4	9
E02	8	4
E03	2	10
E04	5	8
E05	6	4
E06	7	5

**Table 1**

- i) Calculate the Euclidean distances between each point and the cluster centers.  
 ii) Determine the new clusters (i.e. the examples/points belonging to each cluster)  
 iii) Determine the centers of the new clusters.  
 iv) Show the clusters after the first epoch and the new centroids.
- (c) Determine, using SVM (*Support Vector Machine*) algorithm, the equation of the hyperplane, give the following support vectors: 6
- $$S_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, S_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, S_3 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$
- where  $S_1, S_2$  belong to negative class and  $S_3$  belongs to positive class.

-----XXXXX-----