Course Code:	CST414-2	

Seventh Semester B.E. (Computer Science and Engineering) Examination

Machine Learning (Elective-II)

Time: 02 Hours]	[Max. Marks: 40

Instructions to Candidates:

1. All Questions carry marks as indicated.
2. Assume suitable data wherever necessary.

<table>
<tr><th colspan="2">Que.</th><th>Description of Question</th><th>Marks</th><th>CO</th></tr>
<tr><td>1</td><td>(a)</td><td>Consider the dataset given below for classification of tree as Oak and pine. Use ID3 algorithm to construct the full decision tree.
<table>
<tr><th>Example</th><th>Density</th><th>Grain</th><th>Hardness</th><th>Class</th></tr>
<tr><td>Example #1</td><td>Heavy</td><td>Small</td><td>Hard</td><td>Oak</td></tr>
<tr><td>Example #2</td><td>Heavy</td><td>Large</td><td>Hard</td><td>Oak</td></tr>
<tr><td>Example #3</td><td>Light</td><td>Large</td><td>Soft</td><td>Oak</td></tr>
<tr><td>Example #4</td><td>Light</td><td>Large</td><td>Hard</td><td>Pine</td></tr>
<tr><td>Example #5</td><td>Heavy</td><td>Small</td><td>Soft</td><td>Pine</td></tr>
<tr><td>Example #6</td><td>Heavy</td><td>Large</td><td>Soft</td><td>Pine</td></tr>
</table></td><td>07</td><td>CO1</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>2</td><td>(a)</td><td>What do you mean by instance based learning? Also write the drawbacks of K-NN algorithm.Apply K-NN for K=3 and classify the new query instance.
<table>
<tr><td>Fruit</td><td>Sweetness</td><td>Sourness</td><td>Fruit_Type</td></tr>
<tr><td>Lemon</td><td>1</td><td>9</td><td>Sour</td></tr>
<tr><td>Grapefruit</td><td>2</td><td>8</td><td>Sour</td></tr>
<tr><td>Orange</td><td>3</td><td>7</td><td>Sour</td></tr>
<tr><td>Raspberry</td><td>2</td><td>8</td><td>Sour</td></tr>
<tr><td>Cherry</td><td>6</td><td>4</td><td>Sweet</td></tr>
<tr><td>Banana</td><td>9</td><td>1</td><td>Sweet</td></tr>
<tr><td>Grapes</td><td>8</td><td>2</td><td>Sweet</td></tr>
<tr><td>Watermelon</td><td>9</td><td>1</td><td>Sweet</td></tr>
<tr><td>Avocado</td><td>1</td><td>1</td><td>None</td></tr>
<tr><td>Strawberry</td><td>5</td><td>5</td><td>Sour</td></tr>
<tr><td>Fig</td><td>7</td><td>3</td><td>???</td></tr>
</table></td><td>04</td><td>CO2</td></tr>
<tr><td></td><td>(b)</td><td>Prove that the Conjunction of Boolean literal is PAC learnable.
For example, if a consistent learner attempts to learn a target concept described by conjunctions of up to 10 boolean literals, and we desire a 94% probability that it will learn a hypothesis with error less than 0.2.Justify the statement.</td><td>03</td><td>CO3</td></tr>
<tr><td>3</td><td>(a)</td><td>Apply Delta learning algorithm for 2-input NAND classification up to four iterations. Take initial weight vector w=[0.3 0.4 0.3] ^{T}.Use bipolar sigmoid activation function and take eta=0.5 [Perform one epoch]</td><td>07</td><td>CO3</td></tr>
</table>

<table>
<tr><td>4</td><td>(a)</td><td>Derive the mathematical formulation of the Naïve Bayes classifier.
Write your assumptions clearly. You are given the following set of training examples.
<table><tr><th>A1</th><th>A2</th><th>A3</th><th>Class</th></tr><tr><td>a</td><td>c</td><td>a</td><td>C1</td></tr><tr><td>c</td><td>a</td><td>c</td><td>C1</td></tr><tr><td>a</td><td>a</td><td>c</td><td>C2</td></tr><tr><td>b</td><td>c</td><td>a</td><td>C2</td></tr><tr><td>c</td><td>c</td><td>b</td><td>C2</td></tr></table>How would a naïve bayes classifier classify the example :
{ A1=a , A2=c . A3=b} Class=??</td><td>06</td><td>CO3</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>(a)</td><td>What do you mean by K-means clustering? List the drawbacks of k-means clustering algorithm.
Assume the following dataset is given:
(2,2),(4,4),(5,5),(6,6),(8,8),(0,4)(4,0).
K-means is run with k=3, to cluster the dataset. Use Euclidean distance measure. K-means initial clusters C1,C2,and C3 are as follows:
C1:{(2,2),(4,4),(6,6)}
C2:{(0,4),(4,0)}
C3:{(5,5),(8,8)}</td><td>06</td><td>CO4</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>6</td><td>(a)</td><td>What do you mean by Ensemble Learning? Differentiate between Bagging and Boosting techniques? Why ensemble works?</td><td>03</td><td>CO4</td></tr>
<tr><td></td><td>(b)</td><td>Find α + and α - of linear Support Vector machine for following data points, Also determine weight and bias parameters.

Positive examples:{(-1,2),(-3,1),(-3,3),(-4,3),(-1,-1)}
Negative examples:{(1,1),(3,-2),(4,-3),(3,-4),(4,-1)}

Graph paper is not required just plot the points by making axis in answer sheet using scale.</td><td>04</td><td>CO4</td></tr>
</table>