

**Tribhuvan University**  
**Institute of Science and Technology**  
**SCHOOL OF MATHEMATICAL SCIENCES**  
**First Assessment 2079**

**Subject: Applied Machine learning**

**Full Marks: 45**

**Course No: MDS 552**

**Pass Marks: 22.5**

**Level: MDS /I Year /II Semester**

**Time: 2hrs**

*Candidates are required to give answer in their own words as far as practicable.*

**Attempt All Questions**

**Group A       $[5 \times 3 = 15]$**

1. Why data normalization is important? Explain Min-max and Z-score data normalization.
2. Discuss the concept of model overfitting and underfitting.
3. Discuss the concept of Bagging in ensemble learning.
4. How K-Means++ differs from K-Means? Explain.
5. Why SVD is used? Discuss its working.

**Group B       $[5 \times 6 = 30]$**

6. Derive weigh update rule for logistic regression using gradient descent.

**OR**

Discuss the concept of locally weighted linear regression. Consider a query point  $x=5$  and let  $x^1=6$ ,  $x^2=4$ , and  $x^3=3$  are three points in the training set. Find Cost function for the locally weighted linear regression.

7. Consider the following dataset. Find Information of gain of each attribute and draw decision tree for first iteration.

<b>Weather</b>	<b>Temperature</b>	<b>Wind Level</b>	<b>Go Out (Class)</b>
Sunny	High	Low	No
Sunny	Normal	Normal	Yes
Cloudy	High	Normal	No
Cloudy	Normal	High	Yes
Sunny	Normal	High	No
Rainy	High	Normal	No
Rainy	Low	High	No

8. Continue question number 7 and construct complete decision tree. Then, predict class label for the tuple {Cloudy, Low, Normal}.
9. When DBSCAN algorithm gives better performance than other clustering algorithms? Discuss its working.

**OR**

Divide the data points  $\{(2,10), ((2,5), (8,4), (5,8), (6,4)\}$  into two clusters using agglomerative clustering.

10. What are the features of principle components? Discuss the working of PCA algorithm.

\*\*\*

**Tribhuvan University**  
**Institute of Sciences and Technology**  
**SCHOOL OF MATHEMATICAL SCIENCES**  
**Second Assessment 2079**

**Subject: Applied Machine Learning**

**Course No: MDS 552**

**Level: MDS /I Year /II Semester**

*Candidates are required to give their answer in their own words as far as practicable.*

**Full Marks: 45**

**Pass Marks: 22.5**

**Time: 2 hrs**

**Group A [5 × 3 = 15]**

1. Discuss about sensitivity and specificity metrics.
2. Explain  $R^2$  metrics.
3. How dynamics of MDP proceeds? Explain.
4. Discuss value iteration algorithm briefly.
5. Discuss deterministic model of neuron.

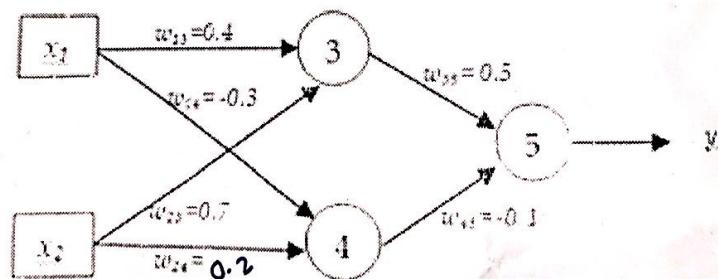
**Group B [5 × 6 = 30]**

6. What is perceptron? Derive weight update rule for perceptron.

**OR**

State XOR problem Can we realize XOR function using perceptron? If yes, realize it ✓ using perceptron. Otherwise, realize XOR function using MLP.

7. Consider a MLP given below. Let the learning rate be 0.7. The initial weights of the network are shown in the MLP. Assume that first training tuple is  $(1, 0)$  and its target output is 1. Calculate weight updates by using back-propagation algorithm. Assume logistic activation function.



8. Discuss working of RNN with suitable block diagram and mathematical formulation.
9. How soft clustering methods differs from <sup>hard</sup> soft clustering methods? Discuss working of EM algorithm for Gaussian mixture models.
10. Discuss working of hierarchical clustering algorithms.

**OR**

Divide the data points  $\{(2,2), (4,3), (4,8), (6,6), (2,7), (3,1)\}$  into two clusters using k-medoid algorithm.

\*\*\*

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 First Assessment 2080

**Subject:** Applied Machine Learning  
**Course No:** MDS 552  
**Level:** MDS / I Year / II Semester

**Full Marks:** 45  
**Pass Marks:** 22.5  
**Time:** 2hrs

*Candidates are required to give their answers in their own words as far as practicable.  
 Attempt ALL questions*

**Group A [5×3=15]**

1. How supervised learning is a costly approach compared to unsupervised?
2. Differentiate between batch and stochastic gradient descent.
3. Describe precision and recall as evaluation metrics for classification task.
4. Write an algorithm for K-Means Clustering.
5. Explain Soft Max Regression over Logistic Regression with suitable example.

**Group B [5×6=30]**

6. Derive the form of Logistic Regression for classification tasks.
7. Explain the various form of evaluation metrics used for regression.
8. What do you understand by ensemble learning? Explain its types.

**OR**

- Consider the following data points  $(1, 0), (0, 1), (-1, 0), (0, -1)$  as negatively labelled data and  $(8, 10), (10, 8), (12, 10), (10, 12)$  as positively labelled data. Determine the equation of hyper plane that divides the above data points into two classes using SVM and predict in which class  $(8, 6)$  belongs.
9. Given a confusion matrix from Iris flower classification as below, compute f1 score and interpret the result.

Confusion Matrix		
		Actual
Predicted	versicolor	2
	virginica	0
setosa		4
	versicolor	5
virginica		3
setosa	versicolor	0
virginica	versicolor	0

10. Execute hierarchical clustering in the provided data as below where rows and columns contains the point and values are the distances between the point. Also, draw dendrogram to illustrate diagrammatically.

Points	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	5
5	11	10	2	5	0

**OR**

Briefly describe how DBSCAN algorithm works. Apply DBSCAN algorithm to the above given data.

\*\*\*

**Tribhuvan University**  
**Institute of Sciences and Technology**  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 First Reassessment 2080

**Subject:** Applied Machine Learning

**Course No:** MDS 552

**Level:** MDS / I Year / II Semester

*Candidates are required to give their answers in their own words as far as practicable.  
Attempt ALL questions*

**Full Marks:** 45

**Pass Marks:** 22.5

**Time:** 2hrs

- Group A [5x3=15]**
1. Explain in brief the concept of Reinforcement Learning.
  2. Describe how gradient descent is used as optimization algorithm.
  3. What do you mean by cross validation? How is it useful in obtaining better model during machine learning?
  4. Explain the concept of Gaussian Mixture algorithm.
  5. Describe the random forest as machine learning algorithms.

**Group A [5x6=30]**

6. Derive the form of weight update rule for linear regression.
7. Explain the concept of Support Vector Machine. Describe different kernels used in Support Vector Machine with their mathematical form?

**OR**

Explain over fitting and under fitting in machine learning with suitable example. How can we mitigate them.

8. Explain how bagging and boosting are useful in obtaining efficient machine learning model.

9. Given a confusion matrix from Iris flower classification as below, compute precision and recall for each class and interpret the result.

Confusion Matrix		
		Predicted
Actual	setosa	2
	versicolor	0
virginica	setosa	4
	versicolor	2
virginica	virginica	3
		5

10. Execute hierarchical clustering in the provided data as below where rows and columns contains the point and values are the distances between the point. Also, draw dendrogram to illustrate diagrammatically.

Points	1	2	3	4	5
1	0	9	3	6	11
2	9	0	7	5	10
3	3	7	0	9	2
4	6	5	9	0	5
5	11	10	2	5	0

**OR**

Briefly describe how DBSCAN algorithm works. Apply DBSCAN algorithm to the above given data.

\*\*\*

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 Second assessment 2080

**Subject:** Applied Machine Learning  
**Course No:** MDS 552  
**Level:** MDS / I Year / II Semester

**Full Marks: 45**  
**Pass Marks: 22.5**  
**Time: 2hrs**

*Candidates are required to give their answers in their own words as far as practicable.*

**Attempt ALL questions**

**Group A [5×3=15]**

1. Differentiate between K-Means and DBSCAN algorithm.
2. Explain the mathematical model of neuron.
3. What do you mean by exploration and exploitation in Reinforcement Learning? How are they important for agent?
4. Considering grey scale image of size 6 x 6, apply convolution filter of edge detection as below:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

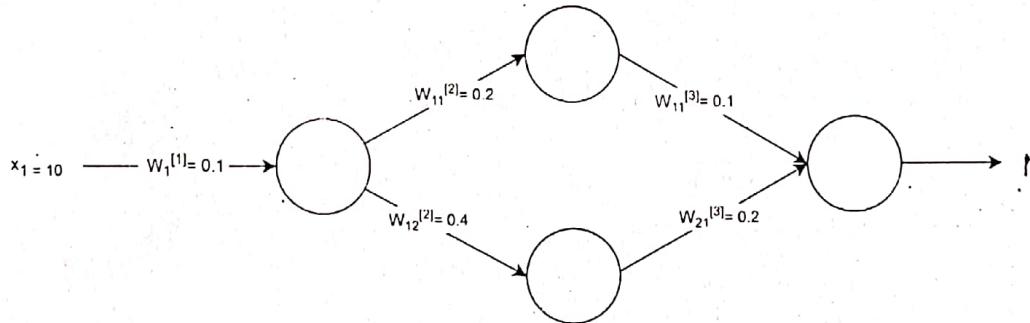
5. Explain the Markov Decision Process used in RL.

**Group B [5×6=30]**

6. What do you mean by dimensionality reduction? How does PCA works? Explain.

OR

7. Explain the Expectation Maximization algorithm for Gaussian Mixture Models.
8. Given a neural network, perform forward propagation considering a classification problem and compute the loss. Consider a sigmoid activation function and learning rate 0.1.



9. Explain the concept of back propagation through time.

OR

Design a convolutional neural network that takes input of 64 x 64 size and the hidden layers are as follows:

Layer 1: 10 filters of size 3 x 3 with padding = 1, Max Pooling 2 x 2

Layer 2: 32 filters of size 5 x 5 with padding = 0, Max Pooling 2 x 2

Layer 3: 5 filters of size 3 x 3 with padding = 0, Average Pooling 2 x 2

Dense Layer: 120 units                      Dense Layer: 5 units

10. What do you mean by Policy? Explain the objective and algorithm of Policy Iteration in RL.

\*\*\*

11. Describe the discretization mostly used in reinforcement learning with its advantages.

Tribhuvan University  
Institute of Science and Technology

2079



Master Level / I Year / IInd Semester/ Science

Data Science (MDS 552)

(Applied Machine Learning)

Candidates are required to give their answers in their own words as far as practicable.

Attempt All Questions

Group A

(5×3=15)

1. Differentiate between supervised and unsupervised learning.
2. Discuss the concept of mini-batch gradient descent algorithm.
3. What are various distance measures used for measuring similarity? Explain briefly.
4. What is Dunn Index? Explain.
5. What is meant by activation function? Discuss ReLU and Leaky ReLU activation functions.

Group B

(5×6=30)

6. Derive weight update rule for logistic regression.

OR

Fit the logistic regression model through the following data. Show one epoch of training.

Assume  $\alpha=0.1$ .

x <sub>1</sub>	0.8	0.4	0.5	0.6
x <sub>2</sub>	0.7	0.3	0.8	0.25
Y	0	1	0	1

7. Use Agglomerative clustering algorithm to divide the data points {(2,10), ((2,5), (8,4), (5,8), (7,5), (6,4))} into two clusters.
8. Derive weight update rule for BPTT?

OR

Consider a 2 bit grey level image with resolution 6x6 and a 3x 3 filter whose all diagonal elements are 1 and non-diagonal elements are zero. Compute feature map and then compute pooled feature map using 3x3 window. Use average pooling.

9. Discuss various measures used in multi-class classification with example.
10. Define the terms value and policy functions? Explain value and policy iteration algorithms.

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 First Assessment 2081

**Subject: Applied Machine Learning**

**Course No: MDS 552**

**Level: MDS /I Year /II Semester**

*Candidates are required to give their answer in their own words as far as practicable.*

**Attempt all questions.**

**Full Marks: 45**

**Pass Marks: 22.5**

**Time: 2 hrs**

**Group A [5 × 3=15]**

1. Describe the applications of reinforcement learning.
2. What is the concept of sequential learning. Describe about BPTT.
3. Given the following scores (logits) for three classes in classification:

$$scores = [2.0, 1.0, 0.1]$$

Calculate the SoftMax for these values.

4. Elaborate the concept of confusion matrix along with precision and recall.
5. Describe the working of K-Nearest Neighbor algorithm.

**Group B [5 × 6=30]**

6. Derive the weight update rule for linear regression using gradient descent.

**OR**

Given a dataset:

X	y
2	11.01
3	13.9
4	17.2

Apply weight update rule for linear regression for two iterations with learning rate 0.01.

7. Derive the Sigmoid function used in logistic regression.
8. Explaining the concept of convolution operations along with padding, stride and pooling illustrate how they are useful on extracting important features from images.

OR

Consider a Convolutional Neural Network having three different convolutional layers in its architecture as:

*Layer-1: Filter Size – 3 X 3, Number of Filters – 10, Stride – 1, Padding – 1*

*Layer-2: Filter Size – 5 X 5, Number of Filters – 20, Stride – 2, Padding – 0*

*Layer-3: Filter Size – 5 X 5, Number of Filters – 40, Stride – 2, Padding – 0*

*Layer-4: Flattening*

*Layer-5: 120 Dense Neuron*

*Layer-6: 3 Dense Neuron*

If we give the input a 3-D image to the network of dimension 32 X 32, then construct a convolutional neural network and compute the number of trainable parameters. (*Hint: Also, count bias term in each filter.*)

9. In a fraud detection system, transactions are classified into three categories: **Legitimate (A)**, **Suspicious (B)**, and **Fraudulent (C)**. The system produced the following confusion matrix based on 200 test transactions:

Actual \ Pred	A	B	C
A	90	10	5
B	12	50	8
C	3	7	15

Compute the macro and micro F1 Score and interpret the result.

10. Given the following data points:

- a) Positively labeled data: (3, 1), (3, -1), (6, 1), (6, -1)
- b) Negatively labeled data: (1, 0), (0, 1), (0, -1), (-1, 0)

How can we apply a linear kernel and derive the decision boundary that separates the positively labeled data from the negatively labeled data?

\*\*\*

Tribhuvan University  
Institute of Science and Technology  
2081  
☆

MDS / I Year / Second Semester/ Science  
**Data Science (MDS 552)**  
**(Applied Machine Learning)**

Full Marks: 45  
Pass Marks: 22.5  
Time: 2 hours.

*Candidates are required to give their answers in their own words as far as practicable.*

The figures in the margin indicate full marks. The symbols have their usual meanings.

**Attempt all questions.**

**Group A**

**[ $5 \times 3 = 15$ ]**

1. In linear regression, what does the coefficient of a feature represent? Elaborate with suitable example 12.5
2. What is the significance of precision and recall in classification algorithm.
3. What are the various attributes used in DBSCAN algorithms. Illustrate using suitable example.
4. How does a random forest model reduce the risk of overfitting compared to the decision tree?
5. What is the objective of back propagation in neural network? Elaborate with suitable example of your own.

**Group B**

**[ $5 \times 6 = 30$ ]**

6. Given points in 2D A(2, 3), B(3, 3), C(6, 5) and D(7, 8) where A and B are positively labelled and C and D are negatively labelled. Find the optimal hyperplane that separates these two classes using SVM Linear Kernel. 5.5

**OR**

When do you think ensemble learning is suitable in Machine Learning? Explain the various types of ensemble learning.

7. Consider a binary classification problem with two features, height and weight, to predict whether a person is Male or Female. The dataset is as follows:

Person	Height (cm)	Weight (kg)	Gender
1	170	65	Male
2	160	55	Female
3	175	70	Male
4	165	58	Female
5	180	75	Male
6	155	50	Female
7	185	80	Male
8	150	45	Female

- 5 Apply Naive Bases to predict the gender (Male or Female) of a person based on their height and weight. Given new observation, height = 167 cm and weight = 63 kg.
8. Apply DBSCAN algorithm to the given data points such that the maximum radius of the neighborhood is 1.5 and the minimum number of points required to form a dense region is 3.

Point	x	y
P1	1	2
P2	1.2	2.1
P3	1.4	2.2
P4	5	8
P5	6	9
P6	7	10
P7	8	8
P8	10	10

OR

Explain the working mechanism of PCA for the dimensionality reduction.

- u.s 9. What do you mean by dynamic programming? Write the Policy Iteration algorithm.
- u.s 10. A group of researchers has trained a model that classifies animals based on their features into Mammals, Birds and Reptiles. The results of the classification in dataset of 150 animals results to the following matrix:

Actual \ Predicted	Mammals (Class 0)	Birds (Class 1)	Reptiles (Class 2)
Mammals (Class 0)	400	50	30
Birds (Class 1)	40	35	80
Reptiles (Class 2)	20	70	45

- a) Calculate the overall accuracy of the model, precision and recall of each class.
- b) Compute the macro-average and Micro-average F1 Score
- c) Interpret the result.

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**

Second Reassessment 2082

**Subject: Machine Learning**

**Course No: MDS 552**

**Level: MDS /I Year /II Semester**

*Candidates are required to give their answer in their own words as far as practicable.*

**Full Marks: 45**

**Pass Marks: 22.5**

**Time: 2 hrs**

1. Discuss the difference of Machine Learning from Artificial Intelligence. **Group A [5 × 3 = 15]**
2. What is multi-class classification? How is it evaluated?
3. What are the evaluation metrics for binary classification?
4. What is the importance of exploration and exploitation in reinforcement learning?
5. Write short notes on following in the domain of reinforcement learning:  
 a) Clustering b) Low Rank Approximation

6. What is Naive assumption? Explain the working mechanism of Naive Bayes algorithm. **Group B [5 × 6 = 30]**

**OR**

A customer's decision to Buy a product (Yes / No) depends on three factors:

Customer	Income	Age	Frequency	Buy
1	Low	Young	Rare	No
2	Medium	Adult	Occasional	Yes
3	High	Adult	Frequent	Yes
4	Medium	Senior	Occasional	No
5	High	Senior	Frequent	Yes
6	Low	Adult	Rare	No
7	Medium	Adult	Frequent	Yes

Using the Naive Bayes classifier, predict whether the customer will Buy the product if Income = Medium, Age = Senior, Shopping Frequency = Frequent.

7. What do you mean by dimensionality reduction? Explain how PCA works?

**OR**

Cluster the given set of 2D points using Hierarchical Clustering algorithm.

Point	Coordinates (x, y)
P1	(1, 1)
P2	(1, 2)
P3	(2, 1)
P4	(2, 2)
P5	(8, 8)
P6	(8, 9)
P7	(25, 25)

8. Explain and illustrate the working mechanism of forward and backward propagation with suitable example of your own.

9. Given a confusion matrix compute Macro and Micro F1 Score:

Predicted	Actual		
	A	B	C
A	9	1	1
B	2	6	2
C	1	3	2

10. What is value function and Q function? Explain the value iteration algorithm. \*\*\*

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 First Assessment 2082

**Subject:** Applied Machine Learning

**Full Marks:** 45

**Code No:** MDS552

**Pass Marks:** 22.5

**Level:** MDS/ I Year/ II Semester

**Time:** 2hrs

Candidates are required to give their answers in their own words as far as practicable.

Attempt all questions.

**Group A [5 x 3 = 15]**

1. In linear regression, what does the coefficient of a feature represent? Elaborate with suitable Example.
2. What is the significance of precision and recall in classification?
3. What is the objective of back propagation in neural network? Elaborate with suitable example of your own.
4. Describe the common evaluation techniques used for regression techniques.
5. Brief the concept of reinforcement learning with suitable use cases.

**Group B [5 x 6 = 30]**

6. Show that the sigmoid function in logistic regression outputs a probability.

**OR**

Fit a linear regression model to the following data using suitable approach:

X1	Y
2	7.15
5	15.92
7	21.80

7. Given 4 points in 2D as A(2, 3), B(3, 3), C(6, 5) and D(7, 8) where A and B are positively labelled and C and D are negatively labelled. Find the optimal hyperplane that separates these two classes using SVM Linear Kernel.
8. A machine learning model is used to classify whether emails are spam or not spam. The following confusion matrix summarizes the model's performance on a test set:

	Predicted: Spam	Predicted: Not Spam
Actual: Spam	70	30
Actual: Not Spam	10	90

- a) Compute the **Precision**, **Recall**, and **F1 Score** for the model.
  - b) Interpret the meaning of the F1 Score in this context.
9. Explain the concept of a **Markov Decision Process (MDP)** in reinforcement learning, detailing its components, how it models decision-making problems, and the process of finding optimal policies. Additionally, describe the importance of the **Markov property** and how it relates to solving MDPs.
  10. Show the mathematical model of neuron. Explain how the forward propagation works in neural Network?

**OR**

Construct a convolution neural network which takes input image of size  $64 \times 64$  such that the layers of the network have following attributes:

**Layer1:** 16 filters of  $3 \times 3$  size with  $2 \times 2$  max pooling, stride = 1, padding = 1

**Layer2:** 64 filters of  $5 \times 5$  size with  $2 \times 2$  max pooling, stride = 1, padding = 0

**Layer3:** 12 filters of  $3 \times 3$  size, stride = 1, padding = 1

**Layer4:** 8 filters of  $3 \times 3$  size with  $2 \times 2$  average pooling, stride = 1, padding = 0

**Dense Layer:** 256 units

**Dense Layer:** 64 units

**Dense Layer:** 10 units

\*\*\*

Subject: Applied Machine Learning  
 Course No: MDS552  
 Year: MDS/ I Year/ II Semester

Full Marks: 45  
 Pass Marks: 22.5  
 Time: 2hrs

Candidates are required to give their answers in their own words as far as practicable.  
 Attempt all questions.

**Group A [5 x 3 = 15]**

- What do you mean by machine learning? What does machine actually learns?  
 Differentiate between bias and variance in machine learning. Why is the trade-off important?  
 What is the role of the activation function in a neural network? Provide an example showing how a nonlinear activation changes model behavior.  
 What do you mean by exploration and exploitation in RL?  
 Describe the concept of overfitting. What techniques can be used to prevent it?

**Group B [5 x 6 = 30]**

- Derive the weight update rule for logistic regression.

**OR**

- Using gradient descent, fit a linear model to the following dataset. Show at least two iterations of weight updates:

X	Y
1	3.2
4	10.5
9	15.1

A binary classifier produces the following confusion matrix for a fraud detection problem:

**Predicted Fraud Predicted Not Fraud**

Actual Fraud	120	40
Actual Not Fraud	20	320

Compute Accuracy, Precision, Recall, Specificity, and F1-Score.

Explain what a high recall signifies in this scenario.

Explain the architecture of a Feed Forward Neural Network. Discuss the flow of forward propagation and backpropagation. Provide a real-world example where Neural Network is suitable.

Design a CNN for input images of size  $128 \times 128$ , meeting the following requirements:

Layer 1: 32 filters,  $3 \times 3$  kernel, ReLU, stride 1, padding 1

Layer 2: 32 filters,  $3 \times 3$  kernel, max-pooling  $2 \times 2$

Layer 3: 64 filters,  $5 \times 5$  kernel, stride 1

Layer 4: 128 filters,  $3 \times 3$  kernel, average pooling  $2 \times 2$

Final layers:  $512 \rightarrow 128 \rightarrow$  output = 20 classes

Discuss the dimensionality after each layer and the total number of trainable parameters.

Write short notes on:      a) Deterministic and Stochastic Environment      b) Policy

**OR**

Explain the working mechanism of convolutional Neural Network along with the concept of padding, pooling and stride.

\*\*\*

Tribhuvan University  
 Institute of Sciences and Technology  
**SCHOOL OF MATHEMATICAL SCIENCES**  
 Second Assessment 2082

**Subject:** Applied Machine Learning  
**Course No:** MDS 552  
**Level:** MDS /I Year /II Semester

**Full Marks:** 45  
**Pass Marks:** 22.5  
**Time:** 2 hrs

*Candidates are required to give their answer in their own words as far as practicable.  
 Attempt all questions*

**Group A [5x3=15]**

- Differentiate between stochastic gradient descent and batch gradient descent.
- What does support vector mean? Why are they important?
- How is Agglomerative hierarchical clustering different from Divisive clustering?
- What do you mean by back propagation through time?
- Write short notes on following in the domain of reinforcement learning:  
 a) Stochastic Policy    b) Episode

**Group B [5x6=30]**

- What is Ensemble Learning? Explain its type with suitable illustration.

**OR**

A customer's decision to Buy a product (Yes / No) depends on three factors:

Customer	Income	Age	Frequency	Buy
1	Low	Young	Rare	No
2	Medium	Adult	Occasional	Yes
3	High	Adult	Frequent	Yes
4	Medium	Senior	Occasional	No
5	High	Senior	Frequent	Yes
6	Low	Adult	Rare	No
7	Medium	Adult	Frequent	Yes

Using the Naive Bayes classifier, predict whether the customer will Buy the product if Income = Medium,  
 Age = Senior, Shopping Frequency = Frequent.

- Explain K-Means algorithm using suitable example of your own.

**OR**

Cluster the given set of 2D points using DBSCAN algorithm where parameters epsilon = 1.5, Minimum Points = 3.

Point	Coordinates (x, y)
P1	(1, 1)
P2	(1, 2)
P3	(2, 1)
P4	(2, 2)
P5	(8, 8)
P6	(8, 9)
P7	(25, 25)

- Explain working mechanism of the convolutional neural network.
- Given a confusion matrix compute Macro and Micro F1 Score:

		Predicted		
Actual	A	B	C	
A	8	1	1	
B	2	6	2	
C	1	3	6	

- What is value function and Q function? Why does the Bellman equation rely on the Markov property, and what breaks if the environment is non-Markovian?

\*\*\*