GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 5ᵗʰ Semester
Subject Name: Machine Learning (ML)
Subject Code: 4659302

With effective from academic year 2018-19

1. Learning Objectives:
- Basic concepts of various learning methods
- To learning algorithms used in machine learning

2. Prerequisites:
Basics of computer science including algorithms, data structure, Basic Linear algebra and Probability theory.

3. Course Contents:

Unit	Course Content	Weightage Percentage
Unit I	**Introduction to Machine Learning, Model Preparation, Modelling and Evaluation** Human learning versus machine learning, types of machine learning, applications of machine learning, tools for machine learning, Machine Learning Activities, Data structures for machine learning, Data Pre-processing, selecting a model, training a model, model representation and interpretability, evaluating performance of a model, improving performance of a model	25%
Unit II	**Feature Engineering, Bayesian Concept Learning** Introduction to feature engineering, feature transformation, feature subset selection, Importance of Bayesian methods, Bayes' theorem, concept learning through Bayes' theorem, Bayesian Belief Network	20%
Unit III	**Supervised Learning – Classification, Regression** Example of supervised learning, classification model, classification learning steps, common classification algorithms, example of regression, common regression algorithms,.	20%
Unit IV	**Unsupervised Learning – Clustering, pattern finding using association rules** Unsupervised learning versus supervised learning, applications of unsupervised learning, clustering and its types, Apriori algorithm for association rule learning	17%
Unit V	**Neural Network** Understanding the biological neuron, exploring artificial neuron, types of activation functions, early implementation of artificial neural network, architectures of neural network, learning process in artificial neural network, backpropagation, Overview of Deep Learning	18%

Desirable Topics:
Representation Learning, Active Learning, Instance-based learning, Ensemble Learning

4. Text Book:

1) Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, "Machine Learning", Pearson Education

5. Reference Books:

1) Tom M Mitchell, "Machine Learning", McGraw Hill
2) Anuradha Srinivarasaraghavan, Vincy Joseph, "Machine Learning", Wiley India

GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 5th Semester
Subject Name: Machine Learning (ML)
Subject Code: 4659302

With effective from academic year 2018-19

3) Peter Harrington, "Machine Learning in Action", DreamTech
4) Henrik Brink, Joseph Richards, Mark Fetherolf, "Real-World Machine Learning", DreamTech
5) Christopher Bishop, "Pattern Recognition and Machine Learning"
6) Jiawei Han and Michelline Kamber, "Data Mining: Tools and Techniques", 3rd Edition.

6. Chapter wise Coverage from the Text Book:

Unit #	Chapter
I	Chapter 1: 1.1 to 1.8, Chapter 2: 2.1 to 2.6 and Chapter 3: 3.1 to 3.6
II	Chapter 4: 4.1 to 4.3, Chapter 6: 6.1 to 6.5
III	Chapter 7: 7.1 to 7.5 and Chapter 8: 8.1 to 8.3
IV	Chapter 9: 9.1 to 9.5
V	Chapter 10: 10.1 to 10.9

7. Accomplishment of the student after completing the course:
1) Student will be able to understand the concept of Machine learning and range of problems that could be solved by machine learning. They will be able to compare different types of learning algorithms and apply machine learning concepts in real life problems.

Practical List

1. Write a python code to implement **decision tree** for below given dataset. Identify the root node and all subpart or children of node and draw the tree.

Item no	Age	Income	Student	Credit Rating	Buys-Computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Middle	Low	Yes	Excellent	No
7	Senior	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle	Medium	No	Excellent	Yes
13	Middle	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

2. Write a python code to implement **K-nearest neighbourhood** program for the given dataset. Assume that value of K=19.

Syllabus for Master of Computer Applications, 5th Semester
Subject Name: Machine Learning (ML)
Subject Code: 4659302

Item no	Outlook	Temp	Humidity	Windy	Play
1	Sunny	Hot	High	FALSE	No
2	Sunny	Hot	High	TRUE	No
3	Overcast	Hot	High	FALSE	Yes
4	Rainy	Mild	High	FALSE	Yes
5	Rainy	Cool	Normal	FALSE	Yes
6	Rainy	Cool	Normal	TRUE	No
7	Overcast	Cool	Normal	TRUE	Yes
8	Sunny	Mild	High	FALSE	No
9	Sunny	Cool	Normal	FALSE	Yes
10	Rainy	Mild	Normal	FALSE	Yes
11	Sunny	Mild	Normal	TRUE	Yes
12	Overcast	Mild	High	TRUE	Yes
13	Overcast	Hot	Normal	FALSE	Yes
14	Rainy	Mild	High	TRUE	No

3. Write a python code to implement Apriori **a**lgorithm, apply join and prune method and find frequent itemset

Sr#	Item no	Name
1	T1	Bread, butter, milk, soda
2	T2	Coke, egg, milk
3	T3	Bread, butter, egg
4	T4	Break, coke, jam
5	T5	Bread, butter
6	T6	Potato chips, soda
7	T7	Coke, fruit, juice
8	T8	Bread, coke, milk
9	T9	Coke, soda, jam, milk
10	T10	Bread, butter, egg, milk, soda
11	T11	Bread, milk
12	T12	Bread, jam

4. Write a python code to apply **Naive Bayesian** algorithm to clasify that whether a person can buy computer or not based on given test data :

Item no	Age	Income	Student	Credit Rating	Buys-Computer
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Middle	Low	Yes	Excellent	No
7	Senior	Low	Yes	Excellent	Yes

GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 5th Semester
Subject Name: Machine Learning (ML)
Subject Code: 4659302

With effective from academic year 2018-19

8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle	Medium	No	Excellent	Yes
13	Middle	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

Test Data
Age : Youth   Income : LOW Student : No Credit Rating : Fair   Buy Computer - ??

5. Write a python code to calculate gini index and draw the regression tree using **CART** algorithm for the below data set of ASCII characters:

511 975 511 975 496 966 480 967 450 981 420 993 376 1023 337 1053 297 1094 269 1135 248 1177 241 1215 252 1245 280 1264 323 1270 373 1263 431 1238 492 1205 550 1168 604 1126 648 1087 684 1056 710 1031 725 1013 725 1013 728 1025 719 1053 707 1095 690 1150 673 1214 654 1288 636 1368 616 1448 595 1523 574 1595 550 1656 529 1704 503 1731 484 1743 469 1734 463 1705 471 1661 493 1595 528 1524 581 1442 645 1364 709 1294 776 1231 838 1186 747 921 755 903 757 882 744 869 717 866 686 866 637 891 586 915 525 960 471 997 418 1045 379 1090 351 1132 340 1168 347 1195 369 1209 409 1211 459 1202 514 1185 572 1162 632 1127 688 1092 735 1053 777 1014 810 987 835 965 851 957 861 957 862 974 857 999 847 1043 833 1097 816 1159 797 1225 776 1300 753 1377 731 1452 708 1521 685 1581 660 1634 636 1674 611 1704 589 1716 568 1714 552 1704 539 1686 538 1657 538 1622 558 1575 583 1531 622 1485 672 1437 728 1387

6. Given the below training dataset of petal size and flower type, predict flower type for petal of size 2.5 cm. using K-Nearest-Neighbor classification.

Petal Size	Flower Type
1	a
2	b
1	a
2	b
3	c
4	d
3	c
2	b
5	a

7. Consider the below training data and determine if jacket is to be worn when the temperature is 12-degree Celsius using Linear Regression Model.

Outside Temperature	**Wear a Jacket**
30°C	No
25°C	No
20°C	No

15°C	Yes
10°C	Yes

8. Implement a python program that takes interest rate (x), finds the equation that best fits the data and is able to forecast out median home price for given interest rate using the data given below. (Use linear regression)

interest rate (%) (x)	Median home price (y)
10.3	$183,800
10.3	$183,200
10.1	$174,900
9.3	$173,500
8.4	$172,900
7.3	$173,200
8.4	$173,200
7.9	$169,700
7.6	$174,500
7.6	$177,900
6.9	$188,100
7.4	$203,200
8.1	$230,200
7	$258,200
6.5	$309,800
5.8	$329,800

9. Write a python code to predict profit of hotel chain given the population of the area (city) using the data at
https://docs.google.com/spreadsheets/d/1Ks20skBgEefHFU36sFqVzozoFtz2EZE2rxB_IgXOrUg/edit?usp=sharing.

10. Write a python code to predict the price of house given square feet and number of bed rooms in the house for the dataset available at
https://docs.google.com/spreadsheets/d/1DHVK7gKo4TSyj7mFLwofHamj1Sl4SOZma2q51w1ZvyE/edit?usp=sharing

11. Build a classification model in python that classifies if a student gets admission in a course or not given his last two examination scores for the dataset available at
https://docs.google.com/spreadsheets/d/1g0mjTUZ9Ado5prXA1UnAvNjmdzTrV0TzkFkIoU-Lpbk/edit?usp=sharing

12. Build a multivariate logistic regression model to classify glass type of glass given different glass mixture features using the Glass Identification Dataset from UCI Machine Learning Repository.

13. Implement supervised machine learning algorithm (Classification – K Nearest Neighbourhood) in python to classify breast tumour data into malignant breast tumour or benign breast tumour (use breast tumour dataset) and obtain its accuracy level.

14. Implement supervised machine learning algorithm (Classification – K Nearest Neighbourhood) in python to classify iris data into setosa, virginica, versicolor using iris dataset and obtain its accuracy level.

GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for Master of Computer Applications, 5th Semester
Subject Name: Machine Learning (ML)
Subject Code: 4659302

With effective from academic year 2018-19

15. Implement supervised machine learning algorithm (Classification – Support Vector Machine) in python to classify breast tumour data into malignant breast tumour or benign breast tumour (use breast tumour dataset) and obtain its accuracy level.
16. Write a python program to build an email spam classifier using support vector machines for the Spam base dataset from UCI machine learning repository.
17. Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Titanic dataset to cluster data (use Titanic dataset) by removing the class label.
18. Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Breast Tumour dataset to cluster data (use Breast Tumour dataset) by removing the class label.
19. Implement unsupervised machine learning algorithm (Clustering – Hierarchical) in python on Titanic dataset to cluster data (use Titanic dataset).
20. Implement Apriori algorithm in python to find rules which explain association between different products for given transactions at a retail store. (The data is available at https://drive.google.com/file/d/1NUXoptUlHY8z4KcFKpFA6sQN5KnWzk3p/view?usp=sharing)
21. Implement text classification using neural network in python/R on Twenty Newsgroup dataset from UCI machine learning repository.
22. Implement supervised machine learning algorithm (Classification - Naïve Bayes algorithm) in python/R on Pima Indians Diabetes dataset and obtain its accuracy level.
23. classification and prediction algorithms on UCI dataset using Python's scikit-learn library

Desirable Practical Lists

1. For the sentiment analysis dataset given in link https://drive.google.com/file/d/1x6H7_KJjkbDrpgZFS7I2wjsZsILeSJ4S/view?usp=sharing, implement the following in python,
   a. Clean and pre-process the dataset by removing URL, removing HTML tags, handling negation words which are split into two parts, converting the words to lower cases, removing all non-letter characters
   b. Split the dataset into training and testing set
   c. Implement feature extraction technique (to convert textual data to the numeric form)
   d. Build the classification model using Logistic Regression that classifies if a given sentiment text is positive or negative
   e. Obtain the accuracy score of the built model.
2. Implement a content based recommender system in python that recommends movies that are similar to a particular movie using movielens-20m-dataset available at https://kaggle.com.

The practical exercises should be performed in python.

References:

1) Peter Harrington, "Machine Learning in Action", DreamTech
2) Michael Bowles, "Machine Learning in Python", Wiley
3) Gavin Hackeling, Mastering Machine Learning with scikit-learn, Packt
4) Giuseppe Bonaccorso, Machine Learning Algorithms - Second Edition, Packt