Course code | Course Name | L-T-P | Credits | Year of Introduction
--- | --- | --- | --- | ---
CS467 | MACHINE LEARNING | 3-0-0-3 | 2016 |

Course Objectives:
- To introduce the prominent methods for machine learning
- To study the basics of supervised and unsupervised learning
- To study the basics of connectionist and other architectures

Syllabus:
Introduction to Machine Learning, Learning in Artificial Neural Networks, Decision trees, HMM, SVM, and other Supervised and Unsupervised learning methods.

Expected Outcome:
The Students will be able to:

i. differentiate various learning approaches, and to interpret the concepts of supervised learning

ii. compare the different dimensionality reduction techniques

iii. apply theoretical foundations of decision trees to identify best split and Bayesian classifier to label data points

iv. illustrate the working of classifier models like SVM, Neural Networks and identify classifier model for typical machine learning applications

v. identify the state sequence and evaluate a sequence emission probability from a given HMM

vi. illustrate and apply clustering algorithms and identify its applicability in real life problems

References:

Course Plan

<table>
<thead>
<tr>
<th>Module</th>
<th>Contents</th>
<th>Hours</th>
<th>End Sem. Exam Marks %</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Introduction to Machine Learning, Examples of Machine Learning applications - Learning associations, Classification, Regression, Unsupervised Learning, Reinforcement Learning, Supervised learning- Input representation, Hypothesis class, Version space, Vapnik-Chervonenkis (VC) Dimension</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>II</td>
<td>Probably Approximately Learning (PAC), Noise, Learning Multiple classes, Model Selection and Generalization, Dimensionality reduction- Subset selection, Principle Component Analysis</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>III</td>
<td>Classification- Cross validation and re-sampling methods- K-fold cross validation, Boot strapping, Measuring classifier performance- Precision, recall, ROC curves. Bayes Theorem, Bayesian classifier, Maximum Likelihood estimation, Density functions, Regression</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>IV</td>
<td>Decision Trees- Entropy, Information Gain, Tree construction, ID3, Issues in Decision Tree learning- Avoiding Over-fitting, Reduced Error Pruning, The problem of Missing Attributes, Gain Ratio, Classification by Regression (CART), Neural Networks- The Perceptron, Activation Functions, Training Feed Forward Network by Back Propagation.</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>VI</td>
<td>Unsupervised Learning - Clustering Methods - K-means, Expectation-Maximization Algorithm, Hierarchical Clustering Methods , Density based clustering</td>
<td>6</td>
<td>15</td>
</tr>
</tbody>
</table>

**END SEMESTER EXAM**

**Question Paper Pattern**

1. There will be *FOUR* parts in the question paper – A, B, C, D
2. **Part A**
   a. *Total marks : 40*
   b. *TEN* questions, each have *4 marks*, covering all the SIX modules (*THREE* questions from modules I & II; *THREE* questions from modules III & IV; *FOUR* questions from modules V & VI).
   *All the TEN* questions have to be answered.
3. **Part B**
   a. *Total marks : 18*
   b. *THREE* questions, each having *9 marks*. One question is from module I; one question is from module II; one question uniformly covers modules I & II.
   c. *Any TWO* questions have to be answered.
   d. Each question can have *maximum THREE* subparts.
4. Part C  
a. Total marks : 18  
b. THREE questions, each having 9 marks. One question is from module III; one question is from module IV; one question uniformly covers modules III & IV.  
c. Any TWO questions have to be answered.  
d. Each question can have maximum THREE subparts.

5. Part D  
a. Total marks : 24  
b. THREE questions, each having 12 marks. One question is from module V; one question is from module VI; one question uniformly covers modules V & VI.  
c. Any TWO questions have to be answered.  
d. Each question can have maximum THREE subparts.

6. There will be AT LEAST 60% analytical/numerical questions in all possible combinations of question choices.