INT354:MACHINE LEARNING-I

L:2 T:0 P:2 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: Explore Different types of Machine Learning and statistics used for risk minimization.

CO2 :: Analyze the operations of different types of Machine Learning Classifiers.

CO3 :: Examine the performance of Generative models based on Bayesian learning to solve different classification problems.

CO4 :: Develop the model that predict value of continuous variable with regression analysis.

CO5 :: Discuss the methods for Error calculations using different Regression metrics.

CO6 :: Extend the Machine Learning approach to understand the bias complexity tradeoff and algorithm independent machine learning.

Unit I

Introduction to machine learning : Machine Learning, Need of Machine Learning, Types of Learning, Well Posed Learning Problems, Designing a Learning Systems, Statistical Learning Framework, Empirical Risk Minimization, Empirical Risk Minimization with Inductive Bias, PAC Learning

Building good training sets : Data Preprocessing, Dealing with Missing Data, Handling Categorical Data, Partitioning a Dataset in Training and Test Sets, Normalization, Selecting Meaningful Features

Unit II

Machine learning classifiers : Choosing a Classification Algorithm, First Steps with Scikit-Learn, Perceptron Classifier, Stochastic Gradient Descent, Modeling Class Probabilities via Logistic Regression, Maximum Margin Classification with Support Vector Machine, Decision Tree Learning, CART, ID3, C4.5, Density Estimation, Parzen Window, The Nearest Neighbour Rule, KNearest Neighbour Estimation

Unit III

Generative models : Maximum Likelihood Estimator, Bayesian Learning, Bayes Theorem, Brute-Force Concept Learning, Bayes Optimal Classifier, Gibs Algorithm, Naive Bayes Classifier, EM Algorithm

Model evaluation and hyperparameter tuning : Streamlining Workflows with Pipelines, Using k-fold Cross Validation to Access Model Performance, Debugging Algorithms with Learning and Validation Curves, Fine-Tuning Machine Learning Models via Grid Search

Unit IV

Predicting continuous target variables with regression analysis : Introducing Linear Regression, Fitting a Robust Regression Model using RANSAC, Relationship Using a Correlation Matrix, Exploratory Data Analysis, Regularized Methods for Regression, Polynomial Regression, Decision Tree, ARIMA

Unit V

Regression Metrics : R2 Score, Mean Absolute Error, Mean Squared Error, Mean Squared Logarithmic Error, Mean Absolute Percentage Error, Explained Variance Score, D2 Score Visual Evaluation of Regression Models

Unit VI

The bias-complexity tradeoff : No Free Lunch Theorem, Error Decomposition, The VC-Dimension, The Rademacher Complexity, The Natarajan Dimension

Algorithm-Independent machine Learning : Combining Classifiers, Majority Voting Classifier, Re-sampling for Estimating Statistics, Lack of Inherent Superiority of Classifier, Bagging and Boosting Classifier, Random Forest Classifier, Regressor, Support Vector Classifier and Regressor

List of Practicals / Experiments:

List of Practical

- Write a Program to perform missing data handling.
- Write a Program to perform categorical data handling.
- Write a Program to select features from data after normalization.
- Write a Program to perform binary classification using single layer perceptron.

- Write a Program to perform classification using support vector machine.
- Write a Program to perform clustering using K-Nearest Neighbor algorithm.
- Write a Program to evaluate model using K-fold cross validation.
- Write a Program to perform hyper-parameter tuning using grid search.
- Write a Program to perform prediction using Linear regression.
- Write a Program to perform data pre-processing using correlation matrix.
- Write a Program to perform prediction using decision tree.
- Write a Program to evaluate the performance of any ml algorithm using R2 score, mean absolute error metrics.
- Write a Program to evaluate the performance of any ml algorithm using mean squared error, mean squared logarithmic error
- Write a program to perform classifier combination and perform classification based on majority voting classifier.
- Write a program to compare different classifier performance like bagging boosting, random forest and support vector classifier.

Text Books:

1. MACHINE_LEARNING_IN_ACTION by PETER HARRINGTON, Manning Publications

References:

1. UNDERSTANDING-MACHINE-LEARNING-THEORY-ALGORITHMS FROM THEORY TO ALGORITHM by SHAI SHALEV-SHAWARTZ AND SHAI BEN-DAVID, CAMBRIDGE UNIVERSITY PRESS
2. MACHINE LEARNING by TOM M. MITCHELL, Mc Graw Hill Education