



**GC University, Lahore**

**Department** **of** Computer Science

**CS-7210** **Machine Learning** **2022-2023**

Credit Hours : 3+0

|  |  |  |  |
| --- | --- | --- | --- |
| Instructor: Office: | **Dr.** **Muhammad Safyan** Room No. 114 | Email: Office Hours:  Extension: | safyanch@gcu.edu.pk Mon-Fri 8am – 4pm  270 |

**Course Description**

This course provides a broad introduction to machine learning and statistical pattern recognition. Topics include: supervised learning (generative/discriminative learning, parametric/non-parametric learning, neural networks, and support vector machines); unsupervised learning (clustering, dimensionality reduction, kernel methods); learning theory (bias/variance tradeoffs, practical advice); reinforcement learning and adaptive control. The course will also discuss recent applications of machine learning, such as to robotic control, data mining, autonomous navigation, bioinformatics, speech recognition, and text and web data processing.

Course is dedicated to Research and application

**Preliminary Syllabus**

* Introduction to machine Learning, Demos, Notation, Tools
* Introduction t o Linear Algebra: matrices and vectors, Additon and Scalar Multiplication, Matrix vector multiplication, matrix multiplication properties, inverse and transpose of matrix,
* Python, Jupiter-Anaconda
* supervise vs unsupervised learning, Linear Regression,Model representation, cost function and its intuition, Gardient Descent and its intuition,
* Regression with multiple feature, its gradient desecent, polynomial Regression, Normal equation and its non invertibility.
* Logistic Regression, its hypothesis, detecting decision boundary, cost function and gradient descent, advance optimization techniques
* Regularization: over fitting, cost function, regularized linear and logistic regression
* Neural Network(NN): Non Linear Hypothesis, how the neuron of the brain works, Model representation of NN, cost function , back propagation, Gradient check in NN.
* Learning curve, Error Analysis, Error Metrics and skew classes, Trade off between Precision and Recall.
* Data for Machine Learning
* Attribute types: Nominal, Binary, Ordinal, Numeric Attribute, Interval, Ratio, Discrete vs. Continuous
* Basic Statistical Descriptions of Data: Mean, Median, Mod, Quartiles, Box plot, Range, Major task of Data Preprocessing: Data Cleaning, Incomplete, Noisy, Inconsistence, Data Integration, Correlation Analysis, Chi-Square Pearson’s product moment coefficient
* Data Reduction Data Transformation and discretization, Aggregation, Normalization, Discretization.
* Support Vector Machine(SVM): Optimization Objective, Large Margin Intuition and its Mathematics, Kernel.
* Clustering: Unsupervised Learning , K mean algorithm, optimization objective, Random initialization, Choosing the number of clusters.
* Dimensionality Reduction, Principal Component Analysis (PCA).
* Anomaly Detection Algorithm, Hieratical Clustering, , Random forest ,
* Recommender System: Collaborative filtering Algorithm, Detail mean normalization, vectorization,
* Large Scale Machine Learning: Map Reduce and Data Parallelism, Mini batch Gradient Descent, online learning, stochastic GD and its convergence.
* Introduction to Deep Learning

**Recommended Text Books**

1. No Book is recommend, You may have to go through many books topic by topic.

**Grading / Mark Scheme**

 10/15-Minute Quizzes:

 Assignments:

 Presentations/Research paper

 Mid-Term Exam:

 Final Exam:

**Policy Matters**

 Assignments will be issued which will be due one week from the issue date.

 Quizzes may be conducted in class during the first 5-10 minutes, and late-comers will suffer.

 Missed quizzes cannot be retaken under any circumstances.

 Anyone found assisting or committing plagiarism in any assignment or quiz will have all their assignment and quiz marks cancelled.

 At least 80% attendance needs to be maintained in order to be allowed to sit the Final Exam.