



**GC University, Lahore**

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

**CS-7210** **Machine Learning** Spring-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 |

Lecture Material: <https://github.com/safyanch/Machine-Learning-2023>

**Course Description**

This course provides a broad and Advanced level knowledge about Machine Learning Concepts and underlying Mathematical intuition behind the machine learning models. Purpose of this course is to explore the latest research, work on complex projects, and gain hands-on experience with state-of-the-art tools and frameworks. Sharpen your ability to tackle challenging problems and push the boundaries of what's possible in the world of artificial intelligence.

The course is focused on both research and practical application

**Preliminary Syllabus**

* Introduction to machine Learning, Potential applications of ML/AI ,Demos, Notation, Tools
* Supervise vs Unsupervised learning, Linear Regression, Model representation, cost function and its intuition, Gradient Descent and its intuition,
* Regression with multiple feature, its gradient descent, polynomial Regression, Normal equation and its non invertibility.
* Introduction to Python and its popular libraries like Numpy, Pandas , seaborn, Matplotlib, Tensor flow, ML-editor Jupyiter-Anaconda-Notebook, VSCode Editor.
* Classification: Logistic Regression, its hypothesis, detecting decision boundary, cost function and gradient descent, advance optimization techniques.
* Evaluation Methods: to evaluate Classification and regression Models, Precision, Recall, Accuracy, F1-Score, ROC Curve, R-Square, Adjusted R-Square. Type-1 Error, Type-2 Error.
* Optimization of Machine Learning Models: Regularization: over fitting, underfitting, Bias Variance, cost function, regularized linear and logistic regression, Mini-batch Gradient Descent.
* Decision Tree,
* Introduction to Probability, Gaussian probability, Statistical interpretation of Linear Regression and logistic Regression
* Naive Bayes Classifier
* Ensemble model: Random forest, Adaboost, Gradient Boost, xboost.
* Data Analytics: 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,
* Data Reduction Data Transformation and discretization, Aggregation, Normalization, Discretization.
* 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.
* 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).
* Recommender System:

**Recommended Text Books**

1. Introduction to Machine Learning with Python: A Guide for Data Scientists by Andreas C. Müller & Sarah Guido
2. Pattern Recognition and Machine Learning by Christopher M. Bishop

**Grading / Mark Scheme**

 3 Quizzes:

 3-Assignments/ Research Paper / Presenations

 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 10-15 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.