AMERICAN UNIVERSITY OF CENTRAL ASIA
DEPARTMENT OF ECONOMICS

MACHINE LEARNING

Spring 2023

Instructor
Prof. Nazgul Jenish
jenish_na@auca.kg

Lecture:
Friday 18:30 – 19:45
Seminar: Fri 19:55 –21:10

Office Hours:
Friday: 17:15 – 18:30
Sat: 12:00 -12:45

Course Description:
This course introduces students to the fundamental statistical machine learning methods -- both theory and practice. Over the last decade, the machine learning discipline has experienced a tremendous growth, spawning a wide range of applications in virtually all walks of life, including computer vision (image recognition), voice recognition, text processing, robotics, healthcare, economics, finance, and marketing, to name a few. All statistical methods reviewed in the course will be illustrated with practical examples implemented in Python and students will be taught relevant Python modules and tools. By the end of the course, students will produce an original paper applying the learned methods to a practical machine learning problem.

The first part of the course discusses the supervised learning problems of regression and classification. Specifically, various methods and algorithms for regression and classification are covered, including linear, polynomial and logistic regressions, discriminant analysis, naïve bayes, tree models, random forests and gradient boosting trees, support vector machines and neural networks.

The second part of the course reviews unsupervised learning methods, including principal components and clustering algorithms. The following clustering techniques are introduced and demonstrated by market segmentation applications: K-means clustering, agglomerative clustering, hierarchal clustering.

Finally, the course discusses the basic methods for model selection (optimal model complexity and feature selection), preventing overfitting, regularization, and model evaluation. These methods encompass, inter alia, cross-validation, bootstrap, probabilistic criteria and scoring techniques, such as Akaike information criterion, Bayes Information criterion, and Cp. Model-specific evaluation metrics are introduced and analyzed for all models.

Prerequisites:
Students must have a minimum baseline of programming knowledge and skills in Python, as well as a strong background in Probability and Statistics, Linear Algebra and Multivariate Calculus or Analysis (e.g., optimization methods). The mathematical prerequisite may be satisfied by completion of the AUCA Economics or Mathematics department courses, or if the course was taken outside AUCA, the syllabi must be shown to the Instructor. As for Python,

students are expected to have knowledge of the following Python modules: Numpy, Pandas, SciPy, Matplotlib and Seaborn.

Required Text:
An Introduction to Statistical Learning. Trevor Hastie, Robert Tibshirani, Gareth James, Daniela Witten. Springer, 2nd ed. 2021 Edition

Recommended Text:
The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Second Edition (Springer Series in Statistics) 2nd Edition, 2016

Required Software:
Jupyter Notebook (freely downloadable from Anaconda's website), comes with the latest Python interpreter (Python 3.9 or later version is required)

Requirements:
You are expected to attend all lectures and seminars, as well as to complete eight-to-ten homework assignments during the semester. Assignments will be posted on the e-Course webpage and will be due at the beginning of lectures. No late assignments will be accepted. Completion of the assignments is strongly recommended as they are an integral part of the learning process. They will also have a direct impact on your final grade.

There will be one mid-term in-class exam and one final research paper. No make-up exams will be given. The detailed requirements for final papers will be distributed separately.

Course grades will be determined as follows:

Final Paper	40%
Mid-term exam	40%
Homework assignments	20%
Total	100%

Academic Policies:
The standard AUCA rules and regulations concerning absences, exams, grading, plagiarism, and violations of academic integrity will apply to this course. Please consult the AUCA website.

Course Outline (I may add, drop or change the order of topics):
Objectives and Types of Statistical Learning Problems
Regression: Linear and Polynomial Regressions, k-NN, Ridge, Lasso regularization, PCR regression, PLS regression
Classification: Logistic Regression, Naïve Bayes, Discriminant Analysis, k-NN
Resampling Methods: cross-validation, bootstrap
Model Selection: AIC, BIC, Cp, validation set, cross-validation
Tree-Based Models: decision and regression trees, gradient boosting, random forests
Support Vector Machines
Deep Neural Nets
Unsupervised Learning: PCA, clustering methods