Subject Code and Course Number: *CS251*
Course Title: *Machine Learning*
Number of Credits: *3*
Term/Year: *Spring2018-2019*
Class Schedule: Tuesday-Thursday *9:00-10:20am, Room 413W PAB*
Prerequisites: *IESM 106 (Probability and Statistics), CS 108 (Statistics)*
Co-Requisites: *None*

Instructor Name: *Ashot Harutyunyan*
Email Address: *ashot.harutyunyan@aua.am*
Office Location: *335W PAB*
Office Hours: *Tuesday-Thursday12:00-13:00*

TA Name: *Avetik Arakelyan*
Email Address: *arakelyanavetik@gmail.com*
Office Location: *310W*
Office Hours: Thursday-Friday 17:30-18:30

Course Description
Machine learning links together computers and statistics by teaching machines to act without human interaction. It compiles those methods of data science that automate model building process for computer realization by applying algorithms that iteratively learn from data allowing computers to find hidden insights in data without explicit programming. This course will provide the basic ideas and methods of machine learning. Topics include - supervised learning, unsupervised learning, best practices in machine learning with many examples from real-world applications. It also includes explanations on how to use the well-known R software for application of the learned techniques to practical problems. Three hours of instructor-led class time per week including discussions and problem sets.

Required Materials

Machine Learning:

[1] *G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, with Applications in R by, Springer, 2013.*
[2] *T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data mining, Inference, and Prediction, Springer, 2008.*
[3] *B. Lantz,* Machine Learning with R, *Packt Publishing, Second Edition, 2015.*
[4] *M. Kuhn, and K. Johnson, Applied Predictive Modeling, Springer, 2013.*
[5] *F. Chollet, Deep Learning with Python, Manning Publications, 2018.*
[5*] *F. Chollet, Deep Learning with R, Manning Publications, 2017.*

R programming:

[6] *R. I. Kabacoff, R in Action: Data Analysis and Graphics with R, Manning Publications, Second Edition, 2015.*
[7] *M. J. Crawley, The R Book, A John Wiley & Sons, Ltd., Publication, 2013*
[8] *N. Zumel, and J. Mount, Practical Data Science with R, Manning Publishing, 2014*
[9] *R. Cotton, Learning R, O'Reilly, 2013*
[10] *H. Wickham, GGPlot2, Elegant Graphics for Data Analysis, Springer, Second Edition, 2016*

Schedule & Topics *(subject to possible changes)*

Week	Days	Topic	Reading
1	Jan17	**Introduction** - What is Data Science, AI, Machine Learning and Deep Learning? - Historical overview	**[1]** Ch 1,2 **[5]** Ch 1
2	Jan 22, 24	**Main Concepts** - Supervised vs unsupervised learning - Regression vs classification - Parametric vs non-parametric models - No-free-lunch theorem - Ockham's razor principle	**[1]** Ch 1,2 **[5]** Ch 1
3	Jan 29, 31	**Regression: Model Accuracy** - Error decomposition into reducible and irreducible - Bias-variance decomposition	**[1]** Ch 1,2
4	Feb 5, 7	**Linear Regression** - Simple linear regression - Multiple linear regression - p-value, F-statistic, R squared - Potential problems in regression: non-linearity, correlation of error terms, non-constant variance of the error terms, outliers, high-leverage points, collinearity - Interaction terms - Model selection: AIC, BIC, C_p, adjusted R-squared	**[1]** Ch 3 **[4]** Ch 3 **[3]** Ch 6
5	Feb 12, 14	**Resampling Methods** - The validation set approach - Leave-one-out cross validation - K-fold cross validation - Bootstrap	**[1]** Ch 5 **[4]** Ch 4
6	Feb 19, 21	**Regularization** - Ridge regression - The Lasso - Elastic net	**[1]** Ch 6 **[4]** Ch 6
7	Feb 26, 28	**Mid-term**	
8	**Mar 4–9**	**SPRING BREAK!**	
9	Mar 12, 14	**Classification: Model Accuracy** - Accuracy and error rate of the classifier	**[4]** Ch 11 **[1]** Ch 4

		- Classification measures – TPR, FPR, Precision, Sensitivity, ROC curve, AUC - Baseline models **K-Nearest Neighbors** - Regression - Classification	**[4]** Ch 13 **[3]** Ch 3
10	Mar 19, 21	**Classifiers** - Bayes classifier - Naïve Bayes classifier - Logistic regression	**[1]** Ch 4
11	Mar 26, 28	**Classifiers** - Linear discriminant analysis - Quadratic discriminant analysis	**[1]** Ch 4 **[4]** Ch 12 **[4]** Ch 13
12	Apr 2, 4	**Decision Trees** - Regression - Classification	**[1]** Ch 8 **[4]** Ch 14
13	Apr 9, 11	**Tree based methods** - Bagging - Random forests	**[1]** Ch 8 **[4]** Ch 14
14	Apr 16, 18	**Tree based methods** - Boosting - C5.0 Algorithm	**[1]** Ch 8 **[4]** Ch 14 **[3]** Ch 5
15	Apr 23, 25	**Unsupervised Learning: Dimensionality Reduction** - Principal components analysis	**[1]** Ch 10
16	Apr 30, May 2	**Unsupervised Learning: Clustering** - K-means clustering - Hierarchical clustering - Dendrograms	**[1]** Ch 10
17	May 7, 9 (Victory Day)	**Support Vector Machines** - Maximal margin classifier - Soft margin classifier - Support vector machines	**[1]** Ch 9
18	17 May – 23 May	**FINAL EXAM**	
	May27	**GRADES**	

Student Learning Outcomes

The following chart shows alignment between course-specific and program student learning outcomes and program goals.

Course Learning Outcomes *In this course, students will be able to:*	***Course-based Student Learning Outcomes*** *Students will be able to:*	***Program Goal***

• *Identify the set of right algorithms appropriate for a specific problem. Describe pros and cons of different algorithms applied for the same problem* • *Identify the basic approaches while solving real-world problems* • *Apply classification, regression and clustering algorithms for data understanding and predicting.* • *Estimate model accuracy, adjust model parameters. Interpret the results* • *Use R software for academic research and for solving practical problems. Write comprehensive problem solving program codes*	• *Analyze complex data structures and algorithms* • *Develop and implement software applications in one or more programming languages* • *Develop and test software tools and methods* • *Use common software and information technology to pursue inquiry relevant to their academic and professional fields, and personal interests*	• *Prepare students for development of scientific, engineering and industrial software applications* • *Provide students with a broad foundation of knowledge and skills and cultivate a commitment to life-long learning*

Course Structure

Instructor-led class will meet three times per week. Home tasks include readings, and problem sets. All assignments will undergo in-class discussion. Reading will be assigned by the instructor based on the covered material. All home tasks must be completed before the discussions start.

Method of Evaluation

Midterm (**M**)	40 %	97-100 = A +93-96 = A90-92 = A-
Final Exam (**F**)	40 %	87-89 = B +83-86 = B80-82 = B-
Homework (**HW**)	20 %	77-79 = C +73-76 = C70-72 = C–
Total	100%	67-69 = D +63-66 = D60-62 = D -
		0-59 = F

So the Grade Calculation Formula is:

$$\boldsymbol{Total} = \mathbf{0.2} * \boldsymbol{HW} + \mathbf{0.4} * \boldsymbol{M} + \mathbf{0.4} * \boldsymbol{F}$$

The final numerical score for the course obtained may or may not be scaled before it is converted to a letter grade. The instructor reserves the right to decide whether or not to scale grades until the very end of the semester.

(subject to possible changes)

Home Task	
Deadline:	strictly before the stated deadline – **NO EXTENSIONS**
Setup:	R programming language software
Format:	application of appropriate models and statistical methods by R
Submission:	electronically

Midterm Exam	
Content:	application of appropriate models and statistical methods by R
Duration:	2 hours
Setup:	computer with R programming language
Submission:	electronically
RIGID RESTRICTION: no communication	
Final Exam	
Content:	application of appropriate models and statistical methods by R
Duration:	2 hours
Setup:	computer with R programming language
Submission:	electronically
RIGID RESTRICTION: no communication	

Evaluation Rubrics*(subject to possible minor changes)*

Task Type	Completion Quality	Score
Exam task	Any collaboration	0%
	Wrong algorithm – wrong implementation	up to 40%
	Wrong algorithm – correct implementation	~ 60%
	Correct algorithm – wrong implementation	~80%
	Correct algorithm – correct implementation	~100%
Home task	Unstated duplication	max 0%
	Stated collaborative work (per collaborator)	max 40%
	Wrong/nonworking solution	max 40%
	Correct/working solution	100%

Bonus

The instructor reserves a right to assign bonus for creative solutions regardless of the correctness.

Exams:
The course will include a mid-term exam covering topics from weeks 1 to 7 as well as a comprehensive final exam covering all course topics with an emphasis on topics covered in weeks 9-17.

The purpose of the mid-term exam is to assess students' progress in learning how to use terminology and basic concepts. Students will receive feedback through the in-class mid-term review as well as written feedback on their mid-term exams. Students are encouraged to attend office hours for more individualized guidance.

The purpose of the final exam is to assess students' mastery of concepts and terminology as well as their abilities to select appropriate methods and apply them to analyze data sets and solve complex problems.

Students are welcome to confer with the instructor for more individualized feedback.

Library and Media/Technology Use

Students are encouraged to use supplemental online and reference materials to enhance their overall learning in the course. If students have any questions or need additional support in using library resources or technology, they should confer with library staff, ICT, or the instructor.

Late Policy

No late works.

Make-up Procedures

Make-up assignment, exam, and quiz will be given at the instructor's discretion. Students must submit convincing evidence of a medical or other emergency that makes completing an assignment or taking an exam or quiz at the scheduled time impossible.

Policy on Grade Appeal

Students are entitled to appeal grades in line with the university's Grades Policies policy which is available online athttp://policies.aua.am

Standards for Academic Integrity

Students are required to conduct themselves in an academically responsible and ethical manner in line with the Student Code of Ethics. Acts of academic dishonesty impair the academic integrity of AUA and create an unfair academic advantage for the student involved and other member(s) of the academic community. These acts are subject to disciplinary measures as prescribed in the AUA Code of Student Ethics http://policies.aua.am/policy/10

The Student Code of Conduct can be found at http://policies.aua.am/policy/101

Special Needs

Students requiring special accommodations for learning should contact the Center for Student Success by the end of the Drop/Add period with such requests.studentsuccess@aua.am,http://studentsuccess.aua.am/disability-support-services/