CUNY School of Professional Studies

Syllabus
School of Professional Studies
DATA 622: Introduction to Machine Learning: Supervised Learning

Instructor Name: Raman Kannan
Instructor Email Address: Raman.Kannan@sps.cuny.edu
Degree Program: M.S. in Data
Science **Credits**: 3 graduate credits
Prerequisites: *605,606, 621*
Type of Course: Elective

Course Description

The course will develop basic understanding of supervised learning techniques, a generic process method to execute classification exercise, leveraging the Statistics and Linear Algebra students have acquired in 605 and 606.

This is a clinical course and emphasis is on computational techniques: designing, running and refining supervised learning models introduced in the course. We will be primarily using R as our programming environment. All assignments will need to be submitted in as R-Markdown documents or executable R scripts on cloud computing resources the instructor will provide free of cost. Students are required to demonstrate

> *1. good understanding of what supervised learning/classification models are; and*

> *2. proficiency and experience in running classification exercise on any given dataset*

> *3. Familiarity and mastery over technical terminology expected of a practitioner.*

Program Learning Outcomes

1. Conceptual Understanding of Classical Supervised Learning Techniques. Understand the statistical/algebraic foundation of classification algorithms, relative strengths and weakness of various techniques, theoretical and practical criteria in adopting a model.

2. Data Understanding. Collect, describe, model, explore and verify data.

3. Engineering. Use industry standard environment and process to conduct repeatable and reproducible classification exercise.

4. Experimentation and Analysis: Run prescribed process to optimize model using multiple classification algorithms.

5. Presentation: deliver summary results of the experiments and explain key decisions they made in designing the model and model output.

Learning Objectives:

1. Develop deep understanding of both parametric/non-parametric classification algorithms.

2. Prepare datasets for classification algorithms, plan and conduct modeling exercise on given datasets.

3. Apply techniques introduced to reduce overfitting and mitigate the adverse effects of variance and bias.

Assignments and Grading

Grades in this course are calculated by the percentage of points obtained.

Course assignments	Percentage of Final Grade	Points
Homework – Run the model exercise	30.00%	30
✓ There will be 2 individual modeling assignments (15% each) ✓ Grading based on process discipline (3%) and ability to prepare data (6%) and run the model (6%)	15 x 2% = 30%	
Class Discussion	12.00%	12
✓ You must participate in weekly forums and discussions. ✓ Discussions are applied analysis from the texts. ✓ You must post a response by Wednesday at midnight (ET) ✓ You must respond to at least one of your colleagues' contributions by Saturday at midnight (ET), ✓ You should provide meaningful feedback on the analysis.	12 x 1% = 12%	
Homework – demonstrate ability to improve performance	40%	40
✓ There will be 2 individual modeling assignments (20% each) ✓ Grading based on computing specified performance measures (12%), and 8% for systematically seeking to improve performance using prescribed methods	20x2=40	
✓ Open book, open notes review test – t1	9.00%	9
✓ Open book, open notes review test – t2	9	9
Total	**100%**	**100**

Late Policy for Homework:

Late work will not be graded. All assignment due dates and times are shown in Blackboard.

Course Materials (All Open Source)

TITLE: Understanding Machine Learning, CODE:UML (AGGRESSIVE)

URL:https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf

AUTHORS: Shai Shalev Shwartz and Shai Ben David

(this book is of the class books by Christopher M. Bishop and book by Kevin P Murphy)

URL:http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

TITLE: Elements of Statistical Learning, CODE:ESL (AMBITIOUS STUDENTS)

URL:http://www-bcf.usc.edu/~gareth/ISL/

AUTHORS: Trevor Hastie, Robert Tibshirani, Jerome Friedman

TITLE: An Introduction to Statistical Learning, CODE:ISL

RECOMMENDED

URL:http://faculty.marshall.usc.edu/gareth-james/ISL/ (booksite, pdf,data,R etc)

http://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf (pdf)

AUTHORS: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

SWARM Intelligence – our collective conscience-wisdom

https://www.quora.com/What-is-the-best-book-to-learn-ML

Papers

P1:

P2:

Relevant Software and Other Tools:

Students should have R Studio & R installed in their computers. Relevant libraries that are required will be posted in the assignments and course materials. Students are expected to submit R-Markdown files for their assignments. ***Or submit R Script on IBM Cloud. Instructor will provide access to IBM Cloud once registration is stabilized, fully sponsored by IBM Power Systems Academic Initiative.***

Instructor will demonstrate all work exclusively on IBM PSAI Cloud Node.

Course Meeting Time:

See the course website.

Grade Distribution

Quality of Performance	Letter Grade	Range %	GPA/ Quality Pts.
Excellent - work is of exceptional quality	A	93 - 100	4.0
	A-	90 - 92.9	3.7
Good - work is above average	B+	87 - 89.9	3.3
Satisfactory	B	83 - 86.9	3.0
Below Average	B-	80 - 82.9	2.7
Poor	C+	77 - 79.9	2.3
	C	70 - 76.9	2.0
Failure	F	< 70	0.0

Course Outline:

Please note that this schedule is subject to change depending on our progress, questions, requests, etc.

Week	Week of	Topics	Reading	Assignment	Due Sunday at 11:59pm
1	27-Jan	Introduction to Learning, Machine Learning	R1		2-Feb
2	3-Feb	Refresher Probability/Statistics/Dataset manipulation in R	R2		9-Feb
3	10-Feb	Advanced Probability and Linear Algebra	R3		16-Feb
4	17-Feb	Machine Learning, Supervised Learning, Classification, Method	R4		23-Feb
5	24-Feb	Continuous Numerical data, Linear, Polynomial models and Categorical data	R5		1-Mar
6	2-Mar	Generic Concepts applicable to all supervised learners: Performance, Comparative	R6		8-Mar
7	9-Mar	Fisher's Discriminant Analysis: LDA/QDA	R7		15-Mar
8	16-Mar	Uncorrelated Features	R8		22-Mar
9	23-Mar	Instance Based techniques (no assumptions about the distribution, aka non-	R9		29-Mar
10	30-Mar	Support Vectors SVM: Vapnik LearningTheory Margins, Kernel, Radial	R10		5-Apr
11	6-Apr	Spring Recess			12-Apr
12	13-Apr	Decision Trees (no assumptions about the distribution, aka non-	R11		19-Apr
13	20-Apr	Resampling: Varying dataset Bootstrapping, Cross	R13		26-Apr
14	27-Apr	Comparing Classifier performance	R14		3-May
15	4-May	What causes inferior performance, techniques for performance improvement	R15		10-May
16	11-May	Combining Classifiers: Aggregating variants of one classifier	R16		17-May
17	22-May	Combining Heterogeneous Classifiers Stacking: Voting Mechanisms	R17	Final Exam	22-May

Accessibility and Accommodations

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. For more information, please see:

[Disability Services on the CUNY SPS Website.](#)

Online Etiquette and Anti-Harassment Policy

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: ["Netiquette in an Online Academic Setting: A Guide for CUNY School of Professional Studies Students."](#)

Academic Integrity

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see:

[Academic Integrity on the CUNY SPS Website.](#)

Student Support Services

If you need any additional help, please visit [Student](#) [Support](#) [Services](#)

Notes on Course: