Course Syllabus

Course No. & Title: EEL 6935, Advanced Data Analytics
Level: Graduate   **Credits:** 3
Background: This course is intended for graduate students who major in Electrical Engineering or a related field, and have experience in machine learning and programming.
Term & Schedule: Fall, 2019   Fr 11:00AM – 1:45PM ENG 201
Instructor: Dr. Yasin Yilmaz
                         yasiny@usf.edu
                         Office Hours: Mo, Th 12:00-1:00PM in ENB 251

Course Description: This is a research-oriented course that focuses on Big Data challenges by teaching useful probabilistic and statistical methods, and providing hands-on experience. The topics of interest include dimensionality reduction, mixture models, graphical models, exact inference, variational inference, and deep learning. Students will read and discuss in the class interesting papers from the literature. There will also be homework assignments (coding-based and written), and a final project proposed by the student. It will typically be an application of state-of-the-art techniques to new Big Data problems, or possibly an extension of the existing methods. Project report will be in conference paper format.

Course Objectives: To provide students with state-of-the-art machine learning methods and an ability to apply them to Big Data problems in various applications.

Course Prerequisites: EEE 6935 Data Analytics (at least with a grade B) or instructor's approval

Textbook Info: There is no required textbook. See Supplementary Books towards the end of syllabus.

Test & Grading Info: There will be 5 homework assignments, 5 paper discussions, and a final project. A missed homework, exam, or project will be graded zero. Make-ups and deadline extensions will be allowed at the instructor's consent, only under special circumstances; appropriate documentation in writing must be provided.

Homework: Homework assignments will include theoretical and applied (coding) questions. They will be due in one week after the post date. There will be 5 homework assignments.

Paper Discussions: There will be 5 paper discussions with one presenter, one scribe and two commentators. Presenter will first introduce the paper and present his/her comments in 10 minutes. Each commentator will be asked a question to comment on. Scribe's responsibility is to write a 2-page summary, and submit it in one day after the paper discussion. The presenter, commentators and scribe for each paper will be selected randomly in class at the time of discussion. Hence, you are supposed to read all papers and come ready to discuss in class. You will be in each role once throughout the semester.

Final Project: Final project consists of applying the learned techniques to a complex problem proposed by the student. Proposed datasets must be approved by the instructor to avoid unexpectedly low grades (including zero if the proposed dataset and/or analysis is inappropriate). A one-page preliminary report which introduces the dataset and the proposed analysis is required due 10/11. Students will need to make an oral presentation at the end of the semester (typically during the last couple of weeks). A final report (max. 10 page) documenting the analysis and results is due 12/9.

Programming: Python will be used in this course.

Grading Breakdown:

Homework	35% (7% each)	
Paper Discussion	20% (10% presenting, 5% commenting, 5% scribing)	
Final Project	45% (5% preliminary report + 15% oral presentation + 25% final report)	

Course Topics and Approximate # of Lectures:

Fundamentals (2)
> *Machine Learning, Frequentist and Bayesian Probability, Probability Distributions, Optimization, Python Basics*

Latent Variable Models (2)
> *Probabilistic PCA, Factor analysis, Kernel PCA, Manifold learning, Laplacian eigenmaps, Gaussian Mixture Model, Bayesian Mixture Models, Latent Dirichlet Allocation*

Graphical Models (3)
> *Bayesian Networks, Markov Networks, Gaussian Networks*

Exact Inference (1)
> *Variable Elimination, Message Passing*

Approximate Inference (2)
> *Laplace Approximation, Variational Inference, Expectation Propagation, MCMC Methods, Gibbs Sampling, Importance Sampling*

Deep Learning (1)
> *Variational Autoencoders*

Course Schedule (Tentative)

Date	Class Title	Class Content	Due Items
W1: 8/30	Fundamentals	Review of Machine Learning Concepts (MLPP §1, BRML §13)	
W2: 9/6	Fundamentals	Frequentist and Bayesian Probability, Probability Distributions, Optimization, Python Basics	
W3: 9/13	Latent Variable Models	Factor Analysis, Probabilistic PCA, Kernel PCA, Manifold Learning, Laplacian Eigenmaps (FML §12, MLPP §14.4, BRML §15.7)	HW 1
W4: 9/20	Latent Variable Models	Gaussian Mixture Model, Bayesian Mixture Models (MLPP §11, BRML §20) Latent Dirichlet Allocation (MLPP §27.3, BRML §20)	Paper 1
W5: 9/27	**No Class**		HW 2

W6: 10/4	Graphical Models	Bayesian Networks (PGM §3, MLPP §10, BRML §3)	Paper 2
W7: 10/11	Graphical Models	Markov Networks (PGM §4, MLPP §19, BRML §4) Gaussian Networks (PGM §7, MLPP §10.2,19.4)	Project Preliminary Report
W8: 10/18	Exact Inference	Variable Elimination (PGM §9, MLPP §20, BRML §5), Belief Propagation (PGM §10, MLPP §20, BRML §5)	HW 3
W9: 10/25	Approximate Inference	Laplace Approximation, Quadratic Bounds (MLPP §8.4,21.8, BRML §28.2), Mean Field (MLPP §21.5, BRML §28.4)	Paper 3
W10: 11/1	Approximate Inference	Sampling Methods (MLPP §23,24, BRML §27)	HW 4
W11: 11/8	Deep Learning	Artificial Neural Networks, Backpropagation,	Paper 4
W12: 11/15	Deep Learning	Variational Autoencoders	Paper 5
W13: 11/22	**Final Project Presentations**		
W14: 11/29	**No Class**	Thanksgiving	HW 5
W15: 12/6	**No Class**	Reading Days	
12/9			Final Project Reports

Supplementary Books:
(BRML) *Bayesian Reasoning and Machine Learning,* David Barber, Cambridge University Press, 2012, ISBN 978-0521518147
Free online version is available at the author's webpage: http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/090310.pdf

(MLPP) *Machine Learning: A Probabilistic Perspective,* Kevin Murphy, MIT Press, 2012, ISBN 978-0262018029

(PGM) *Probabilistic Graphical Models*, Daphne Koller, and Nir Friedman, MIT Press, 2009, ISBN 978-0262013192

Foundations of Machine Learning, Mehryar Mohri, et al., MIT Press, 2018, ISBN 9780262039406
https://mitpress.mit.edu/books/foundations-machine-learning-second-edition
Discount Code: MTSR20

Introduction to Deep Learning, Eugene Charniak, MIT Press, 2019, ISBN 9780262039512
https://mitpress.mit.edu/books/introduction-deep-learning
Discount Code: MTSR20

Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016, ISBN 978-0262035613

Detection of Abrupt Changes – Theory and Application, Michele Basseville, and Igor V. Nikiforov, Prentice Hall, 1993, ISBN 978-0131267800
Free online version is available at the author's webpage:
http://people.irisa.fr/Michele.Basseville/kniga/kniga.pdf

Introduction to Algorithms, Thomas H. Cormen, Charles E. Leiserson, Ronald Rivest, and Clifford Stein, MIT Press, 2009, ISBN 978-0262033848

The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Trevor Hastie, et al., Springer, 2009, ISBN 978-0387848570

All of Statistics: A Concise Course in Statistical Inference, Larry Wasserman, Springer, ISBN 978-0387402727

Papers:

1) "Anomaly detection in partially observed traffic networks", Elizabeth Hou, Yasin Yilmaz, and Alfred O. Hero, IEEE Transactions on Signal Processing, 2019
2) "Latent Dirichlet Allocation", David Blei, Andrew Ng, and Michael Jordan, NIPS 2002
3) "Using the Forest to See the Trees: A Graphical Model Relating Features, Objects, and Scenes", Kevin Murphy, Antonio Torralba, William Freeman, NIPS 2004
4) "Variational bounds for mixed-data factor analysis", Mohammad E. Khan, Guillaume Bouchard, Kevin P. Murphy, Benjamin M. Marlin, NIPS, 2010
5) "Tutorial on Variational Autoencoders", Carl Doersch, arXiv

Public Datasets:
http://archive.ics.uci.edu/ml/index.html
https://www.kaggle.com/
http://www.dmoz.org/Computers/Artificial_Intelligence/Machine_Learning/Datasets/
http://funapp.cs.bilkent.edu.tr/DataSets/
https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research
https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning/
http://mldata.org
https://medium.com/@olivercameron/20-weird-wonderful-datasets-for-machine-learning-c70fc89b73d5#.ezscri9e0
http://www.kdnuggets.com/datasets/index.html
http://en.openei.org/wiki/Main_Page (energy)
https://data.nrel.gov (renewable energy)
https://www.rita.dot.gov/bts/data_and_statistics/index.html (transportation)
http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml (taxi)
https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew/data (taxi)
http://www.gapminder.org/data/ (world)
https://archive.stsci.edu/k2/ (space telescopes)

Note: For students taking this course as the Portfolio Course: the final grade may be affected by the evaluation of the student's MSEE Portfolio, as described in the EE Department Portfolio Guidelines.

Academic Integrity:
The faculty of the Electrical Engineering Department is committed to maintaining a learning environment which promotes academic integrity and the professional obligations recognized in the IEEE Code of Ethics (http://ee.eng.usf.edu/about/codeOfEthics.htm). Accordingly, the department adheres to a common Academic Integrity Policy in all of its courses. This policy is to be applied uniformly in a fair and unbiased manner.

University rules regarding academic integrity will be strictly enforced. It is not acceptable to copy, plagiarize or otherwise make use of the work of others in completing homework, project, laboratory report, exam or other course assignments. Likewise, it is not acceptable to knowingly facilitate the copying or plagiarizing of one's own work by others in completing homework, project, laboratory report, exam or other course assignments. It is only acceptable to give or receive assistance from others when expressly permitted by the instructor. Unless specified otherwise, as in the case of all take-home exams, scholarly exchange regarding out-of-class assignments is encouraged. A more complete explanation of behaviors that violate academic integrity is provided at:
 http://www.ugs.usf.edu/catalogs/1112/pdf/AcademicIntegrityOfStudents.pdf.

The minimum penalty for violation of the academic integrity policy stated in the preceding paragraph is the greater of an automatic zero on the assignment or a letter grade reduction in the overall course grade. Student(s) found in violation of the policy on an exam will receive a minimum penalty of an F in the course. All instances of policy violations will be recorded in a letter from the instructor that is kept in the student files held by the department; a copy of the letter will be forwarded to the appropriate (undergraduate or graduate) Dean's office. A second violation of the policy, irrespective of whether it was related to an exam or any other course assignment, will result in a course grade of "FF" and expulsion from the Electrical Engineering Department.

At the instructor's discretion the penalties associated with the EE Department's Academic Integrity Policy may be stricter, in which case further explanation is provided in the following.
Modifications to the Uniform Academic Policy: <none>

Property – you are not granted permission to sell notes or tapes of class lectures.

Students in need of academic accommodations for a disability may consult with the office of Students with Disabilities Services to arrange appropriate accommodations. Students are required to give reasonable notice prior to requesting an accommodation. Contact SDS at 974-4309 or www.sds.usf.edu