FLORIDA ATLANTIC UNIVERSITY

NEW/CHANGE PROGRAM REQUEST
Graduate Programs

UGPC Approval _____
UFS Approval _____
Banner _____
Catalog _____

Department Mathematical Sciences

College Science

Program Name

MS with Major in Data Science and Analytics, Data Science via Scientific Inquiry Concentration.

☐ **New Program***

☑ **Change Program***

Effective Date *(TERM & YEAR)*
Summer 2020

Please explain the requested change(s) and offer rationale below or on an attachment.

In this proposal we request the following changes:

- add STA 5195 (Biostatistics) as required Common Core Course (and therefore remove this course from the list of elective Concentration Courses)

- move CAP 6673 (Data Mining and Machine Learning) to the list of elective Core Courses for Scientific Inquiry Concentration

- remove BSC 6459 (Biomedical Data and Informatics) from the list of elective Core Courses for all concentrations

We are requesting this change in order

- to be consistent with the program change proposed by the supervision committee of the Master Program in Data Science and Analytics

- to make the program more appropriate for our concentration

*All new programs and changes to existing programs must be accompanied by a catalog entry showing the new or proposed changes.

Faculty Contact/Email/Phone

Dr Vincent NAUDOT/vnaudot@fau.edu/561 297 1339

Consult and list departments that may be affected by the change(s) and attach documentation

The supervision committee of the Master Program in Data Science and Analytics

Approved by	***Date***
Department Chair R. Moraga	Feb 4, 2020
College Curriculum Chair Christopher Beetle 2020.03.06 11:40:55 -05'00'	
College Dean William David Kalies	March 9, 2020
UGPC Chair	
UGC Chair	
Graduate College Dean	
UFS President	
Provost	

Email this form and attachments to UGPC@fau.edu 10 days before the UGPC meeting.

Master of Science with Major in Data Science and Analytics (New program effective fall 2019.)

The Master of Science with Major in Data Science and Analytics (MSDSA) is a multi-college interdisciplinary program jointly administered by the Department of Mathematical Sciences in the Charles E. Schmidt College of Science, the Department of Computer & Electrical Engineering and Computer Science in the College of Engineering and Computer Science, the Department of Information Technology and Operations Management in the College of Business and the Department of Political Science in the Dorothy F. Schmidt College of Arts and Letters. The program aims to prepare students with essential skill sets needed to analyze small, fast, big, massive and complex data. To allow for maximum flexibility in career aspirations, students may select from four concentrations:

- Data Science via Scientific Inquiry Concentration, Department of Mathematical Sciences.
- Data Science and Engineering Concentration, Department of Computer & Electrical Engineering and Computer Science.
- Data Science in Business Concentration, Department of Information Technology and Operations Management.
- Data Science in Society Concentration, Department of Political Science.

Admission Requirements

To be admitted to the MSDSA program, applicants must:

1. Have obtained a bachelor's degree from an accredited institution and possess a minimal background consisting of MAC 2233 (Methods of Calculus) or equivalent and STA 2023 (Introductory Statistics) or equivalent. ~~and computer programming (COP 2220 or MAD 2502) or equivalent~~. Students applying to the Data Science and Engineering concentration must have completed a college-level introductory programming course with the grade of C or better. Knowledge of Python and statistical packages such as R, as well as coursework in linear algebra are recommended for all concentrations;

2. Have an undergraduate GPA of 3.0 or higher in the last 60 credits of undergraduate coursework;

3. Submit two letters of recommendation for all the concentrations, except the Data Science and Engineering concentration;

4. Have attained scores of at least 151 (verbal) and 151 (quantitative) on the Graduate Record Examination (GRE). GRE scores more than five years old are not acceptable normally. The Data Science and Engineering concentration requires the submission of the GRE score (verbal and quantitative sections), but no minimum values are required;

5. Be proficient in written and spoken English. International students from non-English-speaking countries must present a score of at least 500 (paper-based test) or 213 (computer-based test) or 79 (internet-based test) on the Test of English as a Foreign Language (TOEFL) or a score of at least 6.0 on the International English Language Testing System (IELTS); and

6. Meet other requirements of the FAU Graduate College.

Curriculum Requirements

The MSDSA program offers both thesis and non-thesis options. Both options require a minimum of 30 credits. Students are required to take one common core course, two additional core courses, four concentration courses and three elective courses for the total of 30 credits. The exact courses taken are to be determined by the students and their advisory committee. The thesis option requires only one elective course and 6 thesis credits. Students selecting the thesis option must complete and defend a written thesis successfully.

Data Science via Scientific Inquiry Concentration

Common Core Courses		
Introduction to Data Science	CAP 5768	3
Biostatistics	STA 5195	3
Take ~~two~~ one additional core courses		
Data Mining and Machine Learning	CAP 6673	3 **or**
~~Biomedical Data and Informatics~~	~~BSC 6459~~	~~3 or~~
Introduction to Business Analytics and Big Data	ISM 6404	3 **or**
Special Topics (Quantitative Methods)	POS 6934	3
Take four concentration courses		
Computer Data Security	CIS 6370	3
Cyber Security: Measurement and Data Analysis	CTS 6319	3
Introduction to Cryptology and Information Security	MAD 5474	3
Applied Computational Topology	MTG 6329	3
Graph Theory	MAD 6307	3
Cryptanalysis	MAD 6478	3
Biostatistics	STA 5195	3
Statistical Computing	STA 6106	3
Survival Analysis	STA 6177	3
Regression Analysis	STA 6236	3
Mathematical Statistics	STA 6326	3
Applied Time Series Analysis	STA 6857	3
Take three elective courses from the Electives Table. Thesis option requires only one elective course and 6 thesis credits.		

Data Science and Engineering Concentration

Common Core Courses		
Introduction to Data Science	CAP 5768	3
Data Mining and Machine Learning	CAP 6673	3
Take ~~two~~ one additional core courses		
~~Biomedical Data and Informatics~~	~~BSC 6459~~	~~3 or~~
Biostatistics	STA 5195	3 **or**

Introduction to Business Analytics and Big Data	ISM 6404	3 **or**
Special Topics (Quantitative Methods)	POS 6934	3
Take four concentration courses		
Introduction to Neural Networks	CAP 5615	3
Social Networks and Big Data Analytics	CAP 6315	3
Data Mining for Bioinformatics	CAP 6546	3
Machine Learning for Computer Vision	CAP 6618	3
Deep Learning	CAP 6619	3
Information Retrieval	CAP 6776	3
Web Mining	CAP 6777	3
Advanced Data Mining and Machine Learning	CAP 6778	3
Big Data Analytics with Hadoop	CAP 6780	3
Computational Advertising and Real-Time Analytics	CAP 6807	3
Computer Performance Modeling	CEN 6405	3
Take three elective courses from the Electives Table. Thesis option requires only one elective course and 6 thesis credits.		

Data Analytics in Business Concentration

Common Core Course~~s~~		
Introduction to Data Science	CAP 5768	3
Introduction to Business Analytics and Big Data	ISM 6404	3
Take ~~two~~ one additional core course~~s~~		
~~Biomedical Data and Informatics~~	~~BSC 6459~~	~~3 or~~
Biostatistics	STA 5195	3 **or**
Data Mining and Machine Learning	CAP 6673	3 **or**
Special Topics (Quantitative Methods)	POS 6934	3
Take four concentration courses		
Quantitative Communication Research	COM 6316	3
Data Mining and Predictive Analytics	ISM 6136	3
Database Management Systems	ISM 6217	3
Advanced Business Analytics	ISM 6405	3
Social Media and Web Analytics	ISM 6555	3
Data Management and Analysis with Excel	QMB 6303	3

Data Analysis for Managers	QMB 6603	3
Take three elective courses from the Electives Table. Thesis option requires only one elective course and 6 thesis credits.		

Data Science in Society Concentration

Common Core Courses		
Introduction to Data Science	CAP 5768	3
Special Topics (Quantitative Methods)	POS 6934	3
Take ~~two~~ one additional core courses		
~~Biomedical Data and Informatics~~	~~BSC 6459~~	~~3 or~~
Biostatistics	STA 5195	3 **or**
Data Mining and Machine Learning	CAP 6673	3 **or**
Introduction to Business Analytics and Big Data	ISM 6404	3
Take four concentration courses		
Advanced Anthropological Research 2	ANG 6092	3
Quantitative Reasoning in Anthropological Research	ANG 6486	3
Social Networks and Big Data Analytics	CAP 6315	3
Quantitative Communication Research	COM 6316	3
Social Media and Web Analytics	ISM 6555	3
Seminar in Political Behavior	POS 6208	3
Research Design in Political Science	POS 6736	3
Seminar in Advanced Research Methods	SYA 6305	3
Take three elective courses from the Electives Table. Thesis option requires only one elective course and 6 thesis credits.		

Electives Table		
Business Analytics		
Data Mining and Predictive Analytics	ISM 6136	3
Database Management Systems	ISM 6217	3
Introduction to Business Analytics and Big Data	ISM 6404	3
Advanced Business Analytics	ISM 6405	3
Social Media and Web Analytics	ISM 6555	3
Data Management and Analysis with Excel	QMB 6303	3

Data Analysis for Managers	QMB 6603	3
Database and Cloud Computing		
Multiprocessor Architecture	CDA 6132	3
Cloud Computing	CEN 5086	3
New Directions in Database Systems	COP 6726	3
Theory and Implementation of Database Systems	COP 6731	3
Database Management Systems	ISM 6217	3
Data Mining and Machine Learning		
Introduction to Neural Networks	CAP 5615	3
Social Networks and Big Data Analytics	CAP 6315	3
Data Mining for Bioinformatics	CAP 6546	3
Machine Learning for Computer Vision	CAP 6618	3
Deep Learning	CAP 6619	3
Data Mining and Machine Learning	CAP 6673	3
Information Retrieval	CAP 6776	3
Web Mining	CAP 6777	3
Advanced Data Mining and Machine Learning	CAP 6778	3
Big Data Analytics with Hadoop	CAP 6780	3
Computational Advertising and Real-Time Analytics	CAP 6807	3
Computer Performance Modeling	CEN 6405	3
Data Mining and Predictive Analytics	ISM 6136	3
Data Security and Privacy		
Computer Data Security	CIS 6370	3
Cyber Security: Measurement and Data Analysis	CTS 6319	3
Management of Information Assurance and Security	ISM 6328	3
Introduction to Cryptology and Information Security	MAD 5474	3
Cryptanalysis	MAD 6478	3
Quantum Mechanics 2	PHY 6646	3
Scientific Applications and Modeling		
Photogrammetry and Aerial Photography Interpretation	GIS 6028C	3
LiDAR Remote Sensing and Applications	GIS 6032C	3
Web GIS	GIS 6061C	3
Geospatial Databases	GIS 6112C	3
Hyperspectral Remote Sensing	GIS 6127	3
Spatial Data Analysis	GIS 6306	3

Special Topics (Quantum Information Processing)	PHY 6938	3
Computational Physics	PHZ 5156	3
Numerical Relativity	PHZ 7609	3
Social Data Science		
Advanced Anthropological Research 1	ANG 6090	3
Advanced Anthropological Research 2	ANG 6092	3
Quantitative Reasoning in Anthropological Research	ANG 6486	3
Social Networks and Big Data Analytics	CAP 6315	3
Quantitative Communication Research	COM 6316	3
Special Topics (Quantitative Methods)	POS 6934	3
Research Design in Political Science	POS 6736	3
Seminar in Advanced Research Methods	SYA 6305	3
Statistics and Data Applications		
~~Biomedical Data and Informatics~~	~~BSC 6459~~	~~3~~
~~Biostatistics~~	~~STA 5195~~	~~3~~
Statistical Computing	STA 6106	3
Survival Analysis	STA 6177	3
Biostatistics - Longitudinal Data Analysis	STA 6197	3
Applied Statistical Methods	STA 6207	3
Regression Analysis	STA 6236	3
Mathematical Statistics	STA 6326	3
Applied Time Series Analysis	STA 6857	3
Applied Computational Topology	MTG 6329	3

MS in Data Science and Analytics Program Changes - Spring 2020

[1] Remove BSC 6459, Biomedical Data and Informatics, from the degree program. This change is supported by the College of Medicine (letter attached).

[2] Add STA 5195, Biostatistics, to the list of common core courses for all concentrations (and remove it from the list of elective courses).

[3] In the Data Science via Scientific Inquiry concentration, list STA 5195 as required common core and move CAP 6673, Data Mining and Machine Learning, as elective common core.

[4] Add MTG 6329 to the elective courses and to the concentration electives for the Data Science via Scientific Inquiry concentration. These courses were newly proposed with the new degree program in Spring 2019 but did not follow the faculty governance process with the new degree for some unknown reason. They were listed in the original curriculum approved by the UFS in April 2019. New course proposals are accompanying this change request.

[5] Remove computer programming (COP 2220 or MAD 2502) from the admission requirements from the degree except for the engineering concentration.

[6] Some cosmetic changes to the catalog arrangement of the core and concentration courses to make it more understandable for students.

Approved by the MSDSA Oversight Committee meeting on March 2, 2020 with Kevin Wagner (chair), Vincent Naudot, Tahgi Koshgaftaar, Tamara Dinev (by phone), and William Kalies (ex-officio).

From: **Kevin Wagner** kwagne15@fau.edu
Subject: Re: MSDSA changes
Date: March 2, 2020 at 2:16 PM
To: William Kalies WKALIES@fau.edu
Cc: Tamara Dinev tdinev@fau.edu, Hanqi Zhuang zhuang@fau.edu, Rainer Steinwandt mathfaculty@lists.fau.edu, Mihaela Cardei mcardei@fau.edu, Vincent Naudot vnaudot@fau.edu, Christopher Beetle cbeetle@fau.edu

Yes, I approve.

KMW

Kevin M. Wagner, J.D., PhD
Professor and Chair, Department of Political Science
President, FAU Faculty Senate
Trustee, FAU Board of Trustees
Director of the Jack Miller Forum
Dorothy F. Schmidt College of Arts and Letters
Florida Atlantic University
777 Glades Road
Boca Raton, FL 33431
tel: 561-252-1794
fax: 561-297-2997
kwagne15@fau.edu
Twitter: @kevinwagnerphd
www.fau.edu/politicalscience

On Mon, Mar 2, 2020 at 2:12 PM William Kalies <WKALIES@fau.edu> wrote:

> Dear Tamara, Kevin, and Hanqi
>
> The attached changes have been approved by the MSDSA steering committee today as well as the Department of Mathematical Sciences. These changes will proceed through the College of Science Graduate Committee this week and submitted to the UGPC by March 9. Please reply to this email supporting these changes.
>
> Thank you
> Bill
>
> —
>
> Bill Kalies
> Associate Dean for Graduate Studies
> Charles E. Schmidt College of Science
> Professor of Mathematical Sciences
>
> Florida Atlantic University
> 777 Glades Rd, SE-43, Room 242
> Boca Raton, FL 33431
> tel: 561-297-1107
>
> —

From: **Tamara Dinev** tdinev@fau.edu
Subject: RE: MSDSA changes
Date: March 2, 2020 at 7:27 PM
To: William Kalies WKALIES@fau.edu, Kevin Wagner kwagne15@fau.edu, Hanqi Zhuang zhuang@fau.edu
Cc: Rainer Steinwandt mathfaculty@lists.fau.edu, Mihaela Cardei mcardei@fau.edu, Vincent Naudot vnaudot@fau.edu, Christopher Beetle cbeetle@fau.edu

TD

ITOM approves, thank you

Best Regards:
Tamara
==========
Tamara Dinev, Ph.D., Department Chair and Professor
Dean's Distinguished Research Fellow
Department of Information Technology and Operations Management, FL 219
College of Business, Florida Atlantic University
Boca Raton, Florida 33431
tel. (561) 297-3181, email: tdinev@fau.edu
Google Scholar: https://scholar.google.com/citations?user=YH8QZ-YAAAAJ&hl=en

---

From: William Kalies <WKALIES@fau.edu>
Sent: Monday, March 2, 2020 2:13 PM
To: Tamara Dinev <tdinev@fau.edu>; Kevin Wagner <kwagne15@fau.edu>; Hanqi Zhuang <zhuang@fau.edu>
Cc: Rainer Steinwandt <mathfaculty@lists.fau.edu>; Mihaela Cardei <mcardei@fau.edu>; Vincent Naudot <vnaudot@fau.edu>; Christopher Beetle <cbeetle@fau.edu>
Subject: MSDSA changes

Dear Tamara, Kevin, and Hanqi

The attached changes have been approved by the MSDSA steering committee today as well as the Department of Mathematical Sciences. These changes will proceed through the College of Science Graduate Committee this week and submitted to the UGPC by March 9. Please reply to this email supporting these changes.

Thank you
Bill

—
Bill Kalies
Associate Dean for Graduate Studies
Charles E. Schmidt College of Science
Professor of Mathematical Sciences

Florida Atlantic University
777 Glades Rd, SE-43, Room 242
Boca Raton, FL 33431
tel: 561-297-1107
—

From: **Taghi Khoshgoftaar** khoshgof@fau.edu
Subject: Re: MS in Data Meeting Documents
Date: March 2, 2020 at 9:39 AM
To: Kevin Wagner kwagne15@fau.edu, William Kalies WKALIES@fau.edu, William Trapani wtrapan1@fau.edu, Mihaela Cardei mcardei@fau.edu, Tamara Dinev tdinev@fau.edu, Vincent Naudot vnaudot@fau.edu

TK

Hello Kevin,

It seems to me everyone agrees with the proposed changes. Do we need to meet? Please advise.

Thanks.

Taghi

Taghi M. Khoshgoftaar, PhD
Motorola Endowed Chair Professor
Graduate Advisor for Big Data Analytics Certificate Program
Co Editor-in-Chief of Journal of Big Data
Director of NSF Big Data Training and Research Lab
Director of Data Mining and Machine Learning Lab
Department of Computer and Electrical Engineering and Computer Science
777 Glades Road
Boca Raton, Florida 33433, USA
https://scholar.google.com/citations?user=-PgNSCAAAAAJ&hl=en
http://www.cse.fau.edu/~taghi/
khoshgof@fau.edu
Tel: (561) 297-3994
Fax: (561) 297-2800

From: Kevin Wagner <kwagne15@fau.edu>
Sent: Thursday, February 20, 2020 3:33 PM
To: William Kalies <WKALIES@fau.edu>; William Trapani <wtrapan1@fau.edu>; Taghi Khoshgoftaar <khoshgof@fau.edu>; Mihaela Cardei <mcardei@fau.edu>; Tamara Dinev <tdinev@fau.edu>; Vincent Naudot <vnaudot@fau.edu>
Subject: MS in Data Meeting Documents

Please find attached the documents for our meeting on March 2.

KMW
Kevin M. Wagner, J.D., PhD
Professor and Chair, Department of Political Science
President, FAU Faculty Senate
Trustee, FAU Board of Trustees
Director of the Jack Miller Forum

From: **William Kalies** WKALIES@fau.edu
Subject: Re: MSDSA changes
Date: March 4, 2020 at 1:29 PM
To: Hanqi Zhuang zhuang@fau.edu
Cc: Rainer Steinwandt rsteinwa@fau.edu

Dear Hanqi

Thank you for getting back to us quickly. I would like to see if there is a way to modify the proposed syllabus so that we can come to some agreement. Could you send me syllabi of the course(s) in your department that are impacted by this course and what specifically the topics of concern are? I appreciate that the deadline for these proposals is short (Monday), and if possible I would like to find a way to proceed.

Thanks,
Bill

—
Bill Kalies
Associate Dean for Graduate Studies
Charles E. Schmidt College of Science
Professor of Mathematical Sciences

Florida Atlantic University
777 Glades Rd, SE-43, Room 242
Boca Raton, FL 33431
tel: 561-297-1107
—

> On Mar 4, 2020, at 11:09 AM, Hanqi Zhuang <zhuang@fau.edu> wrote:
>
> Dear Bill,
>
> Thank you very much for sending us the documents for review. The faculty in our department discussed the course proposal (MAP 6196, Mathematical Foundations for Data Science), and voiced their concern that there was a substantial overlap with a number of the courses currently taught by CEECS in the areas of ML and Data Science. Although the title of the proposed course emphasizes the mathematical foundations, the topics of the courses cover those of ML and Data Science.
>
> Our faculty also reviewed the new course proposal MAP 2190 Mathematics of Data Science, which is included in the new BS DSA program. The course topics in MAP 2190 covers those mathematical tools which are beneficial to students who are interested in Data Science. Clearly, there is a disparity of the topics in the 2 proposed courses, even though they share a similar title at the UG and graduate levels, respectively. The faculty then recommended to revise the topics and course description of MAP 6196 to be aligned with those of MAP 2190. Those mathematical foundations would be very useful for the students in the MSDSA program.
>
> Thanks,
>
> Hanqi
>
> ---
>
> Hanqi Zhuang, Ph.D.
> Professor and Interim Chair
> Department of Computer & Electrical Engineering & Computer Science
> Florida Atlantic University
> Boca Raton, FL 33431
> 561-297-3413
>
> ---
>
> **From:** William Kalies <WKALIES@fau.edu>
> **Sent:** Monday, March 2, 2020 2:12 PM
> **To:** Tamara Dinev <tdinev@fau.edu>; Kevin Wagner <kwagne15@fau.edu>; Hanqi Zhuang <zhuang@fau.edu>
> **Cc:** Rainer Steinwandt <mathfaculty@lists.fau.edu>; Mihaela Cardei <mcardei@fau.edu>; Vincent Naudot <vnaudot@fau.edu>; Christopher Beetle <cbeetle@fau.edu>
> **Subject:** MSDSA changes
>
> Dear Tamara, Kevin, and Hanqi
>
> The attached changes have been approved by the MSDSA steering committee today as well as the Department of Mathematical Sciences. These changes will proceed through the College of Science Graduate Committee this week and submitted to the UGPC by March 9. Please reply to this email supporting these changes.
>
> Thank you
> Bill
>
> —
> Bill Kalies
> Associate Dean for Graduate Studies
> Charles E. Schmidt College of Science
> Professor of Mathematical Sciences
>
> Florida Atlantic University
> 777 Glades Rd, SE-43, Room 242
> Boca Raton, FL 33431
> tel: 561-297-1107
> —

From: **Hanqi Zhuang** zhuang@fau.edu
Subject: Re: Courses similar to the proposed one
Date: March 6, 2020 at 4:11 PM
To: William Kalies WKALIES@fau.edu

HZ

Hi Bill,

Thank you for sending me the revised one. I am forwarding this to our faculty to seek some feedback.

Meanwhile, have a wonderful weekend.

Hanqi

Hanqi Zhuang, Ph.D.
Professor and Interim Chair
Department of Computer & Electrical Engineering & Computer Science
Florida Atlantic University
Boca Raton, FL 33431
561-297-3413

From: William Kalies <WKALIES@fau.edu>
Sent: Friday, March 6, 2020 4:03 PM
To: Hanqi Zhuang <zhuang@fau.edu>
Subject: Re: Courses similar to the proposed one

Dear Hanqi

I have attached a revised syllabus where we emphasize that this course is focused on mathematical proofs in the context of data science. Please let me know if you have any concerns with this syllabus.

Thanks,
Bill

—
Bill Kalies
Associate Dean for Graduate Studies
Charles E. Schmidt College of Science
Professor of Mathematical Sciences

Florida Atlantic University
777 Glades Rd, SE-43, Room 242
Boca Raton, FL 33431
tel: 561-297-1107
—

> On Mar 5, 2020, at 8:15 PM, Hanqi Zhuang <zhuang@fau.edu> wrote:
>
> Dear Bill,
>
> Our CAP 5625 computational foundations of AI covers many topics that are in your proposed course. In addition, many machine learning algorithms proposed in your course are covered in our data mining and machine learning course.
>
> See one of the examples (the syllabus is attached) and the course details are given below

See one of the examples (the syllabus is attached) and the course details are given below.

Please let me know if you have further questions.

Thanks,

Hanqi

More details of the course (CAP 5625)

Lecture 1: Introduction
- Course overview
- Introduction to machine learning
- Linear algebra review
- Probability review
- Single- and multivariate Gaussian distributions and their properties

Lecture 2: Linear regression
- Introduction to multiple linear regression (parametric regression)
- Cost functions and least squares regression
- Convex optimization
- Normal equations
- Hessian matrices
- The "hat" matrix
- Geometric interpretation of least squares regression
- Incorporating qualitative features
- Batch gradient descent
- Mini-batch gradient descent
- Stochastic gradient descent
- Relationship to likelihood model based on Gaussian distributions

Lecture 3: Linear and non-linear regression and model selection
- Modeling non-linear relationships
- Training error vs. test error
- Bias-variance tradeoff for model complexity
- Reducible and irreducible error
- Issues with data that cause bias or inflate variance
- K-nearest neighbors (KNN) regression (non-parametric regression)
- The curse of dimensionality
- Leave-one-out and k-fold cross validation

Lecture 4: Feature selection and regularization
- Feature selection
- Subset selection methods (best subset selection, forward stepwise selection, backward stepwise selection)
- Shrinkage methods (ridge regression, lasso, and elastic net)
- Ridge regression optimization (normal equations, hat matrix, batch, mini-batch, and stochastic gradient descent)
- Singular value decomposition (SVD)
- Rewriting hat matrix for ridge regression as function of SVD decomposition to get degrees of freedom in ridge regression
- Properties of ridge regression
- Properties of lasso
- Compromise between lasso and ridge regression (elastic net)

Lecture 5: Advanced regularization techniques
- Subgradients
- Coordinate descent optimization to find parameter estimates for lasso

- Coordinate descent optimization to find parameter estimates for ridge
- Coordinate descent optimization to find parameter estimates for elastic net
- L1- and L2-norm trend filtering and trend-filtered regression
- Normal equations under L2-norm trend filtered regression
- Hodrick-Prescott filtering
- Fused lasso
- L1-norm graph trend filtering (trend filtering on a graph)
- Grouped lasso

Lecture 6: Principal Components regression
- Principal components analysis (PCA)
- Relationship to eigenvectors and eigenvalues
- PC biplots
- The variance of projected observations
- Variance along a particular eigenvector
- Relationship to singular values
- Principal components regression
- Proportion of variance explained by top PCs
- Scree plots
- Comparison of principal component regression to lasso and ridge regression

Lecture 7: Discriminant analysis
- Classification problems
- Issues with using linear regression for classification (e.g., masking)
- Gaussian assumptions
- Linear discriminant analysis (LDA) to perform linear classification
- Expanding the dimension with LDA to perform non-linear classification
- Quadratic discriminant analysis (QDA) to perform non-linear classification
- Regularize discriminant analysis (tradeoff of QDA and LDA)

Lecture 8: Logistic regression
- Relaxing the Guassian assumption
- Non-differentiability of a sign function to perform classification
- Logistic (binary) regression
- Multinomial (categorical) regression
- Cost function with likelihood
- Batch, minibatch, and stochastic gradient descent to optimize likelihood (cost function)
- K-nearest neighbors classification (KNN) as non-parameteric classification
- Review of Taylor series
- Hessian of the cost function for logistic/multinomial regression
- Newton-Raphson as alternative to optimizing the model parameters or logistic/multinomial regression
- Logistic/multinomial regression with ridge regression penalty
- Trend-filtered logistic/multinomial regression
- Confusion matrices

Lecture 9: Support vector machines
- Properties of separating hyperplanes
- Rosenblatt's perceptron learning algorithm
- Stochastic gradient descent with the perceptron learning algorithm
- Optimal separating hyperplanes
- Quadratic programming
- Lagrangian primal functions
- Lagrangian dual functions
- Karush-Kuhn-Tucker (KTT) conditions for constrained optimization
- Support points
- Maximal margin classifiers
- Slack variables
- Support vectors
- Support vector classifier
- Conversion of support vector classifier optimization problem to ridge penalized logistic regression
- Hinge loss functions and subgradients of hing loss

- Finding support vector classifier using stochastic gradient descent instead of quadratic programming
- Basis functions
- Kernels
- Support vector machines (SVMs)
- Receiver operating characteristic (ROC) curves

Lecture 10: Neural networks
- Limitations of linear classifiers
- Non-linear transformations
- Introduction to feedforward neural networks
- Effects of numbers of hidden units and layers
- Effects of symmetries
- Different activation functions
- Cost functions for regression and classification problems
- Backward propagation to learn parameters in single and multilayer networks
- Regularization with weight decay
- Backward propagation with weight decay
- Early stopping to prevent overfitting
- Effect of network architecture

Lecture 11: Random forests and boosting
- Decision (regression and classification) trees
- Cost function for regression trees
- Recursive binary splitting
- Cost complexity pruning
- Gini and cross-entropy indices
- Node purity for classification trees
- Bootstrap
- Bagging (bootstrap aggregation)
- Out-of-bag error estimates
- Random forests
- Boosting (specifically for regression trees)

Lecture 12: Unsupervised learning
- Dimensionality reduction
- Review principal components analysis (PCA) from earlier
- Prototype clustering in general
- Hard prototype clustering with K-means clustering
- Derivation of within-cluster variation using centroids
- Choosing the number of clusters K
- Gap statistics
- Soft prototype clustering with Gaussian mixture models
- Hierarchical clustering

Final lectures (short 15-minute student presentations and literature review papers), with list from Fall 2019 as:
- Autoencoders
- Generative adversarial networks (GANs)
- Markov chain Monte Carlo (MCMC)
- Convolutional neural networks (CNNs)
- Hidden Markov models (HMMs)
- Anomaly detection
- Basic recurrent neural networks (RNNs)
- Long short term memory (LSTM) neural networks
- Image-to-image processing
- Self-organizing maps

u
<CAP5625_Fa2019_Syllabus.pdf>