

**STA 552 Applied Statistical Machine Learning
Spring Semester**

Professor: _____
Email: _____@wcupa.edu
Office Hours: Day 1
 Day 2
 Day 3

Phone: 610-436-_____
Office: Building Room _____

Prerequisites: STA 503 and 506

Required Materials: None

Course Description: This course introduces commonly used models and algorithms in data science fields. Both supervised and unsupervised machine learning algorithms will be discussed. Specific topics will be selected from supervised learning (probabilistic and linear classification, neural networks, tree-based models), unsupervised learning (clustering and feature extraction) and semi-supervised learning algorithms. This course will introduce both theories and applications.

Applicable Programmatic Student Learning Outcomes:

1. Demonstrate an understanding of probability and statistical inference, including the fundamental laws of classical probability, discrete and continuous random variables, expectation theory, maximum likelihood methods, hypothesis testing, power, and bivariate and multivariate distribution theory.
2. Demonstrated the ability to apply the elementary methods of statistical analysis, namely those based on classical linear models, categorical methods, and non-parametric ideas to perform data analysis for the purposes of statistical inference.
3. Demonstrate proficiency in the effective use of computers for research data management and for analysis of data with standard statistical software packages, particularly SAS.
4. Learn to develop and critically assess design of experimental studies and the collection of data.
5. **Apply one or more methods of statistical inference to a particular area of interest, particularly the program in the elective concentration.**
6. Gain practical experience in statistical consulting and communicating with non- statisticians, culminating with interaction with research workers at a local company as part of the internship practicum.

Course Student Learning Outcomes:

Students will be able to: **(All related to PSLO 5)**

1. Identify relevant real-world problems as instances of canonical machine learning problems.
2. Create features using both heuristic and model-based approaches.
3. Design and implement an effective solution to a prediction or classification problem.
4. Design and implement basic clustering and dimensionality reduction algorithms.
5. Compare and contrast evaluation methods for various predictive tasks with appropriate KPI metrics.

6. Develop and implement effective strategies for preprocessing data representations, partitioning data into training and holdout sets, and tuning hyperparameters.

Meeting & Assessing Student Learning Outcomes:

The course learning outcomes will be evaluated in the following components

- (1). Topic-specific practice assignments (60%)
- (2). Class participation (10%)
- (3). Comprehensive team project (30%)

Attendance Policy: Attendance in class is expected. Attendance will be recorded each class session. One unexcused absence is allowed with no penalty; every unexcused absence after the first will result in a deduction from the participation component of the course grade.

Tentative Course Outline:

Part I: Introduction and Overview (Weeks 1-2)

1. Course Overview
 - 1.1. Why take the introduction to statistical machine learning?
 - 1.2. Course goals
 - 1.3. Logistics
2. Introduction to Statistical Machine Learning
 - 2.1. Types of machine learning problems
 - 2.2. Traditional statistics and machine learning
 - 2.3. Statistical models as machine learning algorithms
 - 2.4. Evaluation of modeling KPIs
 - 2.5. Model building process
3. Featuring Engineering
 - 3.1. Introduction to feature engineering
 - 3.2. Feature transformation – focusing on model-based transformation such as WOE
 - 3.3. Discretization – ad hoc and model-based binning
 - 3.4. Feature selection -focusing on penalization (LASSO and Ridge) and cross-validation

Part II. Learning Algorithms Based on Labeled Data (Weeks 3-7)

Probabilistic and Statistical Models

1. Discrete Regression-based Models/Algorithms
 - 1.1. Discrete regression overview
 - 1.2. Non-linear models
 - 1.3. Kernel regression models
 - 1.4. Non-parametric smoothing regression (GAM)

2. Bayesian Methods
 - 2.1. The logic of Bayesian modeling/learning
 - 2.2. Bayesian prediction
 - 2.3. Bayesian networks (structural equation modeling)
3. Imbalanced Learning
 - 3.1. Description of the question
 - 3.2. Synthetic sampling approaches
 - 3.3. Statistical models under biased sampling plan
 - 3.4. Bias corrected and penalized logit models
4. Sequential Learning
 - 4.1. Problem description
 - 4.2. HMM
 - 4.3. PCI charting
 - 4.4. Machine learning approaches to time series forecasting

Traditional ML and Extensions

1. Tree-based Methods
 - 1.1. Overview of tree algorithms and averaging methods
 - 1.2. Boosting – GBM and XGBoost
 - 1.3. Learning imbalance data with tree-based algorithms
 - 1.4. Cost-sensitive trees for imbalanced problems
2. Neural Networks
 - 2.1. Perceptron
 - 2.2. Multi-layer NNs
 - 2.3. NNs with sigmoid activation vs logistics models
 - 2.4. Non-linear NNs: Introduction
3. Introduction to Support Vector Machines
 - 3.1. Linear SVM
 - 3.2. Nonlinear SVM
 - 3.3. Kernel SVM

Part III. Learning from Data with No Labels (Weeks 8-11)

1. Traditional Clustering Algorithms: An Overview
 - 1.1. K-means
 - 1.2. Hierarchical clustering
2. Mixture models
 - 2.1. Finite mixture model clustering

- 2.2. Bayesian mixture models
- 3. Latent Regression for Dimension and Scoring
 - 3.1. Linear and non-linear PCA and ICA
 - 3.2. Kernel PCA
- 4. Bayesian Methods: An Introduction
 - 4.1. Non-parametric Bayesian Clustering
 - 4.2. Bayesian hierarchical clustering
- 5. Traditional
 - 5.1. Linear and non-linear PCA and ICA
 - 5.2. Kernel PCA

Part IV. Statistical Machine Learning for Financial Risk Modeling (Weeks 12-14)

- 1. Credit risk modeling
- 2. Weight of evidence as a feature transformation tool
- 3. Predicting credit default probability
- 4. Fraud modeling and prediction
- 5. Machine learning models in financial supply chain optimization.

Evaluation & Grading:

A letter grade will be assigned based on performance in the course, according to the following scale:

Grade	Quality Points	Percentage Equivalents	Interpretation
A	4.00		Superior graduate attainment
A-	3.67		
B+	3.33		Satisfactory graduate attainment
B	3.00		
B-	2.67		
C+	2.33		Attainment below graduate expectations
C	2.00		
C-	1.67		
F	0	< 70%	Failure

D grades are not used. Refer to the Graduate Catalog for description of NG (No Grade), W, & other grades.

Statements Common to All WCU Graduate Syllabi:



ACADEMIC & PERSONAL INTEGRITY

It is the responsibility of each student to adhere to the university's standards for academic integrity. Violations of academic integrity include any act that violates the rights of another student in academic work, that involves misrepresentation of your own work, or that disrupts the instruction of the course. Other violations include (but are not limited to): cheating on assignments or examinations; plagiarizing, which means copying any part of another's work and/or using ideas of another and presenting them as one's own without giving proper credit to the source; selling, purchasing, or exchanging of term papers; falsifying of information; and using your own work from one class to fulfill the assignment for another class without significant modification. Proof of academic misconduct can result in the automatic failure and removal from this course. For questions regarding Academic Integrity, the No-Grade Policy, Sexual Harassment, or the Student Code of Conduct, students are encouraged to refer to the Department Graduate Handbook, the Graduate Catalog, the *Ram's Eye View*, and the University website at www.wcupa.edu.

STUDENTS WITH DISABILITIES

If you have a disability that requires accommodations under the Americans with Disabilities Act (ADA), please present your letter of accommodations and meet with me as soon as possible so that I can support your success in an informed manner. Accommodations cannot be granted retroactively. If you would like to know more about West Chester University's Services for Students with Disabilities (OSSD), please visit them at 223 Lawrence Center. The OSSD hours of Operation are Monday – Friday, 8:30 a.m. – 4:30 p.m. Their phone number is 610-436-2564, their fax number is 610-436-2600, their email address is ossd@wcupa.edu, and their website is at www.wcupa.edu/ussss/ossd.

REPORTING INCIDENTS OF SEXUAL VIOLENCE

West Chester University and its faculty are committed to assuring a safe and productive educational environment for all students. In order to meet this commitment and to comply with Title IX of the Education Amendments of 1972 and guidance from the Office for Civil Rights, the University requires faculty members to report incidents of sexual violence shared by students to the University's Title IX Coordinator, Ms. Lynn Klingensmith. The only exceptions to the faculty member's reporting obligation are when incidents of sexual violence are communicated by a student during a classroom discussion, in a writing assignment for a class, or as part of a University-approved research project. Faculty members are obligated to report sexual violence or any other abuse of a student who was, or is, a child (a person under 18 years of age) when the abuse allegedly occurred to the person designated in the University protection of minors policy. Information regarding the reporting of sexual violence and the resources that are available to victims of sexual violence is set forth at the webpage for the Office of Social Equity at <http://www.wcupa.edu/admin/social.equity/>.

EXCUSED ABSENCES POLICY

Students are advised to carefully read and comply with the excused absences policy, including absences for university-sanctioned events, contained in the WCU Graduate Catalog. In particular, please note that the "responsibility for meeting academic requirements rests with the student," that this policy does not excuse students from completing required academic work, and that professors can require a "fair alternative" to attendance on those days that students must be absent from class in order to participate in a University-Sanctioned Event.

EMERGENCY PREPAREDNESS

All students are encouraged to sign up for the University's free WCU ALERT service, which delivers official WCU emergency text messages directly to your cell phone. For more information, visit www.wcupa.edu/wcualert. To report an emergency, call the Department of Public Safety at 610-436-3311.

ELECTRONIC MAIL POLICY

It is expected that faculty, staff, and students activate and maintain regular access to University provided e-mail accounts. Official university communications, including those from your instructor, will be sent through your university e-mail account. You are responsible for accessing that mail to be sure to obtain official University communications. Failure to access will not exempt individuals from the responsibilities associated with this course.