CSCI DASC 6020: Machine Learning
Fall 2023 Course Syllabus
East Carolina University

Contents

1 Class meeting times and office hours

- Meeting days and time: MW, 12:00 - 13:15
- Instructor: V.N. Gudivada
- Office/Hours: MW: 15:15 - 16:30; additional office hours are available by appointment.

2 Course description

This course will empower students with a solid foundation in classical machine learning techniques. It offers a comprehensive exploration of foundational concepts, algorithms, and techniques that underpin machine learning. Students will gain a deep understanding of the mathematical and statistical principles behind various machine learning methods, enabling them to design, implement, and evaluate machine learning models for various applications. Through theoretical lectures, hands-on programming assignments, and practical projects, students will develop the skills to address real-world challenges using the classical machine learning approaches.

3 Prerequisites

Graduate student status in Computer Science, Data Science, or Software Engineering. Instructor permission is needed for other majors. Background in multivariate calculus and linear algebra is beneficial.

4 Course learning outcomes

By the end of the course, students will be able to:

1. **Demonstrate Knowledge**: Understand the fundamental concepts, theories, and terminologies in classical machine learning.
2. **Apply Algorithms**: Implement and analyze a variety of machine learning algorithms to solve diverse problems.
3. **Evaluate Models**: Develop proficiency in evaluating model performance, selecting appropriate evaluation metrics, and interpreting results.
4. **Optimize Models**: Apply techniques for hyperparameter tuning and model selection to improve the performance of machine learning models.
5. **Mathematical Foundations**: Grasp the mathematical foundations of machine learning, including linear algebra, probability, and optimization.

6. **Feature Engineering**: Explore techniques for feature selection, extraction, and transformation to enhance model effectiveness.
7. **Problem Solving**: Adapt and customize machine learning algorithms to address specific problem domains and data types.
8. **Ethical and Social Considerations**: Recognize ethical and societal implications related to data collection, model bias, and fairness in machine learning.

5 Course topics

1. **Introduction to Machine Learning**: definition of machine learning; types of machine learning: supervised, unsupervised, and reinforcement learning; applications of machine learning in various domains.
2. **Mathematical Foundations**: linear algebra for machine learning: matrices, vectors, eigenvectors, and eigenvalues; probability and statistics: distributions, expectation, variance, and Bayes' theorem.
3. **Data Preprocessing**: data cleaning and handling missing values; feature scaling and normalization; one-hot encoding and categorical variable handling.
4. **Supervised Learning Algorithms**: Linear regression and regularization techniques; logistic regression and classification; decision trees and ensemble methods: random forests, gradient boosting.
5. **Unsupervised Learning Algorithms**: clustering techniques: k-means, hierarchical clustering; dimensionality reduction: Principal Component Analysis (PCA), t-SNE.
6. **Model Evaluation and Validation**: train-test split, cross-validation, and overfitting; evaluation metrics: accuracy, precision, recall, F1-score, ROC curves.
7. **Model Selection and Hyperparameter Tuning**: grid search and random search; bias-variance tradeoff and regularization.
8. **Feature Engineering and Transformation**: feature selection methods; feature extraction techniques: PCA, LDA.
9. **Ethical and Social Considerations**: bias and fairness in machine learning; privacy and security concerns in data handling.
10. **Case Studies and Practical Projects**: applying machine learning to real-world datasets; designing and implementing machine learning pipelines; addressing domain-specific challenges through machine learning.

6 Reference books

1. John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies, Second edition, The MIT Press, 2020. ISBN: 978-0262044691. Book website, where you can find sample chapters, solutions to select exercises, code and data, textbook image set, and slide set (PDF and LaTeX formats).

2. Kevin Murphy. Probabilistic Machine Learning: An Introduction, The MIT Press, 2022, 978-0262046824. Book website, where you can find PDF copy of the book, code, figures, and solutions to non-starred exercises.

3. Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 3rd Edition, O'Reilly Media, 2022. ISBN: 978-1098125974. Find book code here.

4. Jake VanderPlas. Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media, 2016. ISBN: 978-1491912058. Find book code here.

5. Hossein Pishro-Nik (2014), Introduction to Probability, Statistics, and Random Processes, Kappa Research, LLC, ISBN: 978-0990637202. Access the online version of the book here.

7 Web resources

1. University of Amsterdam Deep Learning Notebooks.

2. Rafael A. Irizarry. Introduction to Data Science: Data Analysis and Prediction Algorithms with R. Code for the examples in the book is here.

3. coursera, https://www.coursera.org/.

4. edX, https://www.edx.org/.

5. udemy, https://www.udemy.com/.

8 Course assessment and grading scale

Activity	Weight
Assignments	60
Midterm exam	20
Final exam	20

Score range	Letter grade
90.0 - 100	A
80.0 - 89.9	B
70.0 - 79.9	C
< 69.9	F

9 COVID-19 Considerations

COVID-19 continues to be a worldwide health crisis. You are encouraged to continue monitor COVID-19 symptoms. If symptoms occur, you are encouraged to get tested for COVID and remain home until symptoms improve. If you test positive for COVID-19, or are unvaccinated and exposed to COVID-19, you should isolate or quarantine according to CDC isolation and quarantine guidelines. If you live in the residence hall and test positive for COVID-19, you will need to either return home for the isolation period or “isolate in place” (leaving only for food pick-up and medical care while wearing a mask).

If you test positive for COVID-19 or were exposed to COVID-19, you are encouraged to complete the COVID self-report form in PiratePort. Completing this form will allow your teachers to receive a notification that you will not be able to attend classes due to illness or exposure.

If you are more comfortable continuing to wear a mask in indoor settings, then you are encouraged to do so. Students are expected to treat each other with respect as we make individual choices regarding precautions against COVID-19.

10 Respect for Diversity

It is my intent to serve well in this course all students from diverse backgrounds and perspectives. Students’ learning needs will be addressed both in and out of class. The diversity that students bring to this class be viewed as a resource, strength and benefit. I will strive to present the course content and learning activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other student groups.

11 Academic integrity policies

Students are expected to abide by the university’s Student Honor Code. The homework that you do is a critical part of your education. Each student is expected to do his or her own individual work. That does not mean you are not allowed to discuss your ideas with other students. Working together can be beneficial, and I encourage you to talk through ideas with other students. But outright copying is considered plagiarism and is unacceptable. Students who copy other students' work, or who allow their work to be copied, or who copy their work from other sources, such as the Internet, will receive either no credit or negative credit for the assignment, and will be reported to the university for an academic integrity violation.

Other potential academic integrity violations are cheating, falsification, multiple submissions of the same work in different classes, and attempts at any of these violations. Please see for details.

Academic integrity violations will result in a grade penalty up to and including an F for the course.

Bibliography

[1] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge, UK: Cambridge University Press, 2012. ISBN: 978-0521518147.

[2] Dimitri P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability*. 2nd ed. Cambridge, Massachusetts: Athena Scientific, 2008. ISBN: 978-1886529236.

[3] Christopher Bishop. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2007.

[4] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge, England: Cambridge University Press, Apr. 2020.

[5] F.M. Dekking et al. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Berlin, Germany: Springer, 2007. ISBN: 978-1852338961.

[6] Kaiser Fung. *Numbers rule your world: the hidden influence of probability and statistics on everything you do*. New York: McGraw-Hill, 2010. ISBN: 978-0071626538.

[7] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Third. O'Reilly Media, 2022. ISBN: 978-1098125974. URL: https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/.

[8] Saeed Ghahramani. *Fundamentals of Probability: With Stochastic Processes*. 4th ed. Boca Raton, Florida: Chapman & Hall/CRC, 2018. ISBN: 978-1498755092.

[9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Adaptive Computation and Machine Learning series. Cambridge, Massachusetts: The MIT Press, 2016. ISBN: 978-0262035613.

[10] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. Cambridge, Massachusetts: The MIT Press, 2016. ISBN: 978-0262035613.

[11] John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. Cambridge, Massachusetts: The MIT Press, 2020. ISBN: 978-0262044691.

[12] Mykel J. Kochenderfer. *Decision Making Under Uncertainty: Theory and Application*. Cambridge, Massachusetts: The MIT Press, 2015. ISBN: 978-0262029254.

[13] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Second. Cambridge, UK: Cambridge University Press, 2017. ISBN: 978-1107154889.

[14] Leonard Mlodinow. *The Drunkard's walk: how randomness rules our lives*. New York: Vintage, 2009. ISBN: 978-0307275172.

[15] Kevin Murphy. *Machine Learning: A Probabilistic Perspective*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2012. ISBN: 978-0262018029.

[16] Kevin Murphy. *Probabilistic Machine Learning: An Introduction*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts: The MIT Press, 2022. ISBN: 978-0262046824.

[17] Avi Pfeffer. *Practical Probabilistic Programming*. Shelter Island, NY: Manning Publications, 2016. ISBN: 978-1617292330.

[18] Hossein Pishro-Nik. *Introduction to Probability, Statistics, and Random Processes*. Kappa Research, 2014. ISBN: 978-0990637202.

[19] Nate Silver. *The signal and the noise: why so many predictions fail–but some don't*. New York: Penguin Press, 2012. ISBN: 978-0143125082.

[20] Jake VanderPlas. *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media, 2016. ISBN: 978-1491912058.