**COURSE OUTLINE**

AI ML 1

Term –3.3

PGPXP 2020-21

|  |  |
| --- | --- |
| Faculty Name | Dr. Hemachandran K |
| Email ID |  |

**Brief Description and Relevance of the Course**

Based on fundamental knowledge of computer science principles and skills, probability and statistics theory, and the theory and application of linear algebra. This course provides a broad introduction to machine learning and statistical pattern recognition.

Topics include: (1) supervised learning: Bayesian decision theory, Bayes classifier, Minimum error-rate classification, Normal (Gaussian) density – Discriminant functions, Decision surfaces, Maximum-Likelihood estimation, Maximum a posteriori estimation; Gaussian mixture models -- Expectation-Maximization method for parameter estimation; Naive Bayes classifier, Non-parametric techniques for density estimation -- Parzen-window method, K-nearest neighbors method, Hidden Markov models (HMMs) for sequential pattern classification -- Discrete HMMs and Continuous Density HMMs, generative/discriminative learning, parametric/non- parametric learning, neural networks, and support vector machines; (2) unsupervised learning: clustering, dimensionality reduction, kernel methods; (3) learning theory: bias/variance tradeoffs; VC theory; large margins; and (4) reinforcement learning and adaptive control.

The course will also discuss recent applications of machine learning, such as to robotic control, data mining, autonomous navigation, bioinformatics, speech recognition, and text and web data processing.

*\*The course will require the students to develop lab-based projects*

**Intended Learning Outcomes**

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

1. develop an appreciation for what is involved in learning models from data.
2. understand a wide variety of learning algorithms.
3. understand how to evaluate models generated from data
4. apply the algorithms to a real-world problem, optimize the models learned and report on

the expected accuracy that can be achieved by applying the models

1. source and access data from a variety of databases
2. select and apply appropriate tools for data visualization
3. discover trends in analytical data stores using the data mining techniques of clustering,

segmentation, association, and decision trees.

**Recommended and Required Reference Textbooks**

S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)”

J. VanderPlas, “Python Data Science Handbook”

P. Mueller; L. Massaron, “Machine Learning for Dummies”

**Software Requirements for the Course**

Microsoft Office

MacOS 10.15 / Windows 10/ Linux VM

Python – Jupyter Notebook – Anaconda Navigator

**Session-Wise Topics and Reading/References**

**\*Note\* –**

**Readings**: Each session is preceded by a reading assignment. It is important to keep on top of the reading, which will be assumed during the lecture and discussion in class. You should set aside 2 hours to compete each reading. We do not expect you to fully understand everything before coming to class, but the goal is to prepare for class, familiarize yourself with new terminology and definitions, and to determine which part of the subject you want to hear more about. We encourage you to bring questions to class about material that is confusing. Other students might share your confusion.

In addition, you will be provided with handouts by the designated lecturer on a subject related to the session. You are expected to read the paper before class and collate your queries before arriving the lecture hall. The query should be thought provoking about the assigned paper.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sn.** | **Topic** | **Intended Learning Outcomes** | **Reading/Reference** |
| 1 | Introduction: what is ML; Problems, data, and tools; Visualization | 1. Develop an appreciation for what is involved in learning models from data  The student will be able to understand what ML is. | Chapter 1  P. Mueller; L. Massaron, “Machine Learning for Dummies” |
| 2-3 | Linear regression; SSE; gradient descent; closed form; normal equations; features  Overfitting and complexity; training, validation and test data | 1. Understand a wide variety of learning algorithms.  2. Understand how to evaluate models generated from data  The student will be able to understand basic statistics. | Chapter 3 & 4  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)”  Chapter 2, 3  J. VanderPlas, “Python Data Science Handbook” |
| 4 | Classification problems; decision boundaries; nearest neighbor methods | 1. Understand how to evaluate models generated from data  2. Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models  The student will be able to understand KNN. | Chapter 2, 3  J. VanderPlas, “Python Data Science Handbook” |
| 5 | Probability and classification, Bayes optimal decisions, Naive Bayes and Gaussian class-conditional distribution | 1. Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models learned and report on the expected accuracy that can be achieved by applying the models  Source and access data from a variety of databases  The student will be able to understand Naïve Bayes. | Chapter 2, 3  J. VanderPlas, “Python Data Science Handbook” |
| 6 | Linear classifiers: Bayes' Rule and Naive Bayes Model | 1. Understand a wide variety of learning algorithms.  2. Understand how to evaluate models generated from data  3. Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models  4. Source and access data from a variety of databases  The student will be able to understand Gaussian and Naïve Bayes approach. | Chapter 2, 3  J. VanderPlas, “Python Data Science Handbook” |
| 7-8 | Logistic regression, online gradient descent, Neural Networks, Decision tree | 1. Develop an appreciation for what is involved in learning models from data  2. Understand a wide variety of learning algorithms  3. Discover trends in analytical data stores using the data mining techniques of clustering  The student will be able to understand ANN and Gradient Descent | Chapter 2  J. VanderPlas, “Python Data Science Handbook” |
| 9-10 | Ensemble methods: Bagging, random forests, boosting A more detailed discussion on Decision Tree and Boosting | 1. Select and apply appropriate tools for data visualization  2. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand ensemble methods and bagging. | Chapter 18, 19  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)”  Chapter 2  J. VanderPlas, “Python Data Science Handbook” |
| 11-12 | Unsupervised learning: clustering, k-means, hierarchical agglomeration, Markov Chain and Hidden Markov Model | 1. Select and apply appropriate tools for data visualization  2. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand unsupervised learning. | Chapter 2  J. VanderPlas, “Python Data Science Handbook”  Chapter 20  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)” |
| 13-14 | Advanced discussion on clustering and EM | 1. Select and apply appropriate tools for data visualization  2. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand clustering. | Chapter 2  J. VanderPlas, “Python Data Science Handbook” |
| 15 | Latent space methods; PCA. | 1. Select and apply appropriate tools for data visualization  2. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand PCA. | Chapter 2  J. VanderPlas, “Python Data Science Handbook”  Chapter 16 & 17  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)” |
| 16 | Text representations; naive Bayes and multinomial models; clustering and latent space models | 1. Select and apply appropriate tools for data visualization  2. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand multinomial models. | Chapter 2 , 4  J. VanderPlas, “Python Data Science Handbook”  Chapter 9 & 10  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)” |
| 17-18 | VC-dimension, structural risk minimization; margin methods and support vector machines (SVM) and Support Vector Regression (SVR) | 1. Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models  2. Source and access data from a variety of databases  3. Select and apply appropriate tools for data visualization  4. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand SVM and SVR. | Chapter 8 & 9  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)”  Chapter 2  J. VanderPlas, “Python Data Science Handbook” |
| 19-20 | Support vector machines and large-margin classifiers, time series, autoregressive models, gradient descent, feature engineering and hyper parameter tuning | 1. Apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models  2. Source and access data from a variety of databases  3. Select and apply appropriate tools for data visualization  4. Discover trends in analytical data stores using the data mining techniques of clustering,  segmentation, association, and decision trees.  The student will be able to understand autoregressive models and tuning. | Chapter 20  S. Russell and P. Norvig, “Artificial Intelligence: A Modern Approach (3rd edition)”  Chapter 5  J. VanderPlas, “Python Data Science Handbook” |

**Performance Evaluation Components for the Course**

|  |  |  |
| --- | --- | --- |
| **Session No.** | **Marks** | **Evaluation Form** |
| **Continuous Evaluation** | | |
| **2** | 5 | Quiz |
| **4** | 5 | Quiz |
| **6** | 5 | Quiz |
| **8** | 5 | Quiz |
| **10** | 5 | Quiz |
| **12** | 5 | Quiz |
| **14** | 10 | Assignment |
| **16** | 10 | Assignment |
| **18** | 10 | Assignment |
| **20** | 10 | Class Participation |
| **End Term Examination** | | |
| **After Course Completion** | 30 | Written Test |

**Assignment Structure**

All assignments are to be presented in class along with the submission of a written report in soft copy format.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name of Assignment** | **Date of Submission** | **Take-home**  **or in-class** | **Individual-work or Team-work** | **Other Instructions** |
| Choose from:  -Athletic and Sensing Devices  -Audio & Music  -Finance and Commerce | As per Woxsen standards | Take-home | Team-work | You may choose one of the categories available.  Additional instructions will be announced in class when/if necessary |
| Prediction of Wine/Spirits brands quality based on the chemical composition of the products. | As per Woxsen standards | Take-home | Individual work | You may choose to predict the quality of wine or spirit brands considering their chemical composition datasets.  Additional instructions will be announced in class when/if necessary |
| Disease prediction project: predict the next pandemic | As per Woxsen standards | Take-home | Individual work | You shall design a disease prediction system using deep learning.  Additional instructions will be announced in class when/if necessary |

Computational assignments will ask you to develop implementations of algorithms for search, game-solving, constraint satisfaction, knowledge representation and reasoning, and planning, to apply them to different real-world problems, and to analyze the performance. We expect that all code will run, be well-written and be commented appropriately; the course staff is always happy to help explain style and conventions. The written components ask you questions about the concepts and methods that you have learned and to reflect on the performance of your implementations.

Attendance & Punctuality

Learning is an interactive process. Woxsen students are admitted partly based on the experience they bring to the school and what they can add to class discussions. Therefore, attendance is an important aspect of studying here. Students are expected to be present in all the classes. Absence is only appropriate in cases of extreme personal illness, injury, or close family bereavement. Voluntary activities are never valid reasons for missing any class. The faculty, with the assistance of the Faculty Associate, shall keep track of students’ attendance and decide on the nature and extent of penalty for any absence from the class. Penalty may include reduction in grade or repetition of the course.

Late arrival is disruptive to the learning environment; students are expected to be in class before the scheduled commencement time. Students arriving for class after the scheduled commencement time should be turned away unless they have a valid reason to be permitted to attend.

Faculty should not consider attendance of sessions as a component of performance evaluation. The grading system at Woxsen accounts for this.

**External Websites Disclaimer**

Neither the instructor nor Woxsen School of Business is responsible for the content of external websites discussed in the classroom and/or linked to via online course materials, e-mail messages, message boards, or other means. Referred websites are for illustrative purposes only, and are neither warranted nor endorsed by the faculty or Woxsen School of Business. Web pages change frequently, as do ownership of domain names. While every effort is made to ensure proper referencing, it is possible that students may, on occasion, find materials to be objectionable for reasons beyond our control.

**Copyright**

The content provided by the faculty in the class is copy-righted. Students are instructed not to distribute or share content used during courses with external entities without the explicit written consent of the author and/or faculty.

**Student Code of Ethics**

Each student enrolled in this course accepts personal responsibility to uphold and defend academic integrity and to promote an atmosphere in which all individuals may flourish. The Students’ Code of Ethics strives to set a standard of honest behavior that reflects well on students and the school. All students enrolled in these courses are expected to follow the Students’ Code of Ethics, which they have been given at the time of enrolling for the program and pledged to adhere to. Unethical and unfair practices adopted by students may lead to penalties such as having to repeat the course or having the student’s enrollment cancelled.