

# **E E 461P DATA SCIENCE PRINCIPLES (to be updated for Spring 26)**

## **Course Syllabus**

**Instructor:** [Prof. Joydeep Ghosh](mailto:jghosh@utexas.edu), [jghosh@utexas.edu](mailto:jghosh@utexas.edu)

My office hours are Tu/Th 1:30-2:30pm, EER 6.812.

(in person, but fyi my personal zoom link is: <https://utexas.zoom.us/j/9899283077> )

**TAs:**

Office hours:

@ EER 0.814C

## **PREREQUISITES:**

This course is meant for junior/senior level ECE/CS/BME students. **You MUST have taken (with grade of C- or better)**

EE351K (or BME 335 or Math 362K): Probability and Statistics

AND M340L (or equivalent, e.g. M311 or M341): Matrices/Linear Algebra

AND ( EE312): Software Design/Data structures

**Graduate students are not allowed in this class. (Integrated BS/MS students from ECE are allowed).**

## **Course description**

In this course, we will study the fundamental principles of neural nets/deep learning as well as statistical machine learning that are behind data science algorithms and approaches. Understanding these principles is crucial to getting insights into the pros and cons of different predictive modeling techniques so that you can (i) make an informed decision on what approaches to consider when faced with real-life problems requiring predictive modeling, (ii) the basic concepts behind key machine learning techniques, and (ii) how to apply models properly on real datasets so to make valid conclusions. It will focus on basic concepts and not programming/coding, which covered in the complementary Data Science Lab (EE 461J) class.

---

## **Grading information**

|            |                                                              |
|------------|--------------------------------------------------------------|
| 15%        | 6 Assignments                                                |
| 15%        | 6 quizzes. Best 5 scores counted. See canvas for quiz dates. |
| 5+5%       | Attendance + participation in class                          |
| 20+20+20%: | 3 in-class tests (See canvas for test dates.)                |

At the end of the course, you will get a score out of 100 based on the percentages stated above. Your final grade will be solely based on this score. **The grade is primarily based on the curve, i.e. is relative to how the whole class performs;** however entire curve may shift up or down depending on how the class as a whole performs relative to past classes. **Grading is NOT based on absolute thresholds, e.g. 90+ = A etc.**

**Late Submission policy for Assignments:** Due to the class size, exceptions are not made for late submissions (unless you get written permission prior to the deadline). So you will not get any credit for a late submission, and it is better to upload whatever you have done before the deadline.

## **Textbook**

The material for the lectures is taken from a wide variety of sources, my slides will be available via Canvas. I'll also be reading other readings/blogs/videos/.. on Canvas as needed.

The **main textbook** is:

**(BB)** *Christopher M. Bishop, Hugh Bishop*, “Deep Learning: Foundations and Concepts”, Springer. 2024. See <https://www.bishopbook.com/>

Supplementary references are provided in the introductory slides.

The Scikit-Learn websites is also reasonably documented.

## **Course Schedule and Topics.**

Recommended readings for each topic are included in the class notes/slides.

**1. Introduction:** Data Science Process, Types of predictive analytics, Brief History;

*Objective:* Provide overview and context for this class.

**2. Multivariate regression:** Linear regression, bias-variance tradeoff and overfitting; Model tuning; Basis function expansion; Dealing with a large number of features; Ridge Lasso.

*Objective:* Learn to design, understand and implement predictive models where desired outcome is a numeric quantity.

**3. Neural Networks for Regression:** (stochastic) gradient descent, non-linear regression, neural networks and MLP; intro to **deep learning**.

*Objective:* Understand SGD and the power of multilayered non-linear models.

**4. Data Pre-Processing (Brief):** Transformations, Imputations, Outlier detection, dimensionality reduction, PCA and tSNE

*Objective:* Understand that good data quality is a prerequisite for effective models, and study some methods for improving data quality. Also consider ways of representing data in lower dimensions for more effective modeling and visualization.

**5. Classification Theory:** Introduction, Decision Theory; Classifier Performance Metrics and

**6. Learning to Classify:** Leading Direct and Probabilistic (Discriminative and Generative) Approaches to Designing Classifiers.

*Objective:* Learn to design, evaluate and implement predictive models where desired outcome is a class label. Understand design approaches and trade-offs. Specific models considered include Logistic regression, Neural Networks and Decision Trees.

**7. Ensemble Methods:** Model Averaging, Bagging and Random forests, boosting, Gradient boosting.

*Objective:* Understand the benefits of combining multiple predictive models.

## **8. Deep Learning: Why and How:**

*Objectives:* Understand basics of deep learning and associated training issues.

## **9. Convolutional Neural Networks (CNNs):**

Common ML tasks in computer vision; convolution operation, pooling; transfer learning and Keras.

*Objective:* Learn to design, implement, and interpret convolution-based feedforward deep nets that are suitable for data with spatial/temporal locality, such as images, video, and audio.

## **10. Transformers:**

Attention Mechanism; Basic roles of transformers; applications for text; multimodal settings.

*Objective:* Understand the key notions underlying a broadly applicable technology currently powering large language models. Appreciate deployment options and tradeoffs.

---

## **Policy on Using Large Language Models and other internet resources:**

**Homeworks:** For any homework question, unless instructed otherwise, you are welcome to use any resources you can access (other than blindly copying a solution you found on the internet or elsewhere). However, you are responsible for understanding anything that is submitted with your name on it, and may be asked to explain such answers. Moreover, some homework problems may explicitly ask you not to use an LLM, provide history of prompts, to compare with your own approach etc, in which case such instructions should be followed.

**Quizzes/Test: These are strictly closed book, closed notes, and hence also closed to digital assistants.**

For the tests, you will be allowed a cheat sheet as follows: standard 8.5x11" paper, HANDwritten on one side only, and in your own handwriting, with no less than 12 pt times roman font and at least single spacing. This sheet should include your name and UTEID, and be turned in along with your test.

**ACADEMIC DISHONESTY AND POLICIES ON CHEATING:** Faculty at UT are committed to detecting and responding to all instances of scholastic dishonesty and will pursue cases of scholastic dishonesty in accordance with university policy. Scholastic dishonesty, in all its forms, is a blight on our entire academic community. All parties in our community -- faculty, staff, and students -- are responsible for creating an environment that educates outstanding engineers, and this goal entails excellence in technical skills, self-giving citizenry, and ethical integrity. Industry wants engineers who are competent and fully trustworthy, and both qualities must be developed day by day throughout an entire lifetime. Scholastic dishonesty includes, but is not limited to, cheating, plagiarism, collusion, falsifying academic records, or any act designed to give the student an unfair advantage. The fact that you are in this class as an engineering student is testament to your abilities. Penalties for scholastic dishonesty are severe and can include, but are not limited to, a written reprimand, a zero on the assignment/exam, re-taking the

exam in question, an F in the course, or expulsion from the University. Don't jeopardize your career by an act of scholastic dishonesty. Details about academic integrity and what constitutes scholastic dishonesty can be found at the website for the UT Dean of Students Office and the General Information Catalog, Section 11-802.

**Disabilities statement:** "The University of Texas at Austin provides upon request appropriate academic accommodations for qualified students with disabilities. For more information, contact the Office of the Dean of Students at 471-6259, 471-4641 TTY."

**NOTICES:**

- Students with disabilities may request appropriate academic accommodations from the Division of Diversity and Community Engagement, Services for Students with Disabilities, 471-6259, <http://www.utexas.edu/diversity/ddce/ssd/>
- A notice regarding academic dishonesty. UT Honor Code and example of what constitutes plagiarism : <http://registrar.utexas.edu/catalogs/gi09-10/ch01/index.html>
- A notice regarding accommodations for religious holidays. "By UT Austin policy, you must notify me of your pending absence at least fourteen days prior to the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.")

Additional important [\*Student Rights and Responsibilities\*](#)