NEW YORK INSTITUTE OF TECHNOLOGY

**College of Engineering and Computing Sciences**

**DTSC 620: Statistics for Data Science (Fall 2022)**

# DESCRIPTION

This course presents a range of methods in descriptive statistics, frequentist statistics, Bayesian statistics, hypothesis testing, and regression analysis. Topics includes point estimation, confidence interval estimation, nonparametric model estimation, parametric model estimation, Bayesian parametric models, Bayesian estimators, parametric testing, nonparametric testing, simple and multiple linear regression models, logistic regression model.

**Instructor:** Dr. Kiran Balagani, Off: HSH 226A; Tel: 516-686-1302

Email: [kbalagan@nyit.edu](mailto:kbalagan@nyit.edu)

**Teaching Assistant:** Mr. Pranadeep Potti (mpotti@nyit.edu)

**Class meeting:** Monday,5:45PM – 8:25PM

**Office hours:** TBA

**Textbook and Reference Books:**

* Probability and Statistics for Data Science: Math + R + Data (Chapman & Hall/CRC Data Science Series.
* Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python, 2nd Edition, Paperback
* Probability and Statistics for Computer Science 1st ed. 2018 Edition by David Forsyth

**Supplementary Texts and Material**

* R. O. Duda, P. E. Hart, and D. G. Stork, [*Pattern Classification*](http://www.amazon.com/exec/obidos/ASIN/0471056693) (2nd Edition)
* Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).
* Papers and reading material will be posted regularly on Blackboard.

**Upon successful completion of this course, students will be able to:**

1. Explore data with techniques of visualizing data, computing summary quantities, discovering irregularities and outliers in data.
2. Model the data generating process in probabilistic models and estimate the parameters of the model with point estimators and confidence interval estimators, the methods of moments, maximum likelihood estimator.
3. Estimate the underlying distribution of the data with nonparametric methods for empirical cumulative distribution, density estimation.
4. Evaluate estimating results using biasedness and mean square error.
5. Use prior and posterior distributions in Bayesian parametric models for minimum mean-square-error estimation and maximum-a-posterior estimation.
6. Conduct parametric hypothesis testing: setting up null and alternative hypothesis, finding appropriate test-statistics, calculating p-value, controlling two types of errors.
7. Conduct non-parametric hypothesis testing: permutation and bootstrap methods.
8. Fit linear regression models for real-valued outcome: least square estimation, overfitting issues.
9. Fit logistic regression models for binary outcome.

**Syllabus (and Tentative Schedule):**

|  |  |
| --- | --- |
| **Class** | **Topic(s)** |
| 1 | * Introduction * Course description and syllabus * Statistical Pattern Recognition * Introduction to Probability Theory |
| 2 | * Bayes Classification Rule Bayesian * Optimality of Bayes Classification Rule and Expected Loss Functions |
| 3 & 4 | * Classification and Regression Models * Logistic Regression, Simple Linear Regression, Multiple Linear Regression * Project Assignment I |
| 5 | * Descriptive Statistics: * Visualizing data, computing summary quantities, discovering irregularities and outliers in |
| 6 | * Estimation for Parametric Models |
| 7 | * **Midterm Exam** |
| 8 | * Non-parametric Estimation * Project Assignment 2 |
| 9 | * Bayesian Statistics |
| 10 | * Project Assignment 1 Presentations |
| 11-12 | * Hypothesis Testing |
| 13-14 | * Applications I & II |
| 15 | * Presentations |
| 16 | * **Final Exam** |

**Course Requirements:**

|  |  |
| --- | --- |
| Class Participation: | **Regular** attendance is required and class participation is expected. Try not to be late. I reserve the right to decrease your grade if you *miss three or more classes*. |
| Attendance Policy from the NYIT Catalog: | A student is expected to attend each class session on a **regular** and punctual basis in order to obtain the educational benefits, which each meeting affords.  Students shall be informed by their instructors exactly how often they will be allowed to be late or absent during the semester. Students who exceed these limits may be withdrawn from the course by the instructor. In the event of a student’s absence from a test, the instructor will generally determine whether the student will be allowed to make up the work that was missed. Lack of preparation is not an adequate excuse for missing an examination. |
| Plagiarism: | **Plagiarism** will result in an **F** grade on the exam, programming assignment, homework, project, or term paper. If it occurs more than once, the *course grade* will be F. Copying code or text or trivially rephrasing text from other sources (including WWW) for your projects/term papers is **plagiarism.** |
| Late Submission: | Late submissions will be accepted if they are submitted within one week of the due date. Late submissions will be awarded partial credit (75% on the day after the due date, 60% thereafter for the next six days). They will not be accepted after that. There are no exceptions for any reasons. |
| Reading: | Reading assignments should be completed prior to the first class of the week in which they are assigned. It is a good idea to take class notes. While I encourage discussion of assignments among the students, the actual solutions and programs should be done *individually* unless I have approve students working together (such as in group projects and group homework). |
| Homework: | Homework will be assigned. Homework solution will be reviewed in class. |
| Project: | Homework/projects not attempted will be graded as 0 and included in the final average. Students should be ready to demonstrate their projects and programming assignments when asked to do so. |
| Exams: | There will be two exams. **There are no makeup exams.** |
| Grading Criteria: | Midterm: 30%; Final: 30%; Pop-quizzes, homework, mini-presentation, and class participation: 10%; Projects, programming assignments, presentations, and term paper: 30%. |
| Incomplete:  I grade | A grade of incomplete, I, can be given by the instructor after consultation with the Department Chair. It is used when a student, because of some unavoidable circumstance, has been unable to complete all assigned work for the course. The instructor must certify that the student’s work is passing at this point and the student must agree to complete the missing the work. A grade of I will become an F in the following situation: I is given in the fall semester and not made up by the end of the following summer. I is given in the spring semester and not made up by the end of the following fall. |
| Withdrawal:  W grade | Students can withdraw up to the 8th week of the semester and receive a grade of W. After the 8th week deadline, a student may withdraw and receive a W only if the student is passing the course. Otherwise, a student withdrawing after the 8th week will receive a grade of WF. |
| Misc: | |  |  | | --- | --- | | Do not eat in class. Do not sleep in class. Cell phone buzzers off. |  | |