Instructor: Deokgun Park

- Class hours: Tue/Thu 12:30-1:50pm
- Office hours: Tue/Thu 2:00-2:30pm

To schedule a meeting, reserve using the slack message at the following channel:
https://datamining2019spring.slack.com/messages/CF1SG9G3Y/

- Office: ERB 533
- E-mail: deokgun.park.uta.edu
- Homepage: http://crystal.uta.edu/~park/

TA: Fatma Arslan

- Office hours: Tue 3:00 - 5:00pm
- Office: ERB 508
- Email: fatma.arslan@uta.edu

Course Description

This is an introductory course on data mining. Data Mining refers to the process of automatic discovery of patterns and knowledge from large data repositories, including databases, data warehouses, Web, document collections, and data streams. The major topic we study includes the following:

- The fundamentals of the text information retrieval
  - TF-IDF
  - Vector representation of words
  - Word Embedding
- Basics of the supervised learning
  - kNN
  - Naive Bayes
  - Support Vector Machines
  - Dimensionality reduction
  - Recommender system
Student Learning Outcomes:

A solid understanding of the basic concepts, principles, and techniques in data mining; an ability to analyze real-world applications, to model data mining problems, and to assess different solutions; an ability to design, implement, and evaluate data mining software. As a concrete outcome, each students will implement an app that can do following things:

- Search text information
- Build a classifier
- Recommend similar items

Textbook

Introduction to information Retrieval (IR)
Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze

Mining of Massive Datasets (MMDS)
Jure Leskovec, Anand Rajaraman, Jeff Ullman
Available freely at http://www.mmds.org/

Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Course Content</th>
<th>Reading &amp; Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 15</td>
<td>Course introduction. Boolean retrieval</td>
<td></td>
</tr>
<tr>
<td>Jan 17</td>
<td>Boolean retrieval</td>
<td>IR chapter 1</td>
</tr>
<tr>
<td></td>
<td>TF-IDF</td>
<td>Slides</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quiz 1</td>
</tr>
<tr>
<td>Jan 22</td>
<td>Homepage LinkedIn review</td>
<td>IR chapter 2,6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quiz 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Personal homepage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LinkedIn, Profile</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggested Questions</td>
</tr>
<tr>
<td>Jan 24</td>
<td>Vector space model</td>
<td>IR chapter 7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quiz 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggested Questions</td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
<td>Notes</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Jan 29 | Evaluation                   | IR chapter 8  
Project idea  
Quiz 4  
Suggested Questions |
| Jan 31 | Project idea review          | IR chapter 13  
Project idea presentation  
Quiz 5  
Suggested Questions |
| Feb 5  | Naive Bayes                  | IR chapter 14  
Quiz 6  
Suggested Questions |
| Feb 7  | Sketch review                | IR chapter 15  
Design sketch  
Quiz 7  
Suggested Questions  
Voting |
| Feb 12 | kNN                          | IR chapter 18  
Quiz 8  
Suggested Questions |
| Feb 14 | Dimension reduction          | MMDS chapter 9  
Quiz 9  
Suggested Questions |
| Feb 19 |                             | MMDS chapter 11  
Quiz 10  
Suggested Questions |
| Feb 21 |                             | Read cs 229 lecture note by Andre Ng  
Part V Support Vector Machines  
Quiz 11  
Suggested Questions |
| Feb 26 |                             | Read SVM tutorial blog  
Part1, Part2, Part3 |
| Feb 28 | Feature I review             | Read SVM tutorial blog  
Part4, Part5, Part6  
Develop Phase I  
Quiz 12  
Suggested Questions |
| Mar 5  | Contents based recommender system (RS) | Read Support Vector Machines Succinctly chap 3-6  
Quiz 13 |
<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Suggested Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 7</td>
<td>Collaborative filtering RS</td>
<td>[Read and follow](<a href="https://nlpforhackers.io/word-embeddings/Quiz">https://nlpforhackers.io/word-embeddings/Quiz</a> 14) [Suggested Questions Voting]</td>
</tr>
<tr>
<td>Mar 19</td>
<td>SVM I</td>
<td>[Read and follow](<a href="https://www.tensorflow.org/tutorials/representation/word2vec">https://www.tensorflow.org/tutorials/representation/word2vec</a> Quiz 15) [Suggested Questions]</td>
</tr>
<tr>
<td>Mar 21</td>
<td>Feature II review</td>
<td>Develop Phase II</td>
</tr>
<tr>
<td>Mar 26</td>
<td>SVM II</td>
<td></td>
</tr>
<tr>
<td>Mar 28</td>
<td>SVM III</td>
<td></td>
</tr>
<tr>
<td>Apr 2</td>
<td>SVM IV</td>
<td></td>
</tr>
<tr>
<td>Apr 4</td>
<td>Feature III review</td>
<td>Develop phase III</td>
</tr>
<tr>
<td>Apr 9</td>
<td>Word Embedding</td>
<td></td>
</tr>
<tr>
<td>Apr 11</td>
<td>Project brochure review</td>
<td>Project Brochure</td>
</tr>
<tr>
<td>Apr 16</td>
<td>Visual Analytics</td>
<td></td>
</tr>
<tr>
<td>Apr 18</td>
<td>Project video review</td>
<td>Project show video</td>
</tr>
<tr>
<td>Apr 23</td>
<td>Artificial General Intelligence</td>
<td></td>
</tr>
<tr>
<td>Apr 25</td>
<td>Project presentation review</td>
<td>Project presentation</td>
</tr>
<tr>
<td>Apr 30</td>
<td>Project presentation review</td>
<td>Project presentation</td>
</tr>
<tr>
<td>May 2</td>
<td>Final exam review</td>
<td></td>
</tr>
</tbody>
</table>
Expectations

I expect you to study **three hours per day** including weekends and holidays to survive this class. This will be a challenging course. I don't recommend students taking more than two challenging courses per semester. One common mistake that newly admitted ambitious master students make is taking too much challenging courses just because they would like to quickly get the marketable skills. It can actually ruin your grade and quality of life especially you just arrived USA from other places like India or China. Instead distribute challenging courses wisely. Another advice for the students is to start working on the project as soon as possible so that they can have enough time to study for the final.

We will do **flipped classroom model** in this class. In that, the students must read the reading material and watch the video before they come to the class. The class will be mostly question and answers. Many class hours will be devoted to the project progress report including project idea pitching and final presentation. The rationale is that it is the best use of the course time and the instructor resource, because it can provide **personalized care and feedback** to the individual students. By removing the burden of delivering general contents that students get everywhere including youtube and google nowadays, the instructor can have more resource to give personalized feedback to students. Remember that a personalized service is expensive while a mass produced product is cheap.

Grades

- **Term project** 50%
  - There will be a semester long project. It will have multiple parts that will be graded separately.
- **Final exam** 30%
  - (Time and place: TBA)
- **Quiz** 20%
  - There will be an online-quiz before every lecture. You are required to read the text, watch the lecture video and take quiz before you come to every class.
- **Extra credit** 20%
  - The students can get an extra credit by
    - helping other students in slack channel
    - solving extra credit problem in the quiz

The final letter grades will be based on students' performance. There is no pre-defined cutoffs or distribution of grades. Undergraduate and graduate students are compared in separate groups.
**Term Project**

The goal of term project is building a recommender webapp that you can show to someone with your phone or laptop that will help you look competent. I assume that you will work hard to get a job after graduation. If you want to go to academia and want to work on more academic project, please let me know. We can discuss different project topic. This is the most important part of your coursework. You might survive even if you fail the final exam. But there are no ways you can survive if you fail the term project. Luckily there are two aspects that can save you. First, there will be many small modules in the term project you can do well or not. Second coding the right solution is only half of the output. Your well-written report will be 50% of your grade for each progress. It means if you struggle the coding part, there is still a chance by writing a good report. Below are the modules for the term project

1. Preliminary (10%)
   a. Build your personal homepage
      i. You can use any format you want or reuse the homepage you already have. The only requirement is that you should have all the contents in your resume in your homepage. Put a download link to your resume. Use below template and mimic my homepage if you want.
      ii. [https://sourcethemes.com/academic/](https://sourcethemes.com/academic/)
      iii. [http://crystal.uta.edu/~park/](http://crystal.uta.edu/~park/)
   b. Create a Linkedin profile. You should have all the contents in your resume and the link to your homepage.
   c. Add your photo and one sentence to the directory
      i. [https://docs.google.com/document/d/1z37DneEyJhpjZvxhDswj7z_qGiaMhCo1WjGWwjlB-J8/edit?usp=sharing](https://docs.google.com/document/d/1z37DneEyJhpjZvxhDswj7z_qGiaMhCo1WjGWwjlB-J8/edit?usp=sharing)
   d. Submit
      i. Put a download link for your resume in this google doc. It should have a link to your homepage and Linkedin profile.
   e. Grading rubric
      i. Your resume is reasonably professional (2 point)
      ii. Your homepage has all the info in your resume (3 point)
      iii. Your homepage has a download link for your resume (1 point)
      iv. Your LinkedIn profile has all the info in your resume and (3 point)
      v. Your LinkedIn profile has a link to your homepage (1 point)
      vi. During the class, we will have a voting for the following award. The award winner will get 5 extra credits.
         1. Best homepage award
         2. Best Linkedin profile award

2. Project Idea (10%)
   a. Find the dataset you want to build a recommender system. I recommend that you select a topic you personally like. The reason is that it will be better if you
continue to work on upgrading the app after this course. You might try applying new technique from another course by using this work as a ground work for the project for that course. For example, let's say you built a beer recommender system because you like beer. Then when you are taking neural net course, you can add a new feature that can identify the beer by taking a photo and recommend a similar one. Usually one very good app is better than four mediocre ones. Below are some suggestions for the dataset. But you can always use something you want. Remember I don't recommend you to use a large dataset because it can make things unnecessarily complex for you. Usually less than 500 megabytes seems reasonable nowadays for 8GB laptop. Your dataset must have the user-items ratings matrix. It means that User A have rated item B as C. (A, B, C). Usually when you search with recommender system the data has this matrix.

i. Dataset requirements
   1. Should have a textual component for search
   2. Should have a multiple numerical attributes for classifier
   3. Big data is a bonus point but you can get full credit with small size data


iii. http://shuaizhang.tech/2017/03/15/Datasets-For-Recommend-System/


b. Write one page proposal for your project. It should include the link to the dataset you will use, the features it will support, list of the similar apps in the internet. Most importantly write a compelling "Why someone will need this app?" paragraph.

c. Sketch a three screenshot that shows the major feature of the app.

d. Create a project section in your homepage
   i. Put your why paragraph.
   ii. Link for downloading one page proposal
   iii. Three images of the sketches

e. Submit
   i. Submit the link for the project page in your homepage

f. Grading rubric
   i. Your why is compelling (5 point)
   ii. You found three existing similar apps for it (5 point)
   iii. Your one-page proposal looks reasonable (5 point)
   iv. You have three sketches for the features (5 point)

3. Project idea pitching (10%)
   a. Present your project idea to the class for 2 minutes
   b. Grading rubric
   i. I will grade your presentation myself during the class.
ii. We will do voting for the following award in the class. Award winners will get 5 extra points
   1. Best presentation award
   2. Best idea award
   3. Funniest idea award

4. Design sketch (10%)
   a. Build a design sketch and usage scenario
   b. Create a post in your homepage and send the link
   c. Present it to the class

5. Develop phase I (10%)
   a. Develop search features
   b. You can refer the code or tutorial internet. But main question you have to answer
      is what improvement you made over the existing reference.
   c. Documentation is the half of your work. Write a good blog post for your work and
      step-by-step how to guide for github readme.md
   d. Grading Criteria
      i. Developed search feature 1 point for web app or phone app
         1. no localhost allowed
      ii. Have a readme.md for your github code to explain step by step deployment instruction 1 point
      iii. Show TF-IDF calculation step 1 point
           1. show scores for query words and the search results. Show calculation step for ranking in the web app
           2. Highlight search query in the result
      iv. Have your reference 1 point
      v. Explicitly state what is your contribution over the reference 2 point
         1. For graduate student, Engineering contribution such as changing the version of python, adapting to different
            server platform are not accepted as the contribution. Accepted contributions are things such as implementing
            optimization idea in the text book.
      vi. Describe what was your challenge and how you solved it 1 point
      vii. Have some experiments and explain your finding 1 point
          1. Search results with or without stopwords, stemming, lemmartization
          2. Use of inverted matrix and caching to speed up. Measure the speed difference and plot bar chart
      viii. Have the blog article not attaching the file 1 point
      ix. Explain the basic algorithms 1 point

6. Develop phase II (10%)
   a. Develop classifier features
   b. You can refer the code or tutorial internet. But main question you have to answer
      is what improvement you made over the existing reference.
c. Documentation is the half of your work. Write a good blog post for your work and step-by-step how to guide for github readme.md

d. Grading Criteria
   i. Developed classification model for web app or phone app
      1. no localhost allowed
   ii. Have a readme.md for your github code to explain step by step deployment instruction 1 point
   iii. Show Classification calculation step in the web app 1 point
      1. For example, show probability scores for query and classes if you are using Naive Bayes.
   iv. Have your reference 1 point
   v. Explicitly state what is your contribution over the reference 2 point
      1. For graduate student, Engineering contribution such as changing the version of python, adapting to different server platform are not accepted as the contribution. Accepted contributions are things such as hyperparameter tuning or comparison of different models for performance.
   vi. Describe what was your challenge and how you solved it 1 point
   vii. Have the blog article not attaching the file 1 point
   viii. Explain the basic algorithms 1 point
   ix. Evaluation (2point)
      1. Have the evaluation result 1 point
      2. Have some graphs for tuning the hyperparameter and comparing the results 1 point

7. Develop phase III (10%)
   a. Recommender system features
   b. You can refer the code or tutorial internet. But main question you have to answer is what improvement you made over the existing reference.
   c. Documentation is the half of your work. Write a good blog post for your work and step-by-step how to guide for github readme.md
   d. Grading Criteria
      i. Developed recommender model for web app or phone app
         1. no localhost allowed
      ii. Have a readme.md for your github code to explain step by step deployment instruction 1 point
      iii. Show calculation step in the web app 1 point
      iv. Have your reference 1 point
      v. Explicitly state what is your contribution over the reference 2 point
         1. For graduate student, Engineering contribution such as changing the version of python, adapting to different server platform are not accepted as the contribution. Accepted contributions are things such as hyperparameter tuning or comparison of different models for performance.
      vi. Describe what was your challenge and how you solved it 1 point
vii. Have some experiments that validate your recommendation 1 point
viii. Have the blog article not attaching the file 1 point
ix. Explain the basic algorithms 1 point

8. Project show video (10%)
   a. Build a 1 minute project video advertising your app and upload it to the youtube. Put a video link in this google doc.
   b. Grading rubric
      i. We will watch the video during the class together.
      ii. I will grade your video myself.
      iii. We will do voting for the following award. Award winners will get 5 extra points
         1. Most professional video award
         2. Most beautiful video award
         3. Funniest video award
      iv. Make sure the video is easily discoverable in the homepage project description

9. Final report (10%)
   a. You can reuse the previous writing
   b. Make it like a brochure not the paper
   c. Include all the link for your output including github page

10. Final presentation (10%)

Academic Integrity

This is a graduate/senior level course. I take cheating very seriously. I will give 'F' to the FIRST cheating effort. Please note that some international students may not be familiar with how serious the plagiarism can harm your career. If you cannot follow the coursework, it is better to drop the course or to do your best and get 'B' than do cheating and get a 'F' or suspended for your professional career.

Some students got the D grade because they copied each other's work for the makeup assignment for an extra point. Some of them would have gotten 'B' if they did not submit the assignment at all. You are young and you might not know what is important and what is not yet. For example, GPA is not as important at graduate level. Your potential employer will not hire you because you have a perfect GPA. They will hire you when you demonstrate expert skills and when you are trustworthy. Your GPA will matter when you are going to get PhD. Still, if you get D in one course and get A in every other course, they will give you a chance. Here is my secret. My undergraduate GPA is 2.56 out of 4.3, which is ridiculously low. Still you can recover. Don’t think that you will ruin your life when you get a low grade.
However, let’s say that you cheat and don’t get a bad penalty today. Probably you will try similar behavior later. People do not change easily. Maybe you will not caught for long. But once you caught during the graduate school or during the professional career, you will get an irreversible damage and probably have to find another career in other field. The punishment will be irrationally severe than what you think you deserve. The rationale behind this is because the probability of getting caught is low and they would like to warn other people not to try such a behavior. Remember the expected outcome is the product of the probability of getting caught and the punishment you will get when you caught. Because the probability is low, the punishment should be higher to compensate. Otherwise people will cheat always because the benefit of cheating is higher than the expected penalty. By the way, medicine is the passphrase. You are still young to understand this fully, but as an educator I feel strongly obligated to teach this lesson before it is too late.

Similarly very common plagiarism mistake the students make is copying image from web without giving proper credit. In this case, because it is not strictly or intentionally copying, I gave the report part 0 point. It is still terrible mistake and you should not do that in professional report.

Profile

Even though I am quite bad remembering name, I am trying to get better. Please help me by uploading your name and photo and one sentence here. I put mine as an example. Profiles will be used for the main feedback area for the project.

https://docs.google.com/document/d/1z37DneEyJhpjZvxbDswwj7z_qGlaMhCo1WjGWwjB-J8/edit?usp=sharing

Slack channel

In this class, we will use slack as a main communication medium. If you send me an email, you get a penalty. It is for not reading the syllabus carefully. We’re always experimenting with how to structure our online discussions. I highly encourage you be part of the conversation: speak up with thoughts, links, ideas, updates, and anything that comes to mind. Most importantly, relax and enjoy chatting with others, no pressure.¹ You can join at https://datamining2019spring.slack.com/ or using the following link.

---

¹ Paragraph cited from MIT 6.S099: Artificial General Intelligence Syllabus. https://docs.google.com/document/d/1ZqgghxV1lpZeWuV5zNKOpMUBHyYTw9n6eYzzx918nok/edit#heading=h.4qyhup1515w
The current Slack channels are:

- **#data-mining**: Anything related to data mining and the course Data Mining. This includes code, blog posts, high-level thoughts, etc.
- **#announcements**: Announcements by Deokgun and TA
- **#introductions**: Introduce yourself and say hello to others.
- **#p-recommender**: Project channel for Semester Long Project. You can ask question here. You can earn extra credit for helping other people.
- **#random**: Discussion on anything and everything.

Resolving Grading issues

The course will be graded relatively. About 30% of the students will get A and 40% of the students will get B. The actual threshold or number of the grades will be depend on the instructor. If you have a question about the grade, please meet TA and resolve first. If it does not resolve the issue, then you can contact me. **If you contact me directly without consulting with TA first, you will get 3% automatic grade penalty for the assignment or exam.** It is because you didn’t read the syllabus first.

Frequently asked questions

1. I would like to get research experience in my lab. May I join your lab as a master student?

You can contact me and discuss your interest. However, if you want to get a job after graduation, doing research with me might not be helpful. You might get a better chance for the job by preparing the coding interview. The research takes time to produce a meaningful outcome. And the research is a full-time job. For example, it is common that it takes more than two years before a phd student can produce a paper or meaningful software. However, this can be too late for the master students who need something to show to potential employer during first semester and third semester. Doing research is more beneficial for those who consider phd program after graduation.
**Attendance:** At The University of Texas at Arlington, taking attendance is not required but attendance is a critical indicator in student success. Each faculty member is free to develop his or her own methods of evaluating students’ academic performance, which includes establishing course-specific policies on attendance. As the instructor of this section, I require all students to attend lectures. However, while UT Arlington does not require instructors to take attendance in their courses, the U.S. Department of Education requires that the University have a mechanism in place to mark when Federal Student Aid recipients “begin attendance in a course.” UT Arlington instructors will report when students begin attendance in a course as part of the final grading process. Specifically, when assigning a student a grade of F, faculty report the last date a student attended their class based on evidence such as a test, participation in a class project or presentation, or an engagement online via Blackboard. This date is reported to the Department of Education for federal financial aid recipients.

**Announcements:** Stay tuned and make sure to check Blackboard frequently. Important announcements will be posted there.

**Assignments and Deadlines**
- All the assignments must be submitted through Blackboard. **We will NOT take hardcopy or email submission, unless the university verifies that Blackboard was malfunctioning or unavailable.** If you are not able to submit through Blackboard due to its technical failure, you can email your assignment to us, together with a screenshot showing the technical failure. **We will verify with the university.**
- Everything is due by 11:59pm on the due date. The deadline is automatically managed by Blackboard. You can still turn in assignment after the deadline. However, you automatically lose 5 points per hour after the due time, till you get 0. (Each individual assignment is 100 points.) **We cannot waive the penalty, unless there was a case of illness or other substantial impediment beyond your control, with proof in documents.**
- Do not send email asking for an extension for the following reasons:
Regrading: Regrading request must be made within 7 days after we post scores on Blackboard. TA will handle regrade requests. If student is not satisfied with the regarding results, you get 7 days to request again. The instructor will regrade, and the decision is final.

Drop Policy: Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. By the way, “bazinga” is the word being asked in the quiz 1. Students will not be automatically dropped for non-attendance. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. For more information, contact the Office of Financial Aid and Scholarships (http://www.uta.edu/aaofao/).

Disability Accommodations: UT Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including The Americans with Disabilities Act (ADA), The Americans with Disabilities Amendments Act (ADAAA), and Section 504 of the Rehabilitation Act. All instructors at UT Arlington are required by law to provide “reasonable accommodations” to students with disabilities, so as not to discriminate on the basis of disability. Students are responsible for providing the instructor with official notification in the form of a letter certified by the Office for Students with Disabilities (OSD). Only those students who have officially documented a need for an accommodation will have their request honored. Students experiencing a range of conditions (Physical, Learning, Chronic Health, Mental Health, and Sensory) that may cause diminished academic performance or other barriers to learning may seek services and/or accommodations by contacting:

The Office for Students with Disabilities, (OSD) www.uta.edu/disability or calling 817-272-3364. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability.

Counseling and Psychological Services, (CAPS) www.uta.edu/caps/ or calling 817-272-3671 is also available to all students to help increase their
understanding of personal issues, address mental and behavioral health problems and make positive changes in their lives.

**Non-Discrimination Policy:** The University of Texas at Arlington does not discriminate on the basis of race, color, national origin, religion, age, gender, sexual orientation, disabilities, genetic information, and/or veteran status in its educational programs or activities it operates. For more information, visit uta.edu/eos.

**Title IX Policy:** The University of Texas at Arlington (“University”) is committed to maintaining a learning and working environment that is free from discrimination based on sex in accordance with Title IX of the Higher Education Amendments of 1972 (Title IX), which prohibits discrimination on the basis of sex in educational programs or activities; Title VII of the Civil Rights Act of 1964 (Title VII), which prohibits sex discrimination in employment; and the Campus Sexual Violence Elimination Act (SaVE Act). Sexual misconduct is a form of sex discrimination and will not be tolerated. For information regarding Title IX, visit www.uta.edu/titleIX or contact Ms. Jean Hood, Vice President and Title IX Coordinator at (817) 272-7091 or jmhood@uta.edu.

**Academic Integrity:** Students enrolled all UT Arlington courses are expected to adhere to the UT Arlington Honor Code:

> I pledge, on my honor, to uphold UT Arlington’s tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.

> I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code.

UT Arlington faculty members may employ the Honor Code in their courses by having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System Regents’ Rule 50101, §2.2, suspected violations of university’s standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student’s suspension or expulsion from the University. Additional information is available at https://www.uta.edu/conduct/.

**Electronic Communication:** UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and
events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at http://www.uta.edu/oit/cs/email/mavmail.php.

**Campus Carry:** Effective August 1, 2016, the Campus Carry law (Senate Bill 11) allows those licensed individuals to carry a concealed handgun in buildings on public university campuses, except in locations the University establishes as prohibited. Under the new law, openly carrying handguns is not allowed on college campuses. For more information, visit http://www.uta.edu/news/info/campus-carry/.

**Student Feedback Survey:** At the end of each term, students enrolled in face-to-face and online classes categorized as “lecture,” “seminar,” or “laboratory” are directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student’s feedback via the SFS database is aggregated with that of other students enrolled in the course. Students’ anonymity will be protected to the extent that the law allows. UT Arlington’s effort to solicit, gather, tabulate, and publish student feedback is required by state law and aggregate results are posted online. Data from SFS is also used for faculty and program evaluations. For more information, visit http://www.uta.edu/sfs.

**Final Review Week:** For semester-long courses, a period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week unless specified in the class syllabus. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.
**Emergency Exit Procedures:** Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exit. When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist individuals with disabilities.

**Student Support Services:** UT Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. Resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to resources@uta.edu, or view the information at [http://www.uta.edu/universitycollege/resources/index.php](http://www.uta.edu/universitycollege/resources/index.php).
Lecture Note

Vector Space Model

- Bag of words model
  - I eat a fish
  - Fish eat me.
- Knowledge Graph
  - I eat a red apple at the restaurant.
    - (I, eat, apple)
    - (apple, color, red)
    - (I, be at, restaurant)
- Grounded Language
  - Visual Q&A

Evaluation

- Accuracy
- Precision and Recall
- F-measure
  - F0.5 F2
- Mean average precision
- Train, Dev, Test set
- overfit
- underfit
- Limitation of the search evaluation
  - independence relevance
  - binary relevance
- SQL - Google - Visual Analytics
  - Known unknown
  - Unknown unknown
  - Can you detect anomaly using search?

Naive Bayes

- Frequentist vs Bayesian
The article describes a cancer testing scenario:

- 1% of women have breast cancer (and therefore 99% do not).
- 80% of mammograms detect breast cancer when it is there (and therefore 20% miss it).
- 9.6% of mammograms detect breast cancer when it’s not there (and therefore 90.4% correctly return a negative result).

Put in a table, the probabilities look like this:

<table>
<thead>
<tr>
<th></th>
<th>Cancer(1%)</th>
<th>No Cancer(99%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Pos</td>
<td>80%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Test Neg</td>
<td>20%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>

How do we read it?

- 1% of people have cancer
• If you already have cancer, you are in the first column. There’s an 80% chance you will test positive. There’s a 20% chance you will test negative.
• If you don’t have cancer, you are in the second column. There’s a 9.6% chance you will test positive, and a 90.4% chance you will test negative.
• Visualize it?
  ○ https://gatherplot.firebaseapp.com/#/
  ○
  ○

kNN
•
•