GGS 787
Scientific Data Mining for Geo-informatics

1. General information

Instructor:	Dr. Joon-Seok Kim
Where:	Exploratory Hall 2103
When:	Wednesday, 7:20-10pm. Aug 26 - Dec 18, 2019
Course website:	Blackboard
Credits:	3.0
Prerequisites:	Graduate Students: GGS 550 or permission of instructor
Office hours:	Tuesday and Thursday 5-6pm
Preferred contact method:	Blackboard discussion board (preferred) or email

2. Course Description

Both the current trends in technology such as smart phones, general mobile devices, stationary sensors and satellites as well as a new user mentality of utilizing this technology to voluntarily share information produce a huge flood of geo-spatial data. Using this flood of data is a tremendous market: A "$600 billion potential annual consumer surplus from using personal location data globally" is projected by McKinsey Global Institute. This course covers basic and advanced tools of data science to mine previously unknown and potentially useful knowledge from large sets of spatial and spatiotemporal data. The main focus of this course is a survey on the classical tasks of data-mining: Clustering, Outlier-Detection, Classification, Association-Rule-Mining, and Regression. For each of these data-mining tasks, this course will introduce basic solutions and algorithms. Emphasis is on domain-specific data mining algorithms suitable for spatial and spatiotemporal data with geoscience and geoinformatics applications, including Spatial Outlier-Detection, Frequent Co-location Mining and Spatial Regression. In addition to basic algorithms, a survey on advanced state-of-the-art data mining and machine learning solutions will be given. This survey includes Principal Component Analysis (PCA), latent factor analysis and tensor factorization. Furthermore, the availability of personal data and the mining of potentially sensitive information of individuals and their locations raise serious privacy concerns. An overview of the privacy threats in data publishing will be presented and a survey of the basic privacy guarantees and anonymization techniques will be provided.

3. Learning Outcomes

By the end of the course each student will

- have a broad knowledge-base on fundamentals, theory and techniques of data-mining algorithms,
- be able to articulate and effectively communicate concepts and ideas related to Spatial Data Management and Data Mining to experts, non-experts, and other professionals in a work environment,
- have the ability to appropriately apply the knowledge acquired in the course for various hypothetical and real-world data mining tasks, thus being able to mine new and useful information,
- have experience with free and open-source data mining frameworks, in order to apply data mining solutions to new data sets, to find new and interesting patterns,
- be able to properly interpret data mining results.

4. Format

The course will be taught as a combination of lectures and tutorials.

5. Textbooks

No required textbook. Course slides and reading material will be provided via Blackboard.

6. Technology Requirements

You will need access to a Windows or Macintosh computer with at least 2 GB of RAM and to a fast, reliable broadband Internet connection (e.g., cable, DSL).

For the amount of computer hard disk space required to take an online course, consider and allow for the space needed to: 1) install the required and recommended software and, 2) save your course assignments.

Software

This course will be using Python as a programming language.

Instead of Python, knowledge of any other high-level language is sufficient: It will be possible to do assignments in other high-level languages or even in Pseudo-Code. This course is not about programming skills, it's about understanding and applying concepts and algorithms.

7. Course outline (tentative)

In this course we will cover the following topics (please note that the topics and their order are subjected to change at the discretion of the instructor, any changes will be announced in class):

Week	Date	Topic	Assignment	
			Release	Due
1	8/28	Introduction: Data Mining Tasks Overview		
2	9/4	Basics: Data Types, Feature Spaces, Similarity Functions	1	
3	9/11	Classification: Introduction, Classifier Evaluation, Nearest-Neighbor Classifiers	2	1
4	9/18	Classification: Decision Trees, Bayesian Classifiers, Entropy	3	2
5	9/25	Regression	4	3
6	10/2	Clustering: Introduction, k-means		4
7	10/9	Midterm Exam		
8	10/16	Clustering: DBSCAN, Agglomerative Hierarchical Clustering	5	
9	10/23	Frequent Itemset Mining, Association Rule Mining	6	5
10	10/30	Latent Feature Analysis: Principal Component Analysis, Matrix Factorization, Tensor Factorization	7	6
11	11/6	NO CLASS		
12	11/13	Colocation Data Mining, Trajectory Mining	8	7
13	11/20	Data Privacy	9	8
14	11/27	Thanksgiving recess		
15	12/4	Project Presentations		9
16	12/11	Final exam		

8. Grades

Each assignment and written exam will be given a numerical grade on a 0-100 scale. Some assignments may include bonus tasks. At the end of the term all the marks will be totaled as a <u>weighted average</u> according to the following weights:

Intermediate Assignments	**20%**
Midterm Exam	**40%**
Final Exam	**40%**

Please note that in general all assignments will not have the same weight. The weight of each individual assignment will be indicated on the assignment form. Final grades at the end of the course will be assigned using **absolute achievements only**, independent of the relative standing in the class.

Students are given the option to complete a project, instead of taking the final exam. When a student chooses to write the project instead of the exam, then the final grade, expressed as a weighted average of the assignments, midterm and project, becomes as follows:

Intermediate Assignments	**20%**
Midterm Exam	**40%**
Project	**40%**

Grades will be based on the following cut of values.

A+	100%	B+	87-89%	C+	77-79%	D	60-69%
A	93-99%	B	83-86%	C	73-76%	F	<60%
A-	90-92%	B-	80-82%	C-	70-72%		

NOTE: Your final percentage will round to the nearest whole number, e.g., 89.2=89 but 89.6=90.

9. Exams

The course includes mandatory written midterm and final exams. The material covered in the exams will be announced in class. A student who cannot write a course examination or complete a course assignment because of an incapacitating illness, severe domestic affliction or other compelling reasons can apply for extension of time to complete an assignment.

10. Project

The project will include python implementations of most data mining tasks discussed in classroom, as well as an experimental evaluation, performed on real datasets. Each student will be provided with the datasets and will be assigned to implement a specific set of data mining tasks, by the instructor. The student is required to implement the tasks, test them on real data, visualize the results, and write a short final report describing the methods used, the insights and knowledge mined from the data. The project requires very good programming skills. The grade of the project will be based on the readability and performance of the python code, the visualization of results, and the written report. Students always retain the option to take the final exam, instead of the project.

11. Assignments

The course will include several written assignments on selected topics from the material covered in class and in the assigned reading. Assignments may include tasks such as analysis of data using Python, discussion/analysis of theoretical concepts and test cases of algorithms. All assignments are mandatory. Typically, two weeks will be allocated for every assignment.

Assignments should be done **through the Blackboard course website**.

Please note: Assignments should be submitted only through the Assignment submission section of the Blackboard system - DO NOT email assignments directly to the instructor.

12. Late papers submission

Papers submitted **after the due date will not be accepted**. Exceptions to this policy may be made given serious circumstances at the discretion of the Instructor.

Please note: Deferral of term work is a privilege and not a right; there is no guarantee that a deferral will be granted. Please make sure you notify the instructor as soon as you know a deferral is required.

13. General guidelines for ASSIGNMENT preparation and submission

a. Grades of assignments will be based on:
- **Academic merit** of your answers.
- **Conciseness** and **completeness** of your answers. Please write to the point and explicitly address the question or task. Avoid using unnecessary graphics (figures, tables, graphs etc.) unless they serve a specific purpose. Make sure to use captions and to refer to the graphics you include in your written answer. Graphics without any reference or accompanying explanation will be disregarded.
- **Organization** and **presentation**. Remember that your assignment report is a reflection of your thinking and learning process. Please organize your report in a logical fashion so that your answers could be easily identified. A general format for your presentation should, as a minimum, include the following components: (1) Question number, (2) Your written answer and/or description and discussion of your results, and (3) Visualization of your results, e.g. images, graphs, tables, as necessary.

b. Please remember that your assignment is a **professional document**, and should therefore be formatted and constructed accordingly. All assignments are to be typed. Hand-written assignments will not be accepted.

c. Submission of a hardcopy will be made in class; submission of a softcopy will be made through Blackboard.

d. The electronic submission of your assignment report has to be in **PDF format**.

e. If more than one file is submitted, you may submit a single **ZIP** file containing all the assignment files.

f. Each assignment submission should include a cover page with the following information: assignment title, assignment number, student name, and submission date.

g. Please make sure you have a backup of all the materials you submit.

14. Course website

The course has a Blackboard website. This website will provide you a single portal through which you may obtain lecture notes, retrieve assignment data and, review links to additional materials, and receive special announcements. You are required to visit the course website **once per day**. Please notify ITU (and, if necessary, the instructor) if you encounter any problems accessing this website.

15. Electronic communication

All course related email correspondence, including submission of assignments, should be made through the course Blackboard website. Please DO NOT send emails to the instructors' @gmu.edu address.

16. Student Expectations

- **Academic Integrity**: Students must be responsible for their own work, and students and faculty must take on the responsibility of dealing explicitly with violations. The tenet must be a foundation of our university culture. [See https://oai.gmu.edu/].
- **Honor Code**: Students must adhere to the guidelines of the George Mason University Honor Code [See https://oai.gmu.edu/mason-honor-code/full-honor-code-document/].
- **MasonLive/Email (GMU Email)**: Students are responsible for the content of university communications sent to their George Mason University email account and are required to activate their account and check it

regularly. All communication from the university, college, school, and program will be sent to students solely through their Mason email account.

- [See https://masonlivelogin.gmu.edu].
- **Patriot Pass**: Once you sign up for your Patriot Pass, your passwords will be synchronized, and you will use your Patriot Pass username and password to log in to the following systems: Blackboard, University Libraries, MasonLive, myMason, Patriot Web, Virtual Computing Lab, and WEMS. [See https://password.gmu.edu/index.jsp].
- **University Policies**: Students must follow the university policies.
- [See http://universitypolicy.gmu.edu]. Responsible Use of Computing - Students must follow the university policy for Responsible Use of Computing.
- [See https://universitypolicy.gmu.edu/policies/responsible-use-of-computing/].
- **University Calendar**: Details regarding the current Academic Calendar.
- [See https://registrar.gmu.edu/calendars/fall-2019/].
- **Students with Disabilities**: Students with disabilities who seek accommodations in a course must be registered with the George Mason University Office of Disability Services (ODS) and inform their instructor, in writing, at the beginning of the semester [See http://ods.gmu.edu].
- Students are expected to follow courteous Internet etiquette at all times; see http://www.albion.com/netiquette/corerules.html for more information regarding these expectations.

17. Student Services

- University Libraries: University Libraries provides resources for distance students.
- [See https://library.gmu.edu/for/online].
- **Writing Center**: The George Mason University Writing Center staff provides a variety of resources and services (e.g., tutoring, workshops, writing guides, handbooks) intended to support students as they work to construct and share knowledge through writing. [See http://writingcenter.gmu.edu]. You can now sign up for an Online Writing Lab (OWL) session just like you sign up for a face-to-face session in the Writing Center, which means YOU set the date and time of the appointment! Learn more about the Online Writing Lab (OWL).
- **Counseling and Psychological Services**: The George Mason University Counseling and Psychological Services (CAPS) staff consists of professional counseling and clinical psychologists, social workers, and counselors who offer a wide range of services (e.g., individual and group counseling, workshops and outreach programs) to enhance students' personal experience and academic performance [See http://caps.gmu.edu].
- **Family Educational Rights and Privacy Act (FERPA)**: The Family Educational Rights and Privacy Act of 1974 (FERPA), also known as the "Buckley Amendment," is a federal law that gives protection to student educational records and provides students with certain rights. [See http://registrar.gmu.edu/privacy].

Disclaimer: Any typographical errors in this Course Outline are subject to change and will be announced in class. The date of the final examination is set by the Registrar and takes precedence over the final examination date reported by the instructor.

Note: Recording is permitted only with the prior written consent of the professor or if recording is part of an approved accommodation plan.