Math 285 – Selected Topics in High Dimensional Data Modeling

Time and location: 9-10:15am T & TH, MacQuarrie Hall 320

Instructor: Dr. Guangliang Chen

Contact information: Office: MH 417 Phone: 924-5131 Email: guangliang.chen@sjsu.edu

Office hours: 10:15-11:15am (right after class), Tuesdays and Thursdays, and by appointment

Course introduction: Many real data sets have very high ambient dimensions but are intrinsically low-dimensional. This observation is often exploited to effectively model such high dimensional data. This course will survey the different models (linear, nonlinear and sparse) that have been developed during the last decade. The instructor plans to spend 2 to 3 weeks on each of the following topics in the specified order:

- *Singular Value Decomposition (SVD)*
- *Spectral Clustering*
- *Subspace Clustering*
- *Compressive Sensing (CS)*
- *Dictionary Learning (DL)*

Because this is an experimental course, the instructor reserves the right to change or add topics at any time during the semester.

Prerequisites: Math 164 and Math 129A. The course introduces cutting-edge machine learning research in the big data setting and will have a significant theory component, so a grade of B or better in each of the two courses is required.

Programming language: Familiarity of MATLAB is required. The course will have an extensive computing component, and students are expected to use MATLAB to implement many of the ideas and test them on various data sets.

Textbook: None, but reading material will be suggested/distributed from time to time in class.

Course learning outcomes: Upon successful completion of this course, students will be able to

- Understand the different models used for representing high dimensional data
- Know how to use the corresponding software packages
- Apply the models and techniques learned in this course to real data analysis tasks
- Develop a solid background in the foundations of data science

Requirements and grading: Course requirements will include 5 homework assignments and a final project that will involve open-ended data analysis tasks on real data. On the weekly basis,

you are expected to spend 6+ hours outside class time to prepare for lectures, work on homework and read the assigned material.

The homework assignments will typically have both a theoretical component and a programming component. Note that you may collaborate on homework but you must write independent solutions. Homework must be submitted on time in order to receive full credit.

As part of the final project, we will provide real data and a number of open-ended data analysis tasks on these data, as high-level objectives. The students will work towards the completion of these goals in the last few weeks of the semester and will report their findings in a report to be submitted on the last day of classes.

The weights in determining the semester average are:

- **Homework**: 60%
- **Final project** (including report and presentation): 40%

Letter grades will be computed from the semester average. Maximum lower bound cutoffs for A, B, C and D grades are 90%, 80%, 70%, and 60%, respectively. These bounds may be moved lower at the instructor's discretion.

Special accommodations: If you anticipate needing any special accommodation during the semester (for example you have a disability registered with SJSU's Accessible Education Center), please let me know as soon as possible.

Instructor feedback: I strive to teach in the best ways to facilitate your learning. To achieve this, it is very helpful for me to receive immediate feedback from you. You may send me an email, or leave an anonymous note in my mailbox. Please submit constructive criticism about things you would like me to change, as well as positive feedback about things in the class you are happy with.