**CS 43105 Data Mining Techniques**

**Homework 1**

**Instructor:** Xiang Lian

**Due Date:** Please refer to the course website

1. Based on your understanding, please explain what the data mining is. **[10 points]**

2. Please list 3 real applications for the data mining. Explain the data mining problems and solutions in these applications (give citations in reference format, if you use any online resources) **[30 points]**

3. Please discuss whether or not the following problems are data mining tasks. Explain why. **[30 points]**

3(a). Retrieve students' records from a relational table with grade = "A". [5 points]

3(b). From the table of students' information, check if attributes last name and address have any correlations. [5 points]

3(c). Find all the documents from the text database containing keywords "data mining". [5 points]

3(d). Divide the text database into several groups, each group containing near-duplicate or similar documents. [5 points]

3(e). Based on historical stock data, as well as other attributes (e.g., gold price, gas price, etc.) for the past few days, predict the trend of a stock tomorrow. [5 points]

3(f). Please provide your own example of the data mining. [5 points]

4. Please choose a specific data mining problem (e.g., classification, clustering, regression, etc.), and discuss the corresponding solutions to this data mining problem. You can search for these problems via Google from the Internet (e.g., Wikipedia, Web pages, research papers, etc.), and explain the problem definition and solutions. **[30 points]**

**Submission**

Submit an electronic copy of your homework solution to the [Blackboard](https://learn.kent.edu/).

18