Course title and number: ECEN 489/689 Special Topics in Data Mining and Analysis
Term: Spring 2016
Meeting times and location: ETB 1035 MW 03:00pm-04:15pm

Instructor Information
Name: Dr. Nicholas Duffield
Telephone number: 845-7328
Email address: duffieldng@tamu.edu
Office hours: MW 11:00am-12:00pm
Office location: WEB 332D

Course Description and Prerequisites
This course will provide a broad overview of data mining, integrating related concepts from machine learning and statistics. It will include fundamental topics such exploratory data analysis, pattern mining, clustering, and classification, but also more advanced topics such as kernel methods, high-dimensional data analysis, and complex graphs and networks.

Prerequisite: Previously undergraduate or graduate course in probability or statistics, or approval of instructor.

Learning Outcomes
Acquiring knowledge of foundations and application of methods in data mining and data analysis. The course will prepare students to use the methods and tools of data science in their own research, whether focused on methods themselves, or more on applications.

Grading Policies
Homeworks: 60%
Tests: 20%
Final Exam: 20%

Grading Scale: 90-100 A, 80-89 B, 70-79 C, 60-69 D, below 60 F.

Discussion of homework assignments is encouraged, but homework must be executed independently and copying is not allowed. Assignments must be typeset and handed in on time to receive full credit. No late homework and project proposals will be accepted unless an official document (e.g., doctor’s note) justifies the absence.
Textbook

Online through the TAMU Library: http://proquest.safaribooksonline.com.lib-ezproxy.tamu.edu:2048/9781107779105

Course Topics

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>DATA ANALYSIS FOUNDATIONS: Categorical Attributes, Graph Data, Kernel Methods, High-dimensional Data, Dimensionality Reduction</td>
</tr>
<tr>
<td>5-7</td>
<td>FREQUENT PATTERN MINING: Itemset Mining, Summarizing Itemsets, Sequence Mining, Graph Pattern Mining, Pattern and Rule Assessment</td>
</tr>
<tr>
<td>9-12</td>
<td>CLUSTERING: Representative-based Clustering, Hierarchical Clustering, Density-based Clustering, Spectral and Graph Clustering, Clustering Validation</td>
</tr>
<tr>
<td>12-14</td>
<td>CLASSIFICATION: Probabilistic Classification, Decision Tree Classifier, Linear Discriminant Analysis, Support Vector Machines, Classification Assessment</td>
</tr>
</tbody>
</table>

Americans with Disabilities Act (ADA)

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please contact Disability Services, in Cain Hall, Room B118, or call 845-1637. For additional information visit http://disability.tamu.edu

Academic Integrity

For additional information please visit: http://www.tamu.edu/aggiehonor

“An Aggie does not lie, cheat, or steal, or tolerate those who do.”