**E10.2011 Advanced Topics in Quantitative Methods:**

**Classification and Clustering**

**Marc Scott**

**Spring 2009**

**Lecture:** Tuesdays 3:30-6:10 pm; lab follows 6:20-7:35 pm

**Location:** 194 Mercer Room 304

**Office Hours:** Tuesdays 2:30-3:30 pm, and by appointment

**Office:** 318E Kimball Hall

**Phone:** 212-992-9407

**Text:** There is no required text. Selected chapters from several sources will be made available.

**Software:** SPSS version 15 or 16*. This course will use Blackboard.

---

**COURSE DESCRIPTION:**

**COURSE REQUIREMENTS:**

- **Participation:** 10%
  You are expected to attend class and participate in class discussions.

- **Homework problems:** 20%
  There will be several assigned problems intended to give you practical experience with the methods discussed.

- **Data Analysis Projects:** 70%
  There will be two data analysis projects (worth 35% each). One will be a cluster analysis and the other a problem in classification. Data may come from thesis research, a pilot study, or public datasets.

- **Lab section:**
  A lab section runs from 6:20-7:35pm immediately after class on Tuesdays. The lab provides hands-on guidance for all assignments and for data analysis projects.

---

**COURSE READINGS:**

Handouts will be available on Blackboard by the Monday preceding class. It is the student’s responsibility to print out and review the notes before coming to class.

**Late assignment policy:**

Assignments are to be handed in on time.

---

**SCHEDULE**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar 10</td>
<td>Introduction to classification and clustering; what is a cluster; visualization techniques, including principal components</td>
</tr>
<tr>
<td>17</td>
<td>Spring Break (no class)</td>
</tr>
<tr>
<td>24</td>
<td>Hierarchical clustering; linkage choices; distance measures; the dendogram.</td>
</tr>
<tr>
<td>31</td>
<td>Optimization techniques (k-means); choosing the number of groups; evaluating clusters; intro to finite mixture models</td>
</tr>
<tr>
<td>Apr 7</td>
<td>Model-based clustering (including model selection); Nagin clusters</td>
</tr>
<tr>
<td>14</td>
<td>Use of logistic regression for classification; PROJECT 1 DUE</td>
</tr>
<tr>
<td>21</td>
<td>Discriminant function analysis; adding complexity; interpretation</td>
</tr>
<tr>
<td>28</td>
<td>Use of dissimilarity matrices and the partition about medoids approach; multidimensional scaling</td>
</tr>
<tr>
<td>May 6</td>
<td>PROJECT 2 DUE</td>
</tr>
</tbody>
</table>

---

**Readings**

**March 10:** Everitt et al., Cluster Analysis (4th Ed.), chapters 1 & 2.


---

* A combination of SAS and/or R will be used for finite mixture models/Nagin Clustering; resources will be made available in lab to assist you in applying these methods with this software.

