Instructor: Wei Wang  
Spring 2011

Welcome!

- Instructor: Wei Wang
  - Homepage: [www.cs.unc.edu/~weiwang](http://www.cs.unc.edu/~weiwang)
  - Office: 316 Sitterson Hall
  - Email: weiwang@cs.unc.edu
  - Office hour: by appointment
Overview

• Homepage: www.cs.unc.edu/Courses/comp790-090-s11/
• Time: 11:00-12:15PM Tuesday and Thursday
• Place: SN011
• Credit:
  ▶ 3 for COMP 790-90
  ▶ 1 for BCB 713
• Prerequisite: none
  ▶ Preferred: Database, AI, Machine Learning, Statistics, Algorithms

Overview

• Grading scheme

<table>
<thead>
<tr>
<th></th>
<th>COMP 790-90</th>
<th>BCB 713</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Presentation</td>
<td>25%</td>
<td>N/A</td>
</tr>
<tr>
<td>and discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project</td>
<td>60%</td>
<td>80%</td>
</tr>
<tr>
<td>Attendance and</td>
<td>15%</td>
<td>20%</td>
</tr>
<tr>
<td>participation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

➢ No homework
➢ No exam
Overview

- **Textbook:** none
  - A collection of papers in recent conferences and journals

- **References**

---

Overview

- **Paper presentation (COMP 790-90 student only)**
  - One per student
  - Research paper(s)
    - List of recommendations (will be available)
    - Your own pick (upon approval)
  - Three parts
    - Review of research ideas in the paper
    - Debate
    - Questions and comments from audience
    - Class participation: One question/comment per student
  - Order of presentation: random
  - Please send in your choice of paper(s) by Feb. 3rd.
Overview

- **Project (due Apr 29th)**
  - One project: Individual
  - Some suggestion will be available shortly
    - You are welcome to propose your own
  - Checkpoints
    - Proposal: title and goal (due Feb 15th)
    - Survey of related work: pros and cons (due Feb 15th)
    - Outline of approach (due Feb 15th)
    - Implementation (due Apr 29th)
    - Evaluation (due Apr 29th)
    - Discussion and future directions (due Apr 29th)

Topics

- **Scope: Data Mining**
- **Topics:**
  - Association Rule
  - Sequential Patterns
  - Graph Mining
  - Clustering and Outlier Detection
  - Classification and Prediction
  - Regression
  - Pattern Interestingness
  - Dimensionality Reduction
  - ...
Topics

➢ Applications
  ❖ Bioinformatics
  ❖ Web mining
  ❖ Text mining
  ❖ Graphics
  ❖ Visualization
  ❖ Financial data analysis
  ❖ Security
  ❖ Software Engineering
  ❖ …

KDD References

• Data mining and KDD (SIGKDD: CDROM)
  ➢ Conferences: ACM-SIGKDD, IEEE-ICDM, SIAM-DM, PKDD, PAKDD, etc.
  ➢ Journal: Data Mining and Knowledge Discovery, KDD Explorations

• Database systems (SIGMOD: CD ROM)
  ➢ Conferences: ACM-SIGMOD, ACM-PODS, VLDB, IEEE-ICDE, EDBT, ICDT, DASFAA
  ➢ Journals: ACM-TODS, IEEE-TKDE, JIIS, J. ACM, etc.

• AI & Machine Learning
  ➢ Conferences: Machine learning (ICML), AAAI, IJCAI, COLT (Learning Theory), etc.
  ➢ Journals: Machine Learning, Artificial Intelligence, etc.
KDD References

• **Statistics**
  - Conferences: Joint Stat. Meeting, etc.
  - Journals: Annals of statistics, etc.

• **Bioinformatics**
  - Conferences: ISMB, RECOMB, PSB, CSB, BIBE, etc.
  - Journals: J. of Computational Biology, Bioinformatics, etc.

• **Visualization**
  - Conference proceedings: CHI, ACM-SIGGraph, etc.
  - Journals: IEEE Trans. visualization and computer graphics, etc.