

**COMP790-90 Research Seminar**  
**BCB 713 Module**

**Data Mining**

**Concepts, Algorithms, and Applications**

**Instructor: Wei Wang**

**Spring 2009**

*The UNIVERSITY of NORTH CAROLINA at CHAPEL HILL*



**Welcome!**

- **Instructor: Wei Wang**
  - **Homepage:** [www.cs.unc.edu/~weiwang](http://www.cs.unc.edu/~weiwang)
  - **Office:** 329 Sitterson Hall
  - **Email:** [weiwang@cs.unc.edu](mailto:weiwang@cs.unc.edu)
  - **Office hour:** by appointment
- **TA: Catie Welsh**
  - **Homepage:** [www.cs.unc.edu/~cwelsh](http://www.cs.unc.edu/~cwelsh)
  - **Office:** 323 Sitterson Hall
  - **Email:** [cwelsh@cs.unc.edu](mailto:cwelsh@cs.unc.edu)
  - **Office hour:** by appointment



## Overview

---

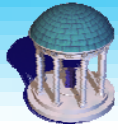
- **Homepage:**  
[www.cs.unc.edu/Courses/comp790-090-s09/](http://www.cs.unc.edu/Courses/comp790-090-s09/)
- **Time: 11:00-12:15PM Tuesday and Thursday**
- **Place: SN011**
- **Credit:**
  - 3 for COMP 790-90
  - 1 for BCB 713
- **Prerequisite: none**
  - Preferred: Database, AI, Machine Learning, Statistics, Algorithms



## Overview

---

- **Textbook: none**
  - A collection of papers in recent conferences and journals
- **References**
  - *Data Mining --- Concepts and techniques*, by Han and Kamber, Morgan Kaufmann, 2006. (ISBN:1-55860-901-6)
  - *Introduction to Data Mining*, by Tan, Steinbach, and Kumar, Addison Wesley, 2006. (ISBN:0-321-32136-7)
  - *Principles of Data Mining*, by Hand, Mannila, and Smyth, MIT Press, 2001. (ISBN:0-262-08290-X)
  - *The Elements of Statistical Learning --- Data Mining, Inference, and Prediction*, by Hastie, Tibshirani, and Friedman, Springer, 2001. (ISBN:0-387-95284-5)
  - *Mining the Web --- Discovering Knowledge from Hypertext Data*, by Chakrabarti, Morgan Kaufmann, 2003. (ISBN:1-55860-754-4)



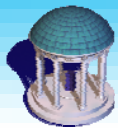
# Overview

---

- **Grading scheme**

	<b>COMP 790-90</b>	<b>BCB 713</b>
<b>Paper Presentation and discussion</b>	<b>25%</b>	<b>N/A</b>
<b>Project</b>	<b>60%</b>	<b>80%</b>
<b>Attendance and participation</b>	<b>15%</b>	<b>20%</b>

- **No homework**
- **No exam**



# 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. 3<sup>rd</sup>.**



## Overview

---

- **Project (due Apr 29<sup>th</sup>)**
  - **One project: Individual**
  - **Some suggestion will be available shortly**
    - ❖ **You are welcome to propose your own**
  - **Checkpoints**
    - ❖ **Proposal: title and goal (due Feb 10<sup>th</sup>)**
    - ❖ **Survey of related work: pros and cons (due Feb 10<sup>th</sup>)**
    - ❖ **Outline of approach (due Feb 10<sup>th</sup>)**
    - ❖ **Implementation (due Apr 29<sup>th</sup>)**
    - ❖ **Evaluation (due Apr 29<sup>th</sup>)**
    - ❖ **Discussion and future directions (due Apr 29<sup>th</sup>)**



## 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.