Dr. Christoph F. Eick

Review List

Midterm2 Exam Data Mining COSC 4335

Thursday, April 7, 2015 in 120K T

Last updated: April 4, 9p

The exam will be “open books and notes” and you can bring your labtop, but use of e-mail is strictly prohibited and will center on the following topics (at least 85% of the questions will focus on material that was covered in the lecture):

1. \*\*\* DBSCAN and density based-clustering (Class transparencies; textbook pages 526-532; <https://en.wikipedia.org/wiki/Density_estimation> )
2. \*\*\*\*\*Decision Trees, Regression Trees and General Topics for Classification (covered class transparencies and textbook pages 145-166; 168-180 (skip 4.4.3), 184-188, 296-297)
3. \*\*\*Nearest Neighbor (class transparencies and pages 223-227)
4. \*\*SVM (class transparencies, <http://en.wikipedia.org/wiki/Kernel_method> , and pages 256-274)

6. \*\*\*\*\*\*\*\*\* R-Programming (class and lab transparencies)

* Write R-functions
* Solve tasks that are similar than those in Assignments 1 and 2

You should have detailed knowledge concerning the following algorithms and approaches: Decision Tree induction algorithm, gain computations, kNN, DBSCAN, SVN hyperplane approach.

Midterm2 will count about 15-19% towards the overall course grade.