Dr. Christoph F. Eick

Review List Midterm2 Exam

COSC 4335: Data Mining

Thursday, April 5, 2017, 11:30a (in GAR 201)

Last updated: March 29, 10a

The exam will be “open books and notes” and you can bring your laptop, 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. \*\*\*\* Hierarchical Clustering, DBSCAN (class transparencies, textbook pages 515-532 )
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. \*\*\*\*SVM (class transparencies, <http://en.wikipedia.org/wiki/Kernel_method> , and pages 256-274)
4. \*\*\* Ensembles (class transpencies; text book pages 276-280, 285-290)
5. \*\*\* Neural networks (class transparencies, textbook pages 246-247, 251-256)

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

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

You should have detailed knowledge concerning the following algorithms and approaches: Decision Tree induction algorithm, gain computations, AGNES, DBSCAN, SVM hyperplane approach, Adaboost, architecture of neural networks and some basis understanding how neural networks learn models.

40-50% of the Midterm2 points are allocated to R-programmming .

Midterm2 will count about 15-18% towards the overall course grade.