Dr. Christoph F. Eick

Review List Midterm2 Exam

COSC 4335: Data Mining

Wednesday, November 7, 2:30p

Last updated: October 30, 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 (class transparencies, textbook pages 515-532 )[[1]](#footnote-1)~~
2. \*\*\*\*\*\*\*Decision Trees, and General Topics for Classification (covered class transparencies and textbook[[2]](#footnote-2) pages 117-156 (skip 3.3.5) and 162(starting with 3.5.4)-168
3. \*\*\*\*SVM (class transparencies, <http://en.wikipedia.org/wiki/Kernel_method> , and pages 276-296
4. ~~\*\*Ensembles (class transparencies; text book pages 296-313)~~1
5. \*\*\*\* Neural networks (class transparencies, textbook pages 249-262)
6. \*\*Nearest Neighbor Classifiers (class transparencies, textbook pages 208-212)
7. \*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 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, information and GINI gain computations, SVM hyperplane approach, kernels (only basic ideas), architecture of neural networks and some basis understanding how neural networks learn models; how nearest nearest neighbor classifier work.

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

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

1. Will be covered in the final exam, not Midterm2. [↑](#footnote-ref-1)
2. All page numbers refer to Second Edition of the Textbook [↑](#footnote-ref-2)