Dr. Christoph F. Eick

Review List Midterm2 Exam

COSC 3337: Data Mining

Wednesday, November 6, 2:30p

Last updated: October 28, 2019, noon

The exam will be “open books and notes” but the use of computers 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. \*\*\*Similarity Assessment class transparencies and 76-81; in general, this subject is not covered well in the textbook; therefore, focus on what is discussed in the transparencies when preparing for the exam.
2. \*\*\*\*\*\*Decision Trees, and General Topics for Classification, particularly overfitting (covered class transparencies and textbook[[1]](#footnote-1) pages 117-157 (skip 3.3.5) and 162(starting with 3.5.4)-169
3. \*\*\*\*\*SVM (class transparencies, <http://en.wikipedia.org/wiki/Kernel_method> , and pages text book 90-94 and 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. \*\* Naïve Bayes Classifier class transparencies and textbook pages 212-226

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 the nearest neighbor classifier and the Naïve Bayes classifier works.

Midterm2 will count about 15-16% towards the overall course grade.

1. All page numbers refer to Second Edition of the Textbook [↑](#footnote-ref-1)