CMPSCI 348 **Exam 2 Review Sheet** Spring 2019

Exam 2 will be given in class on Thursday, April 4. The questions below cover the concepts that will appear on the exam, although the exam questions will require you to use some concepts in combination. Questions on the exam will be drawn from both the lectures *and from the readings associated with each lecture*. Annotations after each topic refer to lecture dates (L1.24 is the lecture on January 24) and their associated readings.

Neural networks (L2.21)

- *Neural networks* — How is a neural network structured and what are its parameters? How does an individual artificial neuron work? What are its inputs and outputs? What are some examples of activation functions?
- *Learning neural networks* — What is backpropagation and how does it work? What is a convolutional layer, and why would you use it? What is a recurrent neural network, and why would you use it?

Classification and regression trees (L3.05, L3.07, L3.21)

- *Classification trees* — What are the components of a classification tree (CT)? What is the shape of possible decision boundaries for CTs?
- *Learning classification trees* — What is the basic method for learning the structure of a CT and estimating its parameters? What are typical hyper-parameters of a CT construction algorithm? What are the tradeoffs of CTs vs. other classifiers? What is cost-complexity pruning and how does it work?
- *Regression trees* — What is a regression tree and how does it differ from a CT?

Feature selection (L3.07)

- *Approaches* — What is best-subset selection and why don't we use it? What is forward selection? What is backward selection?

Overfitting, generalization, and regularization (L3.07, L3.19, L3.21)

- *Estimators in ML* — What are examples of estimators, estimates, and estimands within machine learning? How can estimates of the generalization error of models produced by ML algorithms be decomposed? How do different components of generalization error change as model capacity changes?
- *Model capacity* — What is model capacity? Why don't we always use the highest capacity model?
- *Hyper-parameters* — What are hyper-parameters? What are examples of hyper-parameters for different ML models?
- *Estimating generalization error* — What is generalization error? What is the relationship between the data set used to estimate generalization error and whether that estimate will be biased?

Multiple comparisons (L3.19)

- *Conditions* — What are the three necessary and sufficient conditions for a *multiple comparison procedure* (MCP)?
- *Results* — Is the score of the item selected by an MCP an unbiased estimate of its true score? Why? What can be done to obtain an unbiased estimate of the score of that item?

Bias and variance of learned models (L3.19)

- *Components and definitions* — What are three components of error for a learned statistical model? What are the definitions of those components, both mathematically and conceptually? What is the connection between alternative samples from the population and the bias-variance decomposition?

- *Conditions* — Under what conditions is the bias of a learned model virtually guaranteed to be large? Under what conditions is the variance of a learned model guaranteed to be large?
- *Tradeoff* — Why do we sometimes say that bias and variance can be "traded off" against each other? What is the typical behavior of bias, variance, and noise as model capacity changes?

Cross-validation (L3.19)

- *Operation* — How do you create cross-validation (CV) samples? What do you do with those samples to estimate model accuracy?
- *Uses* — Why is cross-validation useful? What are the alternatives to CV? Why would you use nested CV (with both an inner and outer loop)?

ROC Curves (L3.21)

- *ROC space and constructing curves* — What are the dimensions of ROC space? What do points in ROC space correspond to? What does an ROC curve correspond to? How do you construct an ROC curve for a classifier?
- *Interpreting ROC curves* — What does a diagonal line from (0,0) to (1,1) in ROC space correspond to? What does the curve of a perfect classifier look like in ROC space? What is a common summary measure used to compare ranking classifiers based on their ROC curve?

Ensembles (L3.26)

- *Basics of ensembles* — What is regression toward the mean? What is an ensemble? Why do ensembles improve generalization error? What are some ways that you can construct an ensemble?
- *Bagging* — What is bagging? What is a bootstrap sample? What are random forests, and how are the constructed?
- *Boosting* — What is boosting? How does boosting differ from bagging?
- *Stacking* — What is stacking? What are the key differences between stacking and other ensemble techniques (bagging and boosting)?