# MIDTERM, STUDY GUIDELINE

CIS/STA 3920  Oct 12, 2017

# Exam Structure

*calculator allowed*

- Exam time: 100 minutes
  - Written Exam: 70 min, 2:30----3:40 pm, closed book, room 10-125
  - Lab Exam: 30 min, 4:30---5:00 pm, open book, room 11-175

- Written Exam (85%)

  - 20% True/False, MC questions

  - 65% Short essay questions
    - Partial credit counts!

  - Questions can be both theoretical and applied ones

  - Applied questions may include calculations

  - Some calculation formulas you are supposed to know
    - Odds
    - Euclidian Distance

  - For True/False
    - A statement is true if it is ALWAYS true (not if it could be true or is usually true)

- Lab Exam (15%)

# Written Exam Structure

- You would <u>not</u> be asked to write R code during the written exam

- Also I would <u>not</u> ask questions directly about R code syntax, e.g. which function should we use to generate a histogram?

- However, you should be able to read, understand and interpret R outputs, similar to what we have done in practices or assignments

# Written Exam: Theoretical Questions

- **<u>Sample questions</u>** can be:
  - What is Bias-Variance Tradeoff?
  - What is the difference between Classification and Regression problem?
  - What is the difference between unsupervised learning and supervised learning?
  - What is overfitting?
  - How to apply different resampling methods?
  - How can bootstrapping be used to evaluate a model?
  - In linear regression, how can we perform stepwise/best subset selection?
  - When should we apply different classification methods such as logistic regression and KNN?

*in which situation*

# Written Exam: Applied Questions

- Given data, software outputs, tables, or plots, use your knowledge to solve problems, or interpret results

- **Sample questions** can be:
  - Comment on tables or plots
  - Given the box plot, describe your findings
  - Which model would you select given the outputs?
  - Read out and interpret model from the raw software output
  - How to perform cross validation?
  - How to evaluate the performance for regression or classification methods?
  - Select a better "K" for KNN based on the performance
  - Make a prediction given the model

# Lab Exam

- Use R to answer questions. Questions are similar to what we have in all labs and assignments.

- Based on lab policies, cell phone use is prohibited in lab spaces.

- Please arrive at lab at least 15 minutes before the exam so that you have time to set up software and network connection.

- Lab exam is open book.

- Make sure that you save all the codes related to your answers in one R code file. The file should be saved appropriately so that it could be re-opened.

- You need to submit both the exam paper and the R code file. The grading will be based on both R codes and your answers.