Quiz 2

**Stat. 4869/6620 Statistical Learning in R**  
**Department of Statistics and Biostatistics**  
**CSU East Bay**

May 23, 2018

**Name:**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Class: (4869 or 6620)**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Section: (3 or 1 or 2)**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Instructions:** This is a take-home quiz.

Type your answers in this .docx file and upload it in Blackboard before the end of the day Tuesday May 29.

**Academic Honesty:** As a student at CSU East Bay you are held to the standards stated in the Academic Dishonesty Policy. Copying another student’s work or allowing another student to copy your work is academically dishonest. I expect you to be academically honest while taking the test.

1. Random Forests:
   1. What is an **ensemble method**? Explain.
   2. Explain how **Random Forests** build decision trees.
2. Holdout:
   1. What is the **holdout method**? Explain. Draw a diagram.
   2. What is **repeated holdout**? Explain.
   3. Give the name of two repeated holdout methods.
3. Neural Networks:
   1. Sketch a diagram of a **multilayer feedforward neural network** with one **hidden layer**.
   2. (Stat. 6620) Why are continuous **activation functions** used with neural networks?
4. Market Basket Analysis:
   1. What is **transactional data**?
   2. Why are item matrices **sparse**?
   3. Explain the **Apriori property** that is used in the **Apriori algorithm**?