STA 314H1S: Quiz 3

Time allowed: 10 minutes; Total points: 10

Student Name: _____ Student ID: _____

1. (2 pts) You are working with a small dataset of patients, characterized by a large number of (correlated) genes. In order to build a predictive model using these genes as features, you intend to reduce the dimensionality while retaining as much information as possible. Which technique would be suitable for this task? (*Just write its name*).

2. (2 pts) Suppose you want a classifier that predicts the disease status of a patient using ten features derived from the previous question. If the ten features are conditionally independent given the patient's disease status, which classification method would be the most appropriate? (*Just write its name*).

3. (2 pts) You have created a model from the above scenario to predict whether a patient has a disease or not. If you want to evaluate its performance while considering both sensitivity and specificity in a single metric, which metric would be helpful? (*Just write its name*).

4. (2 pts) You only have a small dataset so you don't want to split them into separate training and testing sets. Which algorithm can you use to obtain an estimate of the testing performance of your metric in the last question using this single dataset? (*Just write its name*)

5. (2 pts) Which topic in the course have you found most challenging so far