# Home Work # 4. AMS 597

# Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_SBU ID:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# The homework is due on Monday, April 13, 2020, at 12noon. Solutions will be emailed to you by 1pm. The homework will not be graded and all complete submission will warrant full points. Quiz 6 will be given on Tuesday, April 14, 2020, in class, covering the same materials – for the entire class duration; and will be graded. So please study hard from now on. Please email your completed homework to your TA at: [song.jiecheng@stonybrook.edu](mailto:song.jiecheng@stonybrook.edu)

# Please include (1) R code

# (2) Output from R

# (3) Answers to all the questions asked

#### Logistic Regression with the bwt Data – Classification Task

The bwt data is a built in R dataset containing 9 variables and 189 cases – please see the following link for detailed description:

<http://garthtarr.github.io/mplot/articles/birthweight.html>

Your task is to split the data randomly into training (75%) and testing (25%), establish an optimal model using the training data, and then use that model to predict whether each infant in the testing data would be born with low birth weight or not.

Please review the following websites for related methods and concepts:

[https://medium.com/analytics-vidhya/a-guide-to-machine-learning-in-r-for-beginners-part-5-4c00f2366b90](https://medium.com/analytics-vidhya/a-guide-to-machine-learning-in-r-for-beginners-part-5-4c00f2366b90" \t "_blank)

[https://en.wikipedia.org/wiki/Receiver\_operating\_characteristic](https://en.wikipedia.org/wiki/Receiver_operating_characteristic" \t "_blank)

1. For the training data, please find a model that best predicts whether the infant would be born with low birth weight or not using the stepwise variable selection method, considering all 2-way interactions, and using the BIC. Please report the final model and the associated BIC value.

1. Please compute the Confusion matrix for the training data – please report the sensitivity, specificity and the overall accuracy based on the training data.
2. Please use the above model to predict whether each infant in the testing data would be born with low birth weight or not. Please compute the Confusion matrix for the testing data based on the default threshold of 0.5 – please report the sensitivity, specificity and the overall accuracy based on the testing data.
3. Please plot the ROC curve for the prediction of the testing data. What is the AUC value of your ROC curve?
4. (Extra credit) Can you perform the prediction using the Jackknife cross-validation (that is, the leave-one-out) method? Please compute the Confusion matrix for the results based on the default threshold of 0.5 – please report the sensitivity, specificity and the overall accuracy.