CSSE 386 – Data Mining with Programming
Rose-Hulman Institute of Technology

Worksheet 08

Name (Print):_____   Section:_____

1. Fill in the blanks:

    i. Logistic regression is designed to predict and explain a _____ variable.

    ii. If p is greater than 0.5, we predict the outcome _____ .

    iii. If p is less than 0.5, we predict the outcome _____ .

2. Provide the formula for logistic regression:

3. In Machine Learning, Logistic regression is often referred to as _____.

4. Provide the name for the fitting model method that minimizes **the sum of all the squared distances from the points to the line**. Note: review slides on linear regression

    Method: _____

5. Provide the main reason why we cannot use the Method of Least Squares for Logistic Regression:

6. What does ROC stand for?

7. The default probability threshold for logistic regression in scikit-learn is set at _____.

8. Explain why the logistic regression can be used for classification tasks:

9. Comparison between MLE and ROC-AUC. Provide the purpose, usage, and focus for each (consult the table in slides):