King Mongkut's University of Technology **Name:** _____

Machine Learning **I.D. Number:** _____

Suthep Madarasmi, Ph.D.

Take Home Quiz 6 Due Sat Mar 18, 2023 **Score:** _____ / 70

---

1. *2.5 hours.* Modify "Logistic Regression Gradient Descent.ipynb" for Iris classification to use the cost function and Jacobian function and the "BFGS" optimization method in SciPy library to:
   a. 20 points. Report the total classification accuracy (score) for the Test data by finding the highest probability class as the output.
   b. 5 points. Print the confusion matrix for the test data of the 3 classes found by your own implementation of the logistic regressor. You can use Sklearn's library; here's an example of how:

```
>>> from sklearn.metrics import confusion_matrix
>>> y_true = [2, 0, 2, 2, 0, 1]
>>> y_pred = [0, 0, 2, 2, 0, 2]
>>> confusion_matrix(y_true, y_pred)
array([[2, 0, 0],
       [0, 0, 1],
       [1, 0, 2]])
```

   c. 5 points. Print the confusion matrix for the test data found by sklearn's logistic regressor.
   d. 10 points. For each mis-classified data sample, show the classification probability for class 0, 1, and 2 (we want to see that the probabilities are hovering around the 0.2 to 0.9 range for these data points).

2. *30 points.* 3 hours. Using the Kaggle open source dataset to predict Attrition (Will an employee continue to stay with our company?) based on several factors using Logistic Regression. Make sure you show the **confusion matrix.** In the case of a Yes/No classification like this problem it shows false positives and false negatives. The data is available in:

   https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset