**Logistic Regression for binary classification**

(NumPy, Pandas and data visualization packages are allowed.)

(SKLearn regression models are allowed!)

Reference code: 2\_Logostic\_ExSKLearn\_Demo.py in blackboard

1. S**elect the pima diabetes dataset with binary target values from**

<https://machinelearningmastery.com/standard-machine-learning-datasets/>

1. **Use pandas to read CSV file as dataframe. (1pt)**

**e.g. The following code helps import pima diabetes dataset**

**﻿***col\_names = ['pregnant', 'glucose', 'bp', 'skin', 'insulin', 'bmi', 'pedigree', 'age', 'label']*

*# load dataset*

*pima = pd.read\_csv("pima-indians-diabetes-database.csv", header=None, names=col\_names)*

1. **Select 5 (if not possible then select 4) features from the chosen dataset. (1pt)**

**List all features you selected in your report.**

For example, the following code will select two features

*﻿feature\_cols = ['pregnant', 'age']*

*X = pima[feature\_cols]*

1. Use ﻿“﻿train \_test\_split” from sklearn.model\_selection to split test and training data by 40% testing + 60% training.  **(1pt)**
2. Fit your model with training data and test your model after fitting.
3. Calculate and print out

the confusion matrix **(1pt)**

precision score, recall score, F score **(3pts)**

**Copy your console output (these scores) to your report**.

1. Plot out the ROC curve and print out the ROC\_AUC score (sklearn.metrics.roc\_curve() and sklearn.metrics.roc\_auc\_score() can be used.) **(3pts)**

*--------------------------------------------------------------------------------------------------------------------*

**Submit your report and your code in two different files.**

**Please include the required figure/plot in your report.**

**e.g.**

**File1: Assignment2\_FirstnameLastname.doc/.pdf (this is the report)**

**+**

**File2: Assignment2\_ FirstnameLastname.py (this is the code. only “.py” files accepted.)**