Logistic Regression

Project Description Document:

1. General Information on dataset :

The name of dataset is (Predict Diabetes), number of classes is 2 , and their labels (0 or 1) that means she has diabetes or not , the total number of samples = 768 in dataset and the size of each (in case of images), and finally the number of samples used in training = 614 and testing =154.

1. Implementation details

* At feature extraction phase: features were extracted = 8 , their names are ['Pregnancies','Glucose','BloodPressure','SkinThickness','Insulin','BMI','DiabetesPedigreeFunctin', 'Age', 'result'] the dimension is (0 or 1) depends on the values of resulted features.
* Hyper parameters used in the model learning rate =0 .5 , max iteration = 1000

1. Results details :

confusion\_matrix is :

[[87 3]

[27 37]]

Test accuracy is : 80.51948051948052

ROC curve :

