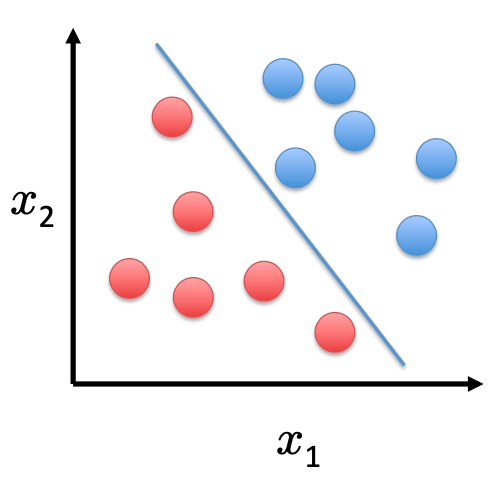
Programming Assignments 1

ICS485 – Machine Learning

1. [40 points] Modify BGD and SGD codes from least squares linear regression for implementing **batch gradient ascent and stochastic gradient ascent for logistic regression** (you can find the codes from ICS485\_IBL.ipynb that has been uploaded in Ms Teams folder). As logistic regression is for binary classification purpose, you cannot use the generated data there [the pairs of x and y that are generated using make\_regression()]. The preferable form of the dataset should be like this:



But, the number of points should be at least 1000 to make it better.

After you generate the data points,

1. Train your model using all data points, and plot the data points using different colors for different classes (there should be two plots, the first one corresponds to the result of batch gradient ascent, and the second one corresponds to the stochastic gradient ascent)
2. Divide the data point into training data (80%) and test data (20%) randomly. Train your model (using both batch and stochastic gradient ascent) using the training data, and measure their classification performances using the following metrics
   1. Accuracy
   2. F1 score
   3. Sensitivity
   4. Precision
3. [30 points] Derive Hessian matrix for cost function of logistic regression. Then implement the Newton’s method for logistic regression. Use this method for updating parameters in logistic regression using the same dataset that you generated in question 1.
4. [30 points] Find ‘data.csv’ which is a diabetes dataset containing 768 number of instances (patients) and 8 features. Open the dataset, you will see a matrix with 9 columns and 768 rows. The last row contains information about classes of each instances (0 or 1). Repeat tasks in question 1(b) using this dataset.