Machine Learning Assignment 2

Que Shuhao(4739124)

March 13, 2018

Real

a

Two methods of semi-supervised learning for LDA classifier

Method1: Self-learning

Self-learning method is implemented to train LDA classifier. First, 25 labeled data randomly selected from the whole dataset are applied to train LDA classifier, based on which, every 2 unlabeled data are used for classifier to perform classification task. The threshold is set as 0.9 (the difference between two confidence scores of two classes g and h). If the confidence score of the label datum assigned by the LDA classifier exceeds the threshold, then this very datum will be added to the labeled dataset. Repeat this procedure.

Method2: Semi-supervised Clustering

Here semi-supervised clustering is applied. First, 25 randomly selected labeled data are used to calculate two means of classes g and h. Then clustering method is applied to classify fixed amount of the unlabeled data [0, 10, 20, 40, 80, 160, 320, 640] based on the Euclidean distance between the unlabeled data and the means. Then these data labeled by clustering method are added to the labeled data. Based on the newly required labeled dataset, the LDA classifier is trained.

b

I repeat the experiments 40 times to get relatively smooth curves. The results are shown in Figure 1

c

The results are shown in Figure 2.

Imaginary

d

The two generated dataset are displayed respectively in Figure 3 and Figure 4

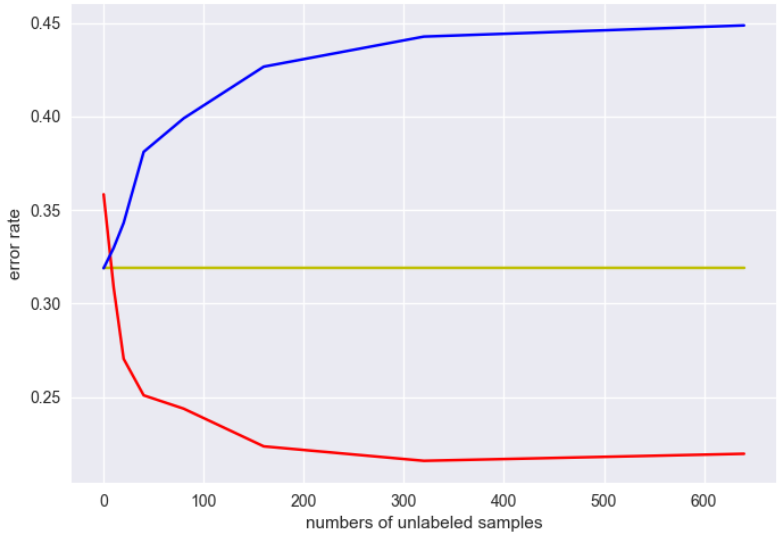

Figure 1: red: self-learning
blue: semi-supervised clustering
yellow: supervised learning

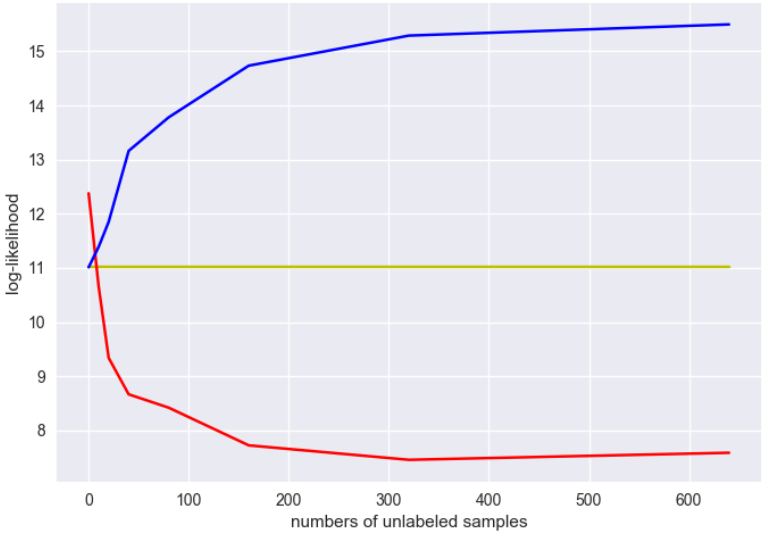

Figure 2: red: self-learning
blue: semi-supervised clustering
yellow: supervised learning

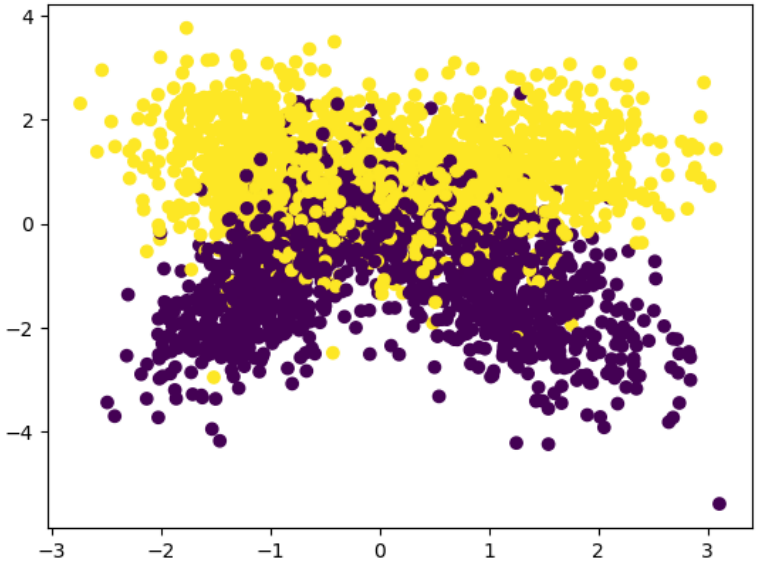

Figure 3: This is the dataset where Self-learning outperforms both supervised Learning and Clustering method

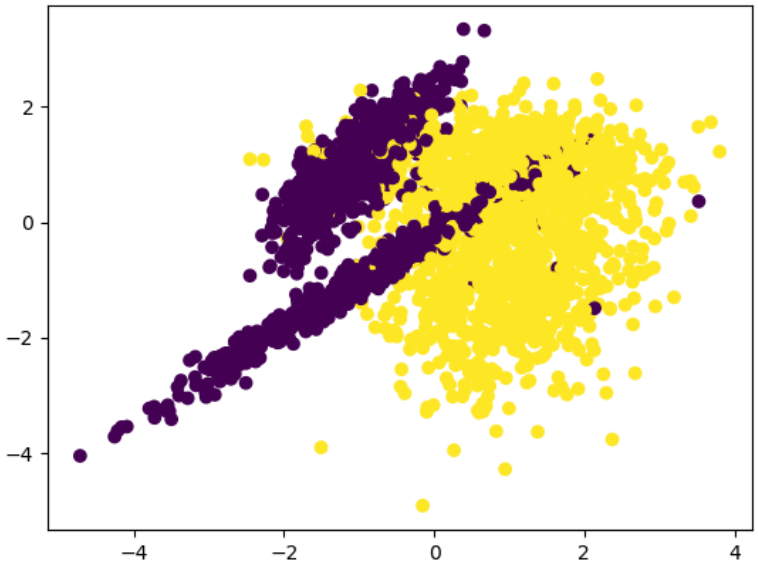

Figure 4: This is the dataset where Clustering method outperforms both supervised Learning and Self-learning