

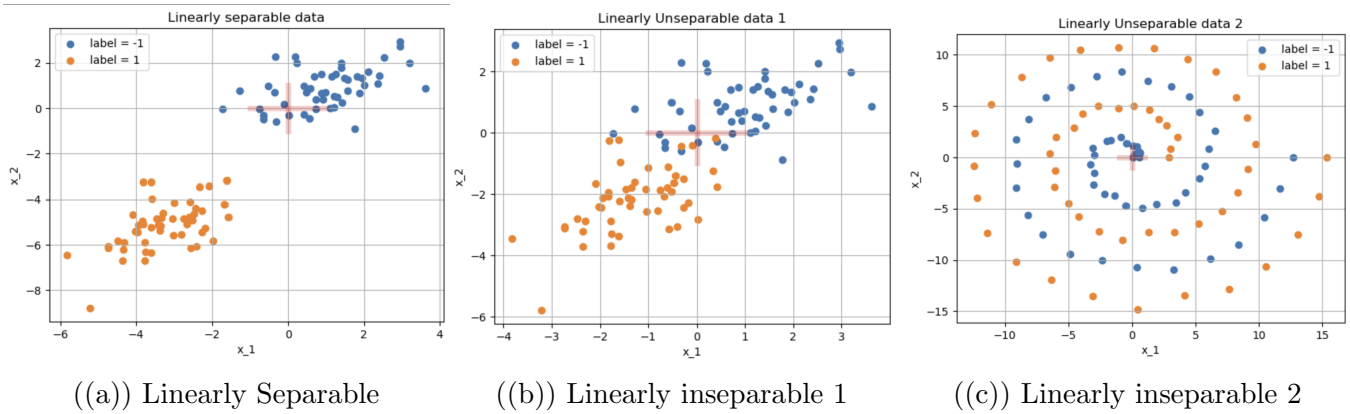
STA4042 24-25T1 Homework 3: SVM

Inquiry: 223040237@link.cuhk.edu.cn

October 29, 2024

Goal

In this assignment, you will implement SVM from scratch. Sklearn (or any other published SVM) is not allowed. We will work on 3 datasets:



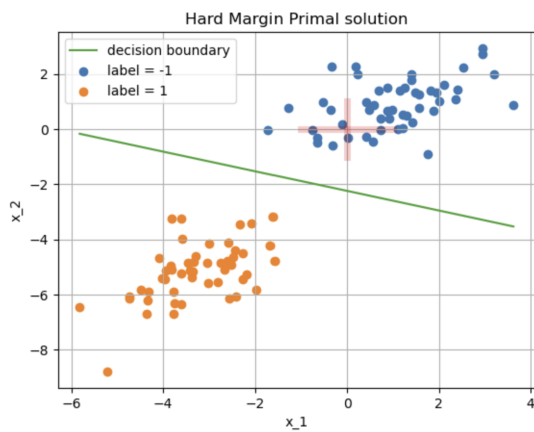
We will not split train and validation sets for this assignment.

Questions

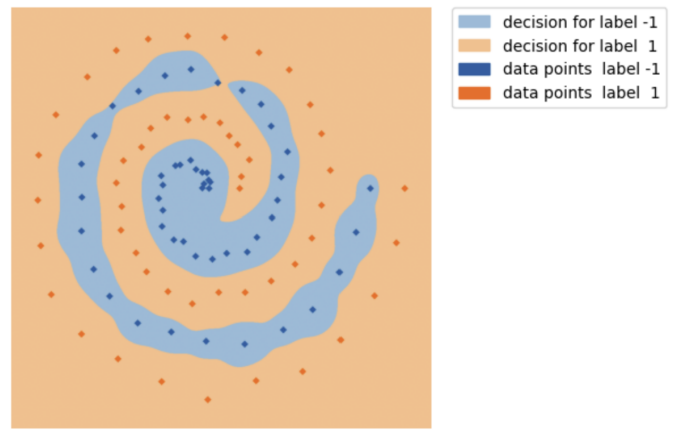
Part 1 - With Linearly separable data (10+5+10+5+5 = 35 points)

1. Solve Primal problem for Hard Margin configuration.
2. Show decision boundary together with the data points. One example is given in Fig 2(a):
3. Solve Dual problem for Hard Margin configuration.
4. Show decision boundary together with the data points.
5. Do Primal and Dual give the same hyperplane ? Show that the difference is of a very small order (e.g. smaller than $1e-8$).

Part 2 - With Linearly inseparable data 1 (5+5+5+10+5 = 30 points)



((a)) Hard Margin SVM



((b)) Kernel SVM (RBF Hard Margin)

Figure 2: Some Decision Boundary display examples

6. Does Hard Margin Primal solution still work? Show decision boundary together with the data points.
7. Does Hard Margin Dual solution still work? Show decision boundary together with the data points.
8. Do Primal and Dual still give the same hyperplane ? Do you have some justifications about why it is no longer the same result ?
9. Solve Dual problem for Soft Margin configuration. Choose your parameters freely as long as the classification works out.
10. Show decision boundary together with the data points.

Part 3 - With Linearly inseparable data 2 (10+5+5+10+5 = 35 points)

11. Choose your own kernel and implement Kernel SVM (either hard or soft margin, the goal is to succeed the classification).
12. Calculate your prediction on train set. Do you have 100% accuracy?
13. Try Spectral Clustering. Does it work with your SVM? You do not need to have very visible results to obtain the points.
14. Find a way to display your decision boundary. One example is given in Fig 2 (b). If you find it difficult to add data points to the plot, you can skip data points and only display the decision boundary.
15. Make some short comments: Given the decision boundary, do you think your model performs well on the data? Will it adapt well for further generalization of data ? Justify your opinion.