DDA2020: Homework II

March 16, 2022

Task Description   In this problem you are asked to write a program that construct support vector machine models with different kernel functions and slack variable. An *svm.ipynb* file is provided, where you have to implement the missing functions.

Datasets   You are provided the training and testing datatset (see *train.txt* and *test.txt*), including 120 training data and 30 testing data, respectively. It covers 3 classes, corresponding to setosa, versicolor, virginica. They are derived from the Iris dataset (`https://archive.ics.uci.edu/ml/datasets/iris`), contains 3 classes of 50 instances each, where each class refers to a type of iris plant. Your task is to classify each iris plant as one of the three possible types.

What you should do

1. Implement standard SVM model. Fit your algorithm on the training dataset, then validate your algorithm on testing dataset. Compute the misclassification error of training and testing datasets.

2. Implement SVM with slack variables. For each $C \in [0, 1]$ with the step 0.1, fit your algorithm on the training dataset, then validate your algorithm on testing dataset. Compute the misclassification error of training and testing datasets. Save the results in a table, depending on the constant C.

3. Implement SVM with kernel functions. As a kernel function, you are supposed to choose: (a) a 2nd-order and a 3rd-order polynomial kernel (b) Radial Basis Function (RBF) kernel with two different kernel sizes.