Homework 2

Student Number: 118033910019
Name: Xingyi Wang

Problem 1. In this assignment, the support vector machine (SVM) and the Min-Max-Module strategy will be used to deal with multi-class classification problems. SVMs usually handle binary classification tasks. If there are several classes to recognize, some strategies are needed. The most common methods are one-vs-one strategy, one-vs-rest strategy, and part-vs-part strategy.

Two problems are given below. The dataset used in this homework is the SJTU Emotion EGG Dataset (SEED), which is a three-class classification problem. Four files are included in the .zip file: train_data.npy, train_label.npy, test_data.npy, and test_label.npy, and 37367 samples are included in the training data, and 13588 samples in the test data.

Solving the three-class classification problem in the given dataset using SVM classifiers and the **one-vs-rest** strategy. SVM classifiers are provided in LibSVM package and other machine learning libraries (sklearn). You can use these libraries to solve this problem.

Solution. In SEED, the labels are 1, 0, and -1. To solve this three-class classification problem using one-vs-rest strategy, we need to train three SVM classifiers. The first one is trained to classify label 1 and the others; the second one is trained to classify label 0 and the others; and the third one is trained to classify label -1 and the others. In the testing stage, the testing sample is classified by all three classifiers, and they return the probability of this sample belonging to the corresponding class. Finally, the testing sample will be classified into the class whose probability is the highest.

The coding is quite simple. For each SVM, we only need to set the corresponding label to 1 and the rest to 0. For example, for the SVM that is to classify label 1 and the others, we set the label 1 as 1, and convert label 0 and -1 to 0. The penalty parameter C of the error term is 1.0; the kernel function is *Radial Basis Function*.

The result is shown below:

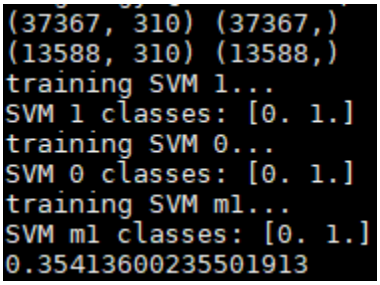

Figure 1: Caption

SVM 1/0/m1 are the classifiers that classify the label 1/0/-1 separately. The SVM using one-vs-rest strategy can obtain an accuracy about 35.41%.

Problem 2. Solving the three-class classification problem using Min-Max-Module SVM and **part-vs-part** task decomposition method. You should divide the three-class problem into three two-class problems using one-vs-rest method and then decompose these imbalance two-class problems into balance two-class problems following random task decomposition and task decomposition with prior knowledge strategies. Please compare the performance of SVMs obtained in Problem one and the Min-Max-Module SVMs here.

Solution. For this problem, we first partition the dataset using one-vs-rest strategy. Then for the *rest* part, we use random task decomposition to achieve part-vs-part strategy. In SEED, there are 12903 samples labeled 1, 12144 samples labeled 0, and 12320 samples labeled -1. For example, if we take class 1 as the *one* part, then the *rest* part is composed of class 0 and class -1. The *rest* part is shuffled and then randomly partitioned into 2 subsets. The reason behind 2 subsets is that we should use balanced dataset in part-vs-part strategy. Then we train two SVM models, one model for class 1 and subset 1, and the other for class 1 and subset 2. The similar dataset processing is implemented for class 0 and class -1, which results in six SVM models. Finally, a Min-Max-Module is implemented based on the prediction probability to generate the final classification results. To be more specific, the results from the same *one* part are handled using Min-Module, and all the results from Min-Module are handled using Max-Module.

The final result is shown below:

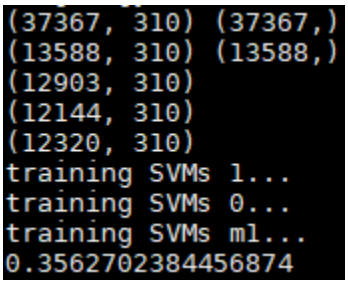

Figure 2: Caption

The SVM using part-vs-part strategy can obtain an accuracy about 35.63%, which is higher than that using one-vs-rest strategy.