CS7327-033-M01 Assignment 2

Liong Khai Jiet (120033990010)

April 14, 2021

Problem 1.

The dataset used in this homework is the SJTU Emotion EEG Dataset (SEED). Out of 37367 training data, the category label distribution are as follows:

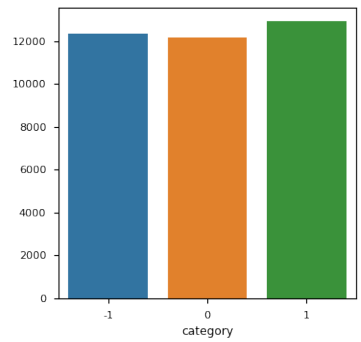

Fig. 1. Training loss of 12 hidden units and 50 hidden units model

1.1 Preprocessing

First, the data are preprocessed using the `sklearn.preprocessing.MaxAbsScaler()` to scale the dataset by its maximum absolute value, so the values are mapped within the range [0,1] as suggested by this experiment study [3]. `MinMaxScaler()` and `StandardScaler()` is not used because both are very sensitive to the presence of outliers. Besides, Principal Component Analysis (PCA) is applied with `n_components = 0.85` helps the model to produce better results by reducing the dimension of the training dataset.

Then, the one-vs-rest strategy is applied to break the 3 class problem into $3 = \frac{3*(3-1)}{2}$ binary classification problems. For every class i, the inputs of all other class are set to -1 and the remaining positive labels are set to 1.

1.2 Training

Each binary classification problem is then feed into an individual SVM classifier which acts as a binary classifier. `svm.LinearSVC` is chosen as the model because it is based on the `libLinear` library where the author [2] proved that it has better performance and efficient training times, especially for datasets with large number of samples compared to the number of features.

1.3 Results

The findings are tabulated in the table 1 as follows:

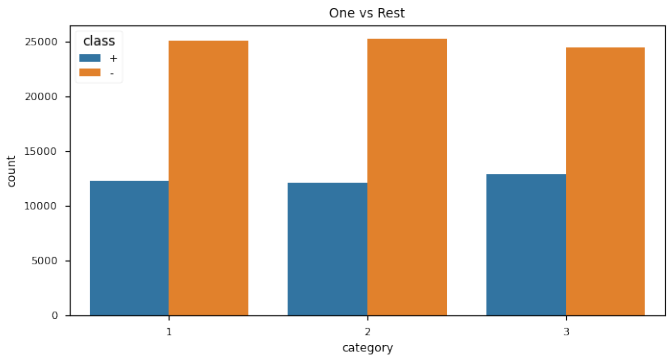

Fig. 2. Positive and Negative class size using One Vs Rest Strategy

Table 1: Classification Report of one-vs-rest strategy and SVM

Class	Precision	Recall	F-score	Support
-1	0.52	0.38	0.44	4480
0	0.52	0.50	0.51	4416
1	0.63	0.81	0.71	4692
micro avg	0.56	0.56	0.55	13588
macro avg	0.56	0.56	0.55	13588
avg	0.56	0.57	0.55	13588
accuracy			0.57	13588

With the help of grid search technique, different value of C (C = 1 (default) ,10, 100, 1000) is also tested, but the results show the regularization does not improve the results anymore.

Problem 2.

2.1 Preprocessing

Similarly, problem 2 is preprocessed in the similar fashion as problem 1 with an additional to one more step. After classifying the training data with the one versus rest strategy, the imbalance data is further broken down into more balance count using the part versus part strategy as follows:

```
def get_partition_size(x, rho = 10000, gamma = 0.8):
    n_samples = x
    total_samples = n_samples * 2
    part_samples = (total_samples) / rho
    if(total_samples > rho and math.fmod(total_samples,rho) <= gamma):
        return math.floor(part_samples)
    return math.ceil(part_samples)
```

where rho is the desired number of training data, $\sum N_{+} + \sum N_{-}$ and $\gamma \in (0, 1)$ for finetuning purpose [1]. In this setting, the number of sub training sets are reduce to 3000 binary classification problem labeled as τ_{ij}, where i, j is the label given to the respective class.

2.2 Training

Same as the previous experiment, the svm model used is a `Liblinear`. Grid Search results also shows that the tuning of hyperparameter C has no effect on the model performance.

A total of 12 SVM models are trained separately on the specific 3000 training data. For example, svm_{ij} is responsible for the τ_i and τ_j. After training all the individual SVMs assigned to the two-class problems, they are integrated into a min-max modular svm network and the prediction of the original three-class problem is obtained by:

```
C = np.argmax(y_i(x),axis=0)
```

where C is the class label $\in \{-1, 0, 1\}$

2.3 Results / Discussion

Table 2 shows the result of the experiment. The f1 score of the m3 model is 0.52 which is slightly lower that the one versus one counter part with the score 0.57. Besides, the time taken to train the m3 model is doubled compared to the first experiment at about 256ms compared to 124ms.

The results obtained is not as expected which could be caused by reasons such as wrong implementation of the part-versus-part strategy or the preprocessed step applied is unsuitable for the problem. The optimization process could not be continued due to insufficient time.

Table 2: Classification Report of part-vs-part strategy and M3-SVM

Class	Precision	Recall	F-score	Support
-1	0.41	0.42	0.42	4480
0	0.48	0.37	0.42	4416
1	0.63	0.75	0.68	4692
macro avg	0.51	0.52	0.51	13588
avg	0.56	0.57	0.55	13588
accuracy			0.52	13588

References

[1] Lu Bao-Liang, Wang Xiao-Lin, Yang Yang, and Zhao Hai. Learning from imbalanced data sets with a min-max modular support vector machine. *Frontiers of Electrical and Electronic Engineering in China*, 6(1):56–71, 2011.

[2] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *the Journal of machine Learning research*, 9:1871–1874, 2008.

[3] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification, 2003.