CSE 847 (Spring 2022): Machine Learning— Homework 4

Han Meng

1 Logistic Regression: Experiment

I set epsilon as 5e-2 and maxiter as 1000 respectively.

Generally speaking, the Accuracy increases as the number of trainning data increases as the below figure shows.

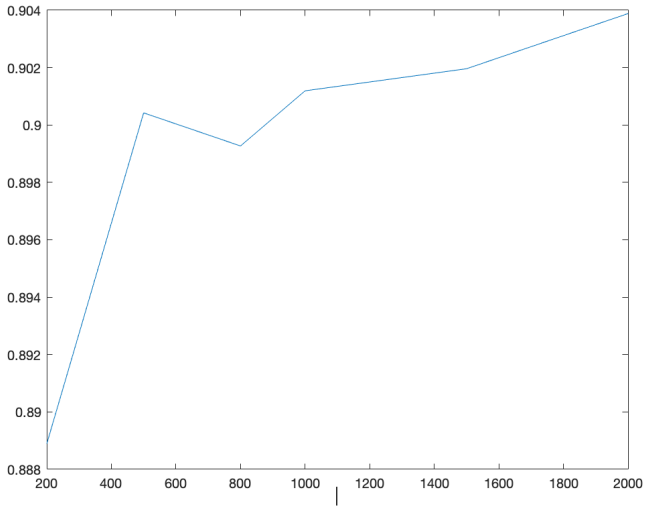

Figure 1: Accuracy vs. Number of Trainning Data

```
200: 0.88889
500: 0.90042
800: 0.89927
1000: 0.90119
1500: 0.90196
2000: 0.90388
```

2 Sparse Logistic Regression: Experiment

In this part, you will perform experiments using sparse logistic regression (ℓ_1-regularized logistic regression). Use the Alzheimer's disease dataset as described in `https://github.com/jiayuzhou/CSE847/tree/master/data/alzheimers`. Sparse logistic regression is then applied to train a linear model on the given training set and prediction is then performed on the given test set. You should use the implementation in SLEP[1], where the sparse logistic regression is the function `LogisticR`[2]. An example of using the sparse logistic regression is as follows:

[1] `https://github.com/jiayuzhou/SLEP/`

[2] Located at `https://github.com/jiayuzhou/SLEP/tree/master/SLEP/functions/L1/L1R`

```
function [w, c] = logistic_l1_train(data, labels, par)
% OUTPUT w is equivalent to the first d dimension of weights in logistic_train
%        c is the bias term, equivalent to the last dimension in weights in logistic_train.

% Specify the options (use without modification).
opts.rFlag = 1;  % range of par within [0, 1].
opts.tol = 1e-6; % optimization precision
opts.tFlag = 4;  % termination options.
opts.maxIter = 5000; % maximum iterations.

[w, c] = LogisticR(data, labels, par, opts);
```

The input `par` is the ℓ_1 regularization parameter, which scales from 0 to 1. Try different values of regularization parameter and report both the AUC (use Matlab code `perfcurve`) and the number of features selected (number of non-zero entries in `w`). A suggested list of parameters is [0, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1], but other choices of parameters are also encouraged. Note that when parameter is 0, the formulation is equivalent to the classical logistic regression.

For both experiments in this homework, submit a brief report. In addition to the report, submit the MATLAB code (do add some comments in your code for others to understand your code) to a public repository under your Github account (the same account of your project) and include the link in the report.