CSCI5525 Assignment 2 Answer

Daihui DOU   dou00005@umn.edu   5514178

October 29, 2018

1 Problem 1

1.1 (b)

The result of different C is listed below.

myDualSVM_result.csv

```
c,error_mean,error_std,sv_num_mean,sv_num_std,margin_mean,margin_std
0.01,0.0165,0.0070,101.3,4.10,78.698,3.700
0.1,0.0140,0.0106,101.0,4.73,78.486,2.786
1,0.0120,0.0078,101.4,4.36,78.287,3.357
10,0.0155,0.0121,100.6,5.67,79.213,5.988
100,0.0145,0.0096,98.9,4.08,79.039,7.413
```

ii. The mean error basically stay the same when C increases, while the error std first increases and than decreases.
iii. The geometric margin $1/||w||$ mean basically do not change as C increases, while the std of the margin increases a bit.
iv. The mean and std of the number of support vectors both basically do not change as C increases.

1.2 (c)

The problem is to get

$$\min \frac{1}{2}||\mathbf{w}^2|| + C\sum_i \xi_i$$

, such that

$$y_i f(x_i) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

The Lagrangian Dual is given by:

$$\mathcal{L} = \frac{1}{2}||\mathbf{w}^2|| + C\sum_i \xi_i - \sum_i a_i(y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 + \xi_i) - \sum_i \mu_i\xi_i$$

Write $\mathcal{L}$ as functions of $\mathbf{w}$, b:

$$\mathcal{L} = \frac{1}{2}||\mathbf{w}^2|| + C\sum_{i=1:N} \xi_i(\mathbf{w}, b)$$

$$\mathcal{L} = \frac{1}{2}||\mathbf{w}^2|| + C\sum_{i=1:N} \max(0, 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b))$$

, such that

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

The margin constraints are only satisfied when $\xi_i = 0$.

2 Problem 2

2.1 Pegasos

This section implemented the Pegagos algorithm introduced in "Pegasos: Primal Estimated sub-GrAdient SOlver for SVM" by S. ShalevShawtrz, Y. Singer, and N. Srebro. The algorithm shows below.
The primal objection function is given by:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \max(0, 1 - y_i \mathbf{w}^T \mathbf{x}_i) + \lambda ||w||$$

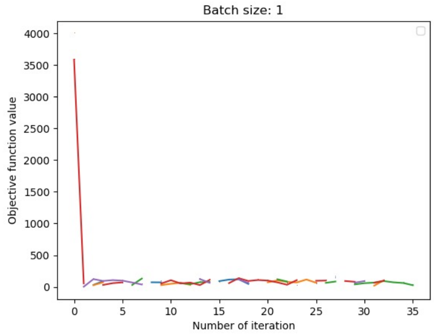
Pegasos Algorithm:
INPUT: S, λ, T, k
INITIALIZE: Choose $\mathbf{w}_1$ s.t. $\|\mathbf{w}_1\| \leq 1/\sqrt{\lambda}$
FOR $t = 1, 2, \ldots, T$
Choose $A_t \subseteq S$, where $|A_t| = k$
Set $A_t^+ = \{(\mathbf{x}, y) \in A_t : y \langle \mathbf{w}_t, \mathbf{x} \rangle < 1\}$
Set $\eta_t = \frac{1}{\lambda t}$
Set $\mathbf{w}_{t+\frac{1}{2}} = (1 - \eta_t \lambda)\mathbf{w}_t + \frac{\eta_t}{k} \sum_{(\mathbf{x},y) \in A_t^+} y \, \mathbf{x}$
Set $\mathbf{w}_{t+1} = \min \left\{ 1, \frac{1/\sqrt{\lambda}}{\|\mathbf{w}_{t+\frac{1}{2}}\|} \right\} \mathbf{w}_{t+\frac{1}{2}}$
OUTPUT: $\mathbf{w}_{T+1}$

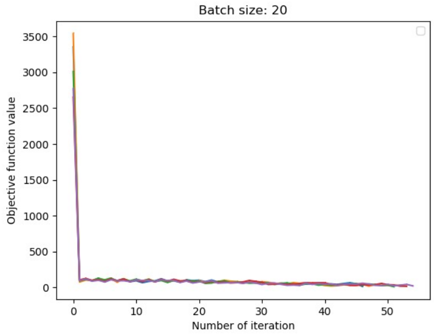
The optimization performance is based on the change rate of weight. The program terminated when the change is smaller than some threshold. The result lists below, noted that the running time increases when k, the batch size, increases.

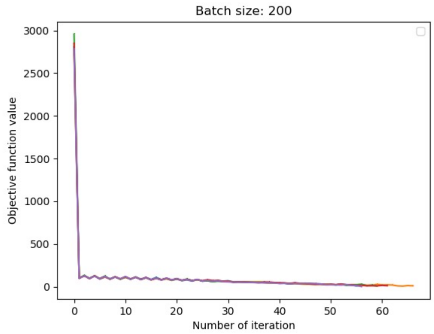
myPegasos_result.csv

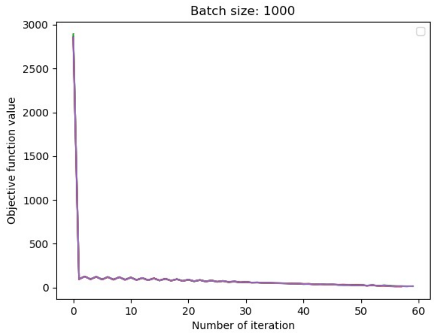
```
k,mean,std
1,0.008799457550048828,0.0022714600904186395
20,0.0601095199584960950,0.015712337706001236
200,0.20170950889587402,0.022749538595149978
1000,1.0287991523742677,0.05117348627559163
2000,1.8022011756896972,0.11140784219178905
```

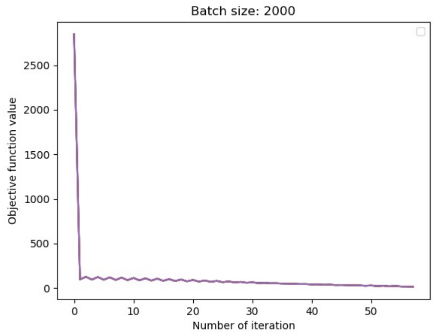
The primal objective function value for each run with increasing number of iterations listed below.











2.2 Softplus

The gradient is given by:

$$\nabla \mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \mathbf{x}_i}{1 + \exp(-(1 - y_i \mathbf{w}^T \mathbf{x}_i)/a)} + 2\lambda ||w||$$

The stochastic gradient descent:

Stochastic gradient descent:

- For $t = 1, \ldots, T$
  - Randomly draw $i_t \in \{1, \ldots, m\}$
  - Compute (sub)gradient $g_{i_t} = \nabla \ell_i(\mathbf{w}_t)$
  - $\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t g_{i_t}$
- Output $\bar{\mathbf{w}}_T = \frac{1}{T} \sum_{t=1}^{T} \mathbf{w}_t$

The optimization performance is based on the norm of gradient. The program terminated when the norm of gradient is smaller than some threshold. The result lists below, noted that when k, the batch size, increases, the running time first increases and then basically do not change.

```
k,mean,std
1,0.06664814949035644,0.03320276156772877
20,1.365898036956787,0.6015587947649584
200,11.616263675689698,1.3419750076627754
1000,12.662287950515747,1.1349438654580997
2000,10.16286244392395,1.1236828749218788
```

The primal objective function value for each run with increasing number of iterations listed below.