$\propto$SVM for Learning with Label Proportions Supplementary Material[1]

Felix X. Yu† yuxinnan@ee.columbia.edu
Dong Liu† dongliu@ee.columbia.edu
Sanjiv Kumar§ sanjivk@google.com
Tony Jebara† jebara@cs.columbia.edu
Shih-Fu Chang† sfchang@cs.columbia.edu
†Columbia University, New York, NY 10027
§Google Research, New York, NY 10011

1. Supplement for alter-$\propto$SVM

1.1. Proof of Proposition 1

Proof 1 *We consider the k-th bag in this proof.*

We first note that the influence of y_i, $\forall i \in \mathcal{B}_k$ to the first term of the objective function, $\sum_{i\in\mathcal{B}_k} L(y_i, \boldsymbol{w}^T\varphi(\boldsymbol{x}_i)+b)$, is independent.

Without loss of generality, we assume $\mathcal{B}_k = \{1\cdots|\mathcal{B}_k|\}$. Also without loss of generality, we assume δ_i's are already in sorted order, i.e. $\delta_1 \geq \delta_2 \geq ... \geq \delta_{|\mathcal{B}_k|}$.

Define $\{i|y_i = 1, i \in \mathcal{B}_k\} = \mathcal{B}_k^+$, and $\{i|y_i = -1, i \in \mathcal{B}_k\} = \mathcal{B}_k^-$. In order to satisfy the label proportion, the number of elements in $\{y_i|i \in \mathcal{B}_k\}$ to be flipped is $\theta|\mathcal{B}_k|$. We are to solve the following optimization problem.

$$\max_{\mathcal{B}_k^+} \quad \sum_{i\in\mathcal{B}_k^+} \delta_i - \sum_{i\in\mathcal{B}_k^-} \delta_i, \quad s.t. \quad |\mathcal{B}_k^+| = \theta|\mathcal{B}_k|.$$

What we need to prove is that $\mathcal{B}_k^+ = \{1, 2, ..., \theta|\mathcal{B}_k|\}$ is optimal.

Assume, on the contrary, there exists $\mathcal{B}_k^{+*}$, and $\mathcal{B}_k^{-*}$, $|\mathcal{B}_k^{+*}| = \theta|\mathcal{B}_k|$, $\mathcal{B}_k^{+*} \neq \{1, 2, ..., \theta|\mathcal{B}_k|\}$, $\mathcal{B}_k^{+*} \cup \mathcal{B}_k^{-*} = \mathcal{B}_k$, $\mathcal{B}_k^{+*} \cap \mathcal{B}_k^{-*} = \emptyset$, such that $\left(\sum_{i\in\mathcal{B}_k^{+*}} \delta_i - \sum_{i\in\mathcal{B}_k^{-*}} \delta_i\right) - \left(\sum_{i=1}^{\theta|\mathcal{B}_k|} \delta_i - \sum_{i=\theta|\mathcal{B}_k|+1}^{|\mathcal{B}_k|} \delta_i\right) > 0$

However, $\sum_{i\in\mathcal{B}_k^{+*}} \delta_i - \sum_{i=1}^{\theta|\mathcal{B}_k|} \delta_i \leq 0$, $\quad \sum_{i=\theta|\mathcal{B}_k|+1}^{|\mathcal{B}_k|} \delta_i - \sum_{i\in\mathcal{B}_k^{-*}} \delta_i \leq 0$. A contradiction.

1.2. Proof of Proposition 2

Proof 2 *As described in the paper, the influences of the bags in the objective function (6) are independent, and for the k-th bag, the algorithm takes $\mathcal{O}(|\mathcal{B}_k|\log(|\mathcal{B}_k|)$, $\forall k = 1\cdots K$.*

Overall, the complexity is $\mathcal{O}(\sum_{k=1}^{K} |\mathcal{B}_k|\log(|\mathcal{B}_k|))$.

[1] This article is the supplementary material of (Yu et al., 2013)

We know that $\sum_{k=1}^{K} |\mathcal{B}_k| = N,\ J = \max_{k=1 \dots K} |\mathcal{B}_k|$.

$$\sum_{k=1}^{K} |\mathcal{B}_k| \log(|\mathcal{B}_k|) \leq \sum_{k=1}^{K} |\mathcal{B}_k| \log(J) = N \log(J).$$

1.3. Justification of The Annealing Loop

We use an annealing loop for alter-$\propto$SVM to alleviate the local minima issues. To justify the requirement of the annealing loop, we keep repeating the alter-$\propto$SVM algorithm with/without the annealing loop, with different random initializations, on the same dataset. We record the smallest objective value found so far. As shown in Figure 1, alter-$\propto$SVM without the annealing loop fails to find a low objective value within a reasonably amount of time, while alter-$\propto$SVM with annealing loop can find a near-optimal solution really fast in about 3 seconds (a few runs). Similar results can be found on other datasets, and other bag sizes. In the experiment section we empirically choose to initialize alter-$\propto$SVM 10 times, which gives us quite stable results.

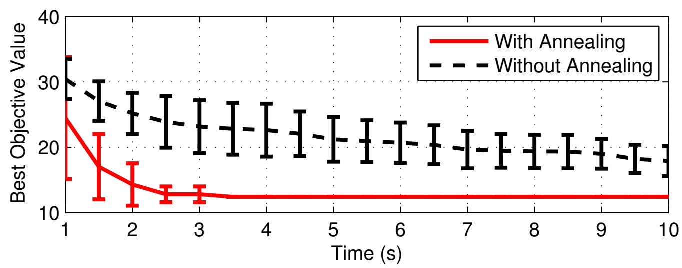

Figure 1. The smallest objective value with/without the annealing loop. The above results are based on experiments on the vote dataset with bag size of 32, linear kernel, $C = 1$, $C_p = 10$.

Due to the usefulness of annealing for $\propto$SVM, deterministic annealing (Sindhwani et al., 2006) can be explored to further improve the algorithm.

2. Supplement for conv-$\propto$SVM

2.1. Proof of Proposition 3

Proof 3 *The proof is identical to the proof of Proposition 2, except that we need to consider* d *dimensions of* $\boldsymbol{x}$*, independently.*

3. Additional Experiment Results

We show additional experiment results in Table 1 and Table 2.

References

Sindhwani, V., Keerthi, S.S., and Chapelle, O. Deterministic annealing for semi-supervised kernel machines. In *Proceedings of the 23rd International Conference on Machine learning*, pp. 841–848, 2006.

Yu, F.X., Liu, D., Kumar, S., Jebara, T., and Chang, S.-F. $\propto$SVM for learning with label proportions. In *Proceedings of the 30rd International Conference on Machine learning*, 2013.

Dataset	**Method**	2	4	8	16	32	64
heart-c	MeanMap	79.69±2.67	77.57±1.27	78.09±1.29	74.89±2.08	74.47±2.48	76.51±2.18
	InvCal	79.81±1.30	78.52±0.66	76.50±2.61	75.91±2.21	72.36±2.77	73.94±1.68
	alter-∝SVM	**81.39±1.19**	**79.93±0.81**	**79.61±1.22**	74.72±3.01	76.00±1.97	**78.11±2.81**
	conv-∝SVM	78.99±1.03	75.59±2.64	77.91±0.98	**77.29±0.48**	**77.99±1.78**	76.71±1.88
breast-cancer	MeanMap	96.49±0.01	96.34±0.18	96.21±0.20	96.20±0.34	96.35±0.36	96.56±0.55
	InvCal	96.02±0.22	96.11±0.61	95.81±0.23	95.61±0.29	95.61±0.12	94.49±1.00
	alter-∝SVM	**96.90±0.20**	**96.87±0.13**	**96.81±0.36**	**96.76±0.28**	**96.82±0.50**	**96.84±0.41**
	conv-∝SVM	93.88±0.16	93.86±0.11	93.82±0.06	95.13±0.33	95.63±0.44	96.12±0.11
credit-a	MeanMap	85.42±0.22	84.79±0.70	83.26±1.58	81.32±1.03	**81.18±2.92**	79.24±4.79
	InvCal	85.51±0.00	85.40±0.41	84.52±0.73	82.69±3.20	79.23±4.31	77.99±5.68
	alter-∝SVM	**85.54±0.12**	**85.51±0.33**	**85.37±0.34**	**83.59±3.17**	80.98±4.68	80.16±4.79
	conv-∝SVM	85.51±0.00	85.24±0.41	82.69±0.90	81.77±1.38	80.13±1.45	**80.79±1.38**
breast-w	MeanMap	96.11±0.06	95.97±0.25	96.13±0.16	96.26±0.32	95.96±0.42	95.80±0.92
	InvCal	95.88±0.33	95.65±0.36	95.53±0.24	95.39±0.57	95.23±0.52	94.31±0.77
	alter-∝SVM	**96.71±0.29**	**96.77±0.13**	**96.59±0.24**	**96.41±0.50**	**96.41±0.21**	**96.25±0.49**
	conv-∝SVM	92.27±0.27	92.25±0.16	92.32±0.13	94.03±0.18	94.60±0.10	94.57±0.21
a1a	MeanMap	81.76±0.89	81.60±0.47	**80.02±0.59**	77.04±1.30	73.19±2.48	72.58±0.95
	InvCal	81.86±0.20	81.35±0.70	78.34±0.70	**77.69±1.36**	73.13±4.86	73.30±1.71
	alter-∝SVM	**82.63±0.36**	**81.72±0.58**	80.00±1.46	76.48±1.08	**76.38±1.31**	**76.09±0.88**
	conv-∝SVM	75.63±0.33	75.39±0.01	75.39±0.01	75.40±0.02	75.37±0.04	75.37±0.05
dna-3	MeanMap	87.57±0.74	83.95±1.34	80.22±0.65	79.14±2.39	75.21±0.89	74.99±1.53
	InvCal	91.77±0.42	89.38±0.41	87.98±0.83	84.28±1.63	79.65±3.55	75.22±5.64
	alter-∝SVM	**93.21±0.33**	**92.83±0.40**	**91.80±0.52**	**88.77±1.10**	**86.94±0.41**	**86.39±1.70**
	conv-∝SVM	91.72±0.26	87.93±1.32	80.13±2.39	73.93±0.46	73.38±0.56	72.87±0.79
satimage-3	MeanMap	94.44±0.25	93.90±0.30	93.66±0.49	92.39±1.64	89.26±0.20	88.77±0.45
	InvCal	94.12±0.33	94.25±0.25	94.08±0.18	93.66±0.31	93.41±0.52	92.34±0.56
	alter-∝SVM	**95.13±0.27**	**95.11±0.32**	**95.09±0.26**	**94.89±0.15**	**94.54±0.22**	**94.46±0.44**
	conv-∝SVM	88.44±0.45	87.18±0.36	86.41±0.47	90.66±0.53	93.17±0.62	93.26±0.51

Table 1. Additional experiments. Accuracy with linear kernel, with bag size 2, 4, 8, 16, 32, 64.

Dataset	**Method**	2	4	8	16	32	64
heart-c	MeanMap	79.98±1.02	79.02±2.23	78.47±2.59	75.94±2.30	74.47±2.79	76.27±2.92
	InvCal	**81.98±1.05**	**80.04±1.41**	78.15±3.50	75.77±2.77	71.30±3.36	72.98±3.35
	alter-∝SVM	81.85±0.74	79.70±0.17	**78.62±1.65**	74.06±0.48	74.07±2.29	73.52±1.95
	conv-∝SVM	81.37±0.69	78.97±0.86	77.98±1.02	**76.84±1.41**	**77.12±0.87**	**77.13±2.39**
breast-cancer	MeanMap	96.69±0.17	96.72±0.22	96.84±0.29	96.60±0.21	96.67±0.18	96.78±0.09
	InvCal	97.07±0.18	97.10±0.22	97.02±0.18	97.08±0.25	96.51±0.25	96.09±0.66
	alter-∝SVM	**97.19±0.12**	**97.10±0.12**	**97.19±0.12**	**97.23±0.25**	**97.09±0.15**	**97.23±0.36**
	conv-∝SVM	96.84±0.17	97.01±0.08	96.84±0.13	96.99±0.12	96.94±0.34	97.13±0.39
credit-a	MeanMap	85.86±0.81	85.04±0.73	84.96±1.25	83.26±1.52	81.14±3.84	76.65±7.00
	InvCal	**86.26±0.65**	85.62±0.12	85.41±0.44	83.79±0.54	82.21±5.15	76.90±6.65
	alter-∝SVM	86.26±0.71	**86.09±0.63**	**85.88±0.22**	**84.86±2.19**	80.89±3.74	80.75±1.33
	conv-∝SVM	85.80±0.58	85.94±0.34	84.26±0.68	83.65±0.95	**82.39±0.78**	**81.56±0.61**
breast-w	MeanMap	96.42±0.18	96.45±0.27	96.20±0.27	96.14±0.46	94.91±1.02	94.53±1.24
	InvCal	96.85±0.23	96.91±0.13	96.77±0.22	96.75±0.22	96.65±0.29	94.58±1.76
	alter-∝SVM	**96.97±0.07**	**97.00±0.18**	**96.94±0.07**	**96.87±0.15**	**96.88±0.25**	**96.70±0.14**
	conv-∝SVM	96.71±0.10	96.60±0.06	96.57±0.08	96.54±0.19	96.77±0.17	96.66±0.14
a1a	MeanMap	76.16±0.33	75.86±0.28	76.44±1.26	76.48±0.55	**75.95±1.06**	**77.03±1.71**
	InvCal	**82.31±0.09**	81.49±0.49	**81.12±0.88**	**78.67±0.74**	75.53±0.22	74.57±1.05
	alter-∝SVM	82.22±0.41	**81.80±0.68**	79.16±1.51	75.77±0.57	75.73±1.80	75.36±0.71
	conv-∝SVM	76.34±0.61	75.39±0.01	75.39±0.01	75.40±0.02	75.37±0.04	75.37±0.05
dna-3	MeanMap	90.99±0.65	89.45±1.12	88.01±0.65	84.30±1.36	79.59±2.49	73.88±4.89
	InvCal	93.23±0.44	91.83±0.63	89.49±0.52	85.47±1.33	78.26±3.57	70.91±3.00
	alter-∝SVM	**94.36±0.31**	**93.28±0.25**	**92.40±0.35**	**90.04±0.65**	**87.89±1.10**	**86.40±1.26**
	conv-∝SVM	91.75±0.45	87.48±2.02	80.41±0.70	75.91±0.29	75.37±1.66	74.63±0.21
satimage-3	MeanMap	95.67±0.15	95.73±0.25	95.36±0.20	94.65±0.49	92.89±1.95	92.05±1.72
	InvCal	96.66±0.19	96.39±0.26	95.99±0.24	95.32±0.33	95.03±0.27	94.07±0.46
	alter-∝SVM	**96.68±0.32**	**96.54±0.24**	**96.16±0.41**	**95.71±0.28**	**95.16±0.17**	**95.05±0.23**
	conv-∝SVM	95.45±0.13	95.34±0.13	95.38±0.49	94.69±0.68	94.69±0.57	94.14±0.70

Table 2. Additional experiments. Accuracy with RBF kernel, with bag size 2, 4, 8, 16, 32, 64.