A Supplementary Material

A.1 Technical Matter

We use the following inequality, defined for $x \in [0,1]$.

$$\sqrt{1-x} + \sqrt{1+x} \leq 2 \ . \tag{12}$$

From concavity of $\sqrt{1+z}$ if follows $\sqrt{1+z} \leq 1 + \frac{1}{2}z$ and thus we have, $\sqrt{1-x} + \sqrt{1+x} \leq 1 - \frac{1}{2}x + 1 + \frac{1}{2}x = 2$.

A.2 Proof of Thm. 1

Proof: Define

$$\Psi_i(\mathbf{v}) = \frac{1}{2}\left\|\mathbf{v}\right\|^2 + \sum_{j=1}^{i} \frac{Z_jQ_j}{2}\left(y_j - \mathbf{v}^\top\mathbf{x}_j\right)^2 \ ,$$

Thm .3 of Cesa-Bianchi et al [11] states,

$$\begin{aligned}
&\frac{1}{2}Z_iQ_i\left(y_i - \hat{p}_i\right)^2\\
=&\inf_{\mathbf{v}} \Psi_{i+1}(\mathbf{v}) - \inf_{\mathbf{v}} \Psi_i(\mathbf{v})\\
&+ \frac{Z_iQ_i}{2}\mathbf{x}_i^\top A_i^{-1}\mathbf{x}_i - \frac{Z_iQ_i}{2}\left(\mathbf{x}_i^\top A_{i-1}^{-1}\mathbf{x}_i\right)\hat{p}_i^2\\
=&\inf_{\mathbf{v}} \Psi_{i+1}(\mathbf{v}) - \inf_{\mathbf{v}} \Psi_i(\mathbf{v}) + \frac{Z_iQ_i}{2}\frac{r_i}{1+r_i} - \frac{Z_iQ_i}{2}r_i\hat{p}_i^2 \ .
\end{aligned}$$

Summing over i,

$$\begin{aligned}
&\frac{1}{2}\sum_i Z_iQ_i\left(y_i - \hat{p}_i\right)^2\\
\leq&\inf_{\mathbf{v}} \Psi_{m+1}(\mathbf{v}) + \sum_i \frac{Z_iQ_i}{2}\frac{r_i}{1+r_i} - \sum_i \frac{Z_iQ_i}{2}r_i\hat{p}_i^2\\
\leq&\frac{1}{2}\left\|\mathbf{v}\right\|^2 + \sum_{i=1}^{m}\frac{Z_iQ_i}{2}\left(y_i - \mathbf{v}^\top\mathbf{x}_i\right)^2\\
&+ \sum_i \frac{Z_iQ_i}{2}\frac{r_i}{1+r_i} - \sum_i \frac{Z_iQ_i}{2}r_i\hat{p}_i^2 \ .
\end{aligned}$$

Expanding the two square terms, and rearranging we get,

$$\begin{aligned}
&\frac{1}{2}\sum_i Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2\right)\\
\leq&\frac{1}{2}\left\|\mathbf{v}\right\|^2 + \sum_{i=1}^{m}\frac{Z_iQ_i}{2}\mathbf{v}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{v} - \sum_{i=1}^{m} Z_iQ_iy_i\mathbf{v}^\top\mathbf{x}_i\\
=&\frac{1}{2}\mathbf{v}^\top\left(\mathbf{I} + \sum_{i=1}^{m} Z_iQ_i\mathbf{x}_i\mathbf{x}_i^\top\right)\mathbf{v} - \sum_{i=1}^{m} Z_iQ_iy_i\mathbf{v}^\top\mathbf{x}_i\\
=&\frac{1}{2}\mathbf{v}^\top A_{\mathbf{v}}\mathbf{v} - \sum_{i=1}^{m} Z_iQ_iy_i\mathbf{v}^\top\mathbf{x}_i \ ,
\end{aligned} \tag{13}$$

where we used (9) for the last step.

Since $\mathbf{v}$ is arbitrary we can replace it with a scaled version $c\mathbf{v}$. Using a trivial relation $1 - x \leq \max\{1-x, 0\}$ yields,

$$-Z_iQ_icy_i\mathbf{v}^\top\mathbf{x}_i \leq -cZ_iQ_i + cZ_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right) \ . \tag{14}$$

Re-arraigning, and substituting (14) in (13),

$$\begin{aligned}
&\frac{1}{2}\sum_i Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2 + 2c\right)\\
\leq&\frac{1}{2}c^2\mathbf{v}^\top A_{\mathbf{v}}\mathbf{v} + c\sum_i Z_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right) \ .
\end{aligned} \tag{15}$$

We now split the first sum into two alternatives, depending whether an update error was performed $i \in \mathcal{M}$ or an update which is not an error $i \in \mathcal{U}$. We start with the first case of an error $i \in \mathcal{M}$, in which we have, $-y_i\hat{p}_i = |\hat{p}_i|$, and consider two subcases, depending whether the function $\Theta(|\hat{p}_i|, r_i)$ is positive ($i \in \mathcal{S} \cap \mathcal{M}$) or negative ($i \in \mathcal{A} \cap \mathcal{M}$). In the former subcase Q_i is random variable with expectation $\mathbb{E}\left[Q_i\right] = \frac{2c}{2c + \Theta(|\hat{p}_i|, r_i)}$ and thus

$$\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2 + 2c\right)\right] = 2c\mathbb{E}\left[Z_i\right] \ .$$

In the later subcase, $Q_i = 1$ (be definition), and we bound,

$$\begin{aligned}
&\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2 + 2c\right)\right]\\
&\qquad \geq 2c\mathbb{E}\left[Z_i\right] - \frac{r_i}{1+r_i} \ .
\end{aligned}$$

Now we consider examples for which an update (that is not a mistake) was performed, that is $0 \leq y_i\hat{p}_i$, and by definition $i \in \mathcal{U}$. Such cases occur only when $i \in \mathcal{A}$, that is $i \in \mathcal{U} \cap \mathcal{A}$. Updates in this case are performed when the margin is negative or causing an aggressive update (see Fig. 1), thus

$$0 \leq y_i\hat{p}_i \leq \theta(r_i) \leq \frac{1 - \sqrt{1 - r_i}}{1 + r_i} \ ,$$

where the last inequality follows (12). We thus bound,

$$\begin{aligned}
&\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2 + 2c\\
=&(1+r_i)\hat{p}_i^2 - 2y_i\hat{p}_i + \frac{r_i}{1+r_i} - 2\frac{r_i}{1+r_i} + 2c\\
=&f(y_i\hat{p}_i) - 2\frac{r_i}{1+r_i} + 2c
\end{aligned}$$

where $f(y_i\hat{p}_i) = (1+r_i)\hat{p}_i^2 - 2y_i\hat{p}_i + \frac{r_i}{1+r_i}$ is a quadratic equation with two non-negative roots and a minima, $\frac{1 \pm \sqrt{1-r_i}}{1+r_i}$. Thus, if $y_i\hat{p}_i$ is lower than the smaller root, $y_i\hat{p}_i \leq \frac{1-\sqrt{1-r_i}}{1+r_i}$ then $f(y_i\hat{p}_i) \geq 0$, and we bound,

$$\begin{aligned}
&\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + r_i\hat{p}_i^2 + 2c\right)\right]\\
&\qquad \geq 2c\mathbb{E}\left[Z_i\right] - \frac{2r_i}{1+r_i} \ .
\end{aligned}$$

To summarize,

$$\frac{1}{2}\sum_i \mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2+2|\hat{p}_i|-\frac{r_i}{1+r_i}+r_i\hat{p}_i^2+2c\right)\right]$$
$$\geq c\sum_{i\in\mathcal{M}}\mathbb{E}\left[Z_i\right]+c\sum_{i\in\mathcal{U}}\mathbb{E}\left[Z_i\right]$$
$$-\frac{1}{2}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{M}}\frac{r_i}{1+r_i}\right]-\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{U}}\frac{r_i}{1+r_i}\right] \ . \quad (16)$$

Taking the expectation of (15), using (16) to lower bound the left-hand-side concludes the proof, and the definitions, $M=\sum_{i\in\mathcal{M}}Z_i$ and $U=\sum_{i\in\mathcal{U}}Z_i$ we get,

$$\mathbb{E}\left[M\right]\leq\frac{1}{2}c\mathbf{v}^\top\mathbb{E}\left[A_\mathbf{v}\right]\mathbf{v}+\mathbb{E}\left[\sum_i Z_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right)\right]$$
$$+\frac{1}{2c}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{M}}\frac{r_i}{1+r_i}\right]$$
$$+\frac{1}{c}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{U}}\frac{r_i}{1+r_i}\right]-\mathbb{E}\left[U\right]$$
$$\leq\frac{1}{2}c\mathbf{v}^\top\mathbb{E}\left[A_\mathbf{v}\right]\mathbf{v}+\mathbb{E}\left[\sum_i Z_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right)\right]$$
$$+\frac{1}{c}\mathbb{E}\left[\sum_{i\in\mathcal{A}}\frac{r_i}{1+r_i}\right]-\mathbb{E}\left[U\right] \ .$$

■

A.3 Proof of Thm. 2

Proof: Similar the argument in beginning of Thm. 1, [22, Theorem 1] stated,

$$\sum_i Z_iQ_i\left(y_i-\hat{p}_i\right)^2$$
$$=\min_{\mathbf{v}}\left(b\left\|\mathbf{v}\right\|^2+\sum_i Z_iQ_ia_i\left(y_i-\mathbf{v}^\top\mathbf{x}_i\right)^2\right)$$
$$\leq b\left\|\mathbf{v}\right\|^2+\sum_i Z_iQ_ia_i\left(y_i-\mathbf{v}^\top\mathbf{x}_i\right)^2$$

Expanding the two square terms,

$$\sum_i Z_iQ_i\left(y_i^2-2y_i\hat{p}_i+\hat{p}_i^2\right) \quad (17)$$
$$\leq b\left\|\mathbf{v}\right\|^2+\sum_i Z_iQ_ia_i\left(y_i^2-2y_i\mathbf{v}^\top\mathbf{x}_i+\mathbf{v}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{v}\right) \ .$$

Rearranging (17) and using both $y_i^2=1$ and (11),

$$\sum_i Z_iQ_i\left(1-a_i-2y_i\hat{p}_i+\hat{p}_i^2\right)$$
$$\leq b\left\|\mathbf{v}\right\|^2+\sum_i Z_iQ_ia_i\mathbf{v}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{v}-2\sum_i Z_iQ_ia_iy_i\mathbf{v}^\top\mathbf{x}_i$$
$$=\mathbf{v}^\top A_\mathbf{v}^a\mathbf{v}-2\sum_i Z_iQ_ia_iy_i\mathbf{v}^\top\mathbf{x}_i \ . \quad (18)$$

We use again the relation (14). Substituting in (18) together with the bound [22, Eq. 30],

$$1\leq a_i\leq\frac{b}{b-1} \ , \ a_i-1\leq\left(\frac{b}{b-1}\right)^2\frac{r_i}{1+r_i}$$

we obtain,

$$\sum_i Z_iQ_i\left(-\frac{r_i}{1+r_i}-2y_i\hat{p}_i+\hat{p}_i^2+2ca_i\right)$$
$$\leq c^2\mathbf{v}^\top A_\mathbf{v}^a\mathbf{v}+2\sum_i Z_iQ_i\frac{bc}{b-1}\ell\left(y_i\mathbf{x}_i^\top\mathbf{v}\right)$$
$$+\left(\left(\frac{b}{b-1}\right)^2-1\right)\sum_i Z_iQ_i\frac{r_i}{1+r_i} \ . \quad (19)$$

As before, split the first sum into two alternatives, depending whether an update error was performed $i\in\mathcal{M}$ or an update which is not an error $i\in\mathcal{U}$. We start with the first case of an error $i\in\mathcal{M}$, in which we have, $-y_i\hat{p}_i=|\hat{p}_i|$, and consider two subcases, depending whether the function $\Gamma(|\hat{p}_i|,r_i)$ is positive ($i\in\mathcal{S}\cap\mathcal{M}$) or negative ($i\in\mathcal{A}\cap\mathcal{M}$). In the former subcase Q_i is random variable with expectation $\mathbb{E}\left[Q_i\right]=\frac{2ca_i}{2ca_i+\Gamma(|\hat{p}_i|,r_i)}$ and thus,

$$\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2-2y_i\hat{p}_i-\frac{r_i}{1+r_i}+2ca_i\right)\right]=2c\mathbb{E}\left[Z_ia_i\right]$$
$$\geq 2c\mathbb{E}\left[Z_i\right] \ .$$

In the later subcase, $Q_i=1$ (be definition), and we bound,

$$\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2-2y_i\hat{p}_i-\frac{r_i}{1+r_i}+2ca_i\right)\right]$$
$$\geq 2c\mathbb{E}\left[Z_i\right]-\frac{r_i}{1+r_i} \ .$$

Now we consider examples for which an update (that is not a mistake) was performed, that is $0\leq y_i\hat{p}_i$, and by definition $i\in\mathcal{U}$. Such cases occur only when $i\in\mathcal{A}$, that is $i\in\mathcal{U}\cap\mathcal{A}$. Updates in this case are performed when the margin is negative or causing an aggressive update (see Fig. 1),

$$0\leq y_i\hat{p}_i\leq\gamma(r_i)\leq 1-\sqrt{1-\frac{r_i}{1+r_i}} \ ,$$

where the last inequality follows (12). We thus bound,

$$\begin{aligned}&\hat{p}_i^2 - 2|\hat{p}_i| - \frac{r_i}{1+r_i} + 2ca_i\\ =&\hat{p}_i^2 - 2|\hat{p}_i| + \frac{r_i}{1+r_i} - 2\frac{r_i}{1+r_i} + 2ca_i\\ =&f(y_i\hat{p}_i) - 2\frac{r_i}{1+r_i} + 2ca_i \ ,\end{aligned}$$

where $f(y_i\hat{p}_i) = \hat{p}_i^2 - 2|\hat{p}_i| + \frac{r_i}{1+r_i}$ is a quadratic equation with two non-negative roots and a minima, $\frac{1\pm\sqrt{1-r_i}}{1+r_i}$. Thus, if $y_i\hat{p}_i$ is lower than the smaller root, $y_i\hat{p}_i \leq 1 - \sqrt{1-\frac{r_i}{1+r_i}}$ then $f(y_i\hat{p}_i) \geq 0$, and we bound,

$$\begin{aligned}&\mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2 - 2y_i\hat{p}_i - \frac{r_i}{1+r_i} + 2ca_i\right)\right]\\ &\qquad\qquad \geq 2c\mathbb{E}\left[Z_i\right] - \frac{2r_i}{1+r_i} \ .\end{aligned}$$

To summarize,

$$\begin{aligned}&\frac{1}{2}\sum_i \mathbb{E}\left[Z_iQ_i\left(\hat{p}_i^2 + 2|\hat{p}_i| - \frac{r_i}{1+r_i} + 2ca_i\right)\right]\\ \geq&\frac{1}{2}c\sum_{i\in\mathcal{M}}\mathbb{E}\left[Z_i\right] + \frac{1}{2}c\sum_{i\in\mathcal{U}}\mathbb{E}\left[Z_i\right]\\ &-\frac{1}{2}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{M}}\frac{r_i}{1+r_i}\right] - \mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{U}}\frac{r_i}{1+r_i}\right] \ . \quad (20)\end{aligned}$$

Taking the expectation of (19), using (20) to lower bound the left-hand-side concludes the proof, and the definitions, $M = \sum_{i\in\mathcal{M}} Z_i$ and $U = \sum_{i\in\mathcal{U}} Z_i$ we get,

$$\begin{aligned}\mathbb{E}\left[M\right] \leq&\frac{1}{2}c\mathbf{v}^\top\mathbb{E}\left[A_{\mathbf{v}}^a\right]\mathbf{v} + \frac{b}{b-1}\mathbb{E}\left[\sum_i Z_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right)\right]\\ &+\frac{1}{2c}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{M}}\frac{r_i}{1+r_i}\right] + \frac{1}{c}\mathbb{E}\left[\sum_{i\in\mathcal{A}\cap\mathcal{U}}\frac{r_i}{1+r_i}\right]\\ &+\frac{1}{c}\left(\left(\frac{b}{b-1}\right)^2 - 1\right)\sum_{i\in\mathcal{M}\cup\mathcal{U}}\mathbb{E}\left[Z_iQ_i\frac{r_i}{1+r_i}\right] - \mathbb{E}\left[U\right]\\ \leq&\frac{1}{2}c\mathbf{v}^\top\mathbb{E}\left[A_{\mathbf{v}}^a\right]\mathbf{v} + \frac{b}{b-1}\mathbb{E}\left[\sum_i Z_iQ_i\ell\left(y_i\mathbf{v}^\top\mathbf{x}_i\right)\right]\\ &+\frac{1}{c}\left(\left(\frac{b}{b-1}\right)^2 + 1\right)\mathbb{E}\left[\sum_{i\in\mathcal{M}\cup\mathcal{U}}\frac{r_i}{1+r_i}\right] - \mathbb{E}\left[U\right] \ .\end{aligned}$$

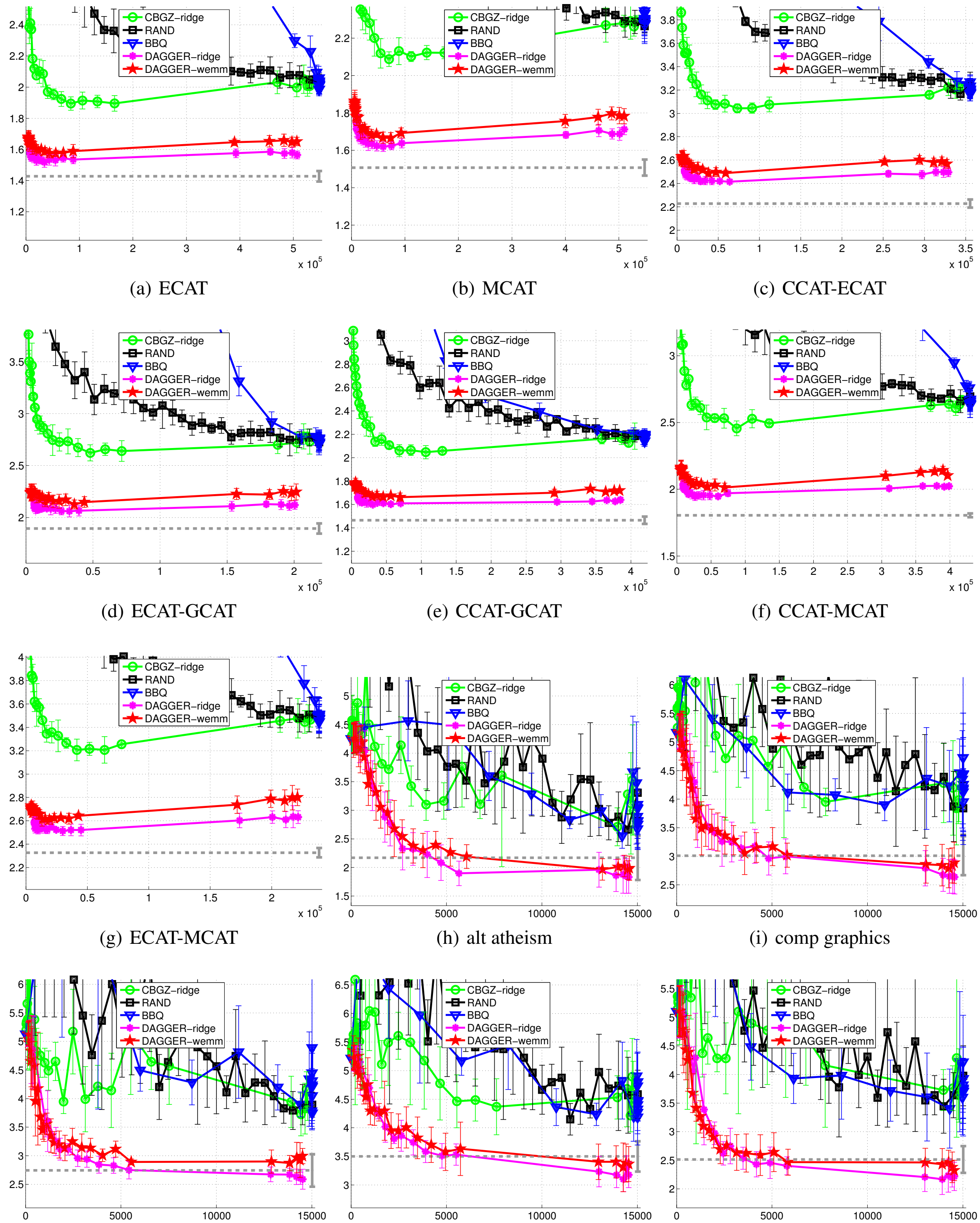
■

Table 4: Average number of queries and test error results for six algorithms for 10 binary classification problems based on the RCV1 dataset.

		RAND		BBQ		CBGZ-ridge		DAGGER-ridge		DAGGER-wemm		AROW	
		queries	error	queries	error	queries	error	queries	error	queries	error	queries	error
queries< 25000	CCAT ECAT	11,809	5.26 (0.30)	14,409	6.98 (0.34)	23,448	3.25 (0.12)	20,831	**2.44** (0.04)	21,014	2.52 (0.03)	355,360	2.23 (0.04)
	CCAT GCAT	14,006	3.69 (0.23)	0	63.54 (0.11)	21,393	2.29 (0.12)	18,741	**1.62** (0.03)	20,279	1.70 (0.03)	419,648	1.47 (0.03)
	CCAT MCAT	14,338	4.38 (0.23)	0	33.57 (0.13)	21,305	2.63 (0.05)	19,262	**1.98** (0.02)	19,411	2.06 (0.04)	429,676	1.80 (0.01)
	ECAT GCAT	14,560	3.95 (0.20)	0	28.16 (0.10)	19,977	2.75 (0.17)	18,518	**2.09** (0.07)	17,030	2.19 (0.06)	218,603	1.89 (0.05)
	ECAT MCAT	15,899	4.89 (0.36)	9,881	10.03 (1.01)	21,039	3.36 (0.10)	20,663	**2.55** (0.04)	19,155	2.61 (0.04)	238,628	2.33 (0.04)
	MCAT GCAT	11,545	2.68 (0.20)	0	46.02 (0.12)	21,313	1.58 (0.03)	17,984	**1.15** (0.02)	20,835	1.23 (0.03)	346,270	1.07 (0.03)
	CCAT	18,271	5.46 (0.13)	3,216	17.52 (2.21)	23,582	3.64 (0.06)	22,926	**2.67** (0.03)	20,791	2.76 (0.01)	548,056	2.37 (0.02)
	ECAT	18,253	3.16 (0.12)	6,708	6.53 (0.21)	19,780	2.07 (0.03)	19,640	**1.53** (0.03)	19,248	1.60 (0.02)	548,056	1.43 (0.03)
	GCAT	18,232	2.42 (0.14)	2,736	6.48 (1.62)	20,340	1.41 (0.03)	19,077	**0.91** (0.02)	18,877	0.97 (0.02)	548,056	0.82 (0.02)
	MCAT	18,369	3.78 (0.23)	8,088	13.62 (1.36)	21,008	2.34 (0.11)	19,062	**1.67** (0.02)	20,701	1.71 (0.05)	548,056	1.51 (0.04)
queries< 50000	CCAT ECAT	35,482	4.29 (0.22)	42,036	5.30 (0.16)	46,529	3.08 (0.05)	42,255	**2.42** (0.04)	38,226	2.49 (0.04)	355,360	2.23 (0.04)
	CCAT GCAT	41,954	3.05 (0.10)	0	63.54 (0.11)	43,187	2.16 (0.06)	36,897	**1.61** (0.01)	36,609	1.68 (0.04)	419,648	1.47 (0.03)
	CCAT MCAT	42,883	3.74 (0.23)	25,460	7.58 (0.76)	43,400	2.54 (0.10)	38,811	**1.96** (0.04)	35,648	2.04 (0.03)	429,676	1.80 (0.01)
	ECAT GCAT	43,751	3.40 (0.22)	39,384	6.31 (0.80)	47,920	2.62 (0.09)	39,539	**2.06** (0.06)	43,612	2.15 (0.06)	218,603	1.89 (0.05)
	ECAT MCAT	39,824	4.37 (0.23)	37,003	7.94 (0.77)	41,509	3.21 (0.11)	36,696	**2.52** (0.06)	35,018	2.62 (0.05)	238,628	2.33 (0.04)
	MCAT GCAT	34,636	2.17 (0.13)	0	46.02 (0.12)	43,120	1.55 (0.04)	40,951	**1.15** (0.03)	37,922	1.22 (0.03)	346,270	1.07 (0.03)
	CCAT	36,703	5.05 (0.19)	3,216	17.52 (2.21)	46,976	3.37 (0.05)	39,700	**2.60** (0.04)	42,266	2.71 (0.01)	548,056	2.37 (0.02)
	ECAT	36,532	2.85 (0.07)	6,708	6.53 (0.21)	40,953	1.97 (0.05)	34,733	**1.52** (0.04)	35,564	1.59 (0.03)	548,056	1.43 (0.03)
	GCAT	36,489	2.38 (0.35)	2,736	6.48 (1.62)	41,383	1.27 (0.04)	40,761	**0.92** (0.02)	34,467	0.98 (0.02)	548,056	0.82 (0.02)
	MCAT	36,542	3.23 (0.24)	8,088	13.62 (1.36)	43,090	2.20 (0.11)	38,523	**1.63** (0.03)	37,315	1.68 (0.04)	548,056	1.51 (0.04)
queries< 100000	CCAT ECAT	82,909	3.79 (0.04)	74,711	4.78 (0.16)	91,057	3.04 (0.05)	63,865	**2.42** (0.02)	59,145	2.49 (0.03)	355,360	2.23 (0.04)
	CCAT GCAT	84,014	2.79 (0.09)	45,729	4.76 (0.43)	85,096	2.06 (0.05)	70,422	**1.61** (0.02)	69,709	1.66 (0.04)	419,648	1.47 (0.03)
	CCAT MCAT	85,732	3.33 (0.15)	58,353	6.10 (0.29)	87,427	2.46 (0.06)	74,771	**1.97** (0.03)	69,960	2.01 (0.04)	429,676	1.80 (0.01)
	ECAT GCAT	65,599	3.19 (0.12)	39,384	6.31 (0.80)	71,451	2.64 (0.11)	39,539	**2.06** (0.06)	43,612	2.15 (0.06)	218,603	1.89 (0.05)
	ECAT MCAT	71,373	3.98 (0.12)	65,837	6.82 (0.51)	77,841	3.25 (0.03)	44,888	**2.52** (0.06)	42,829	2.64 (0.03)	238,628	2.33 (0.04)
	MCAT GCAT	80,748	1.91 (0.06)	81,336	3.49 (0.56)	84,463	1.52 (0.04)	51,039	**1.15** (0.04)	57,702	1.21 (0.04)	346,270	1.07 (0.03)
	CCAT	91,279	4.39 (0.20)	58,233	11.08 (1.39)	95,560	3.27 (0.09)	87,712	**2.56** (0.01)	80,371	2.65 (0.02)	548,056	2.37 (0.02)
	ECAT	72,992	2.81 (0.06)	6,708	6.53 (0.21)	84,122	1.90 (0.05)	69,718	**1.54** (0.03)	70,069	1.58 (0.01)	548,056	1.43 (0.03)
	GCAT	72,986	1.83 (0.09)	64,671	3.28 (0.68)	83,104	1.26 (0.03)	79,874	**0.91** (0.03)	66,697	0.95 (0.01)	548,056	0.82 (0.02)
	MCAT	73,087	3.05 (0.07)	8,088	13.62 (1.36)	88,594	2.13 (0.08)	74,872	**1.62** (0.02)	73,548	1.67 (0.03)	548,056	1.51 (0.04)

Table 5: Average number of queries and test error results for six algorithms for three binary classification problems based on the sentiment dataset.

		RAND		BBQ		CBGZ-ridge		DAGGER-ridge		DAGGER-wemm		AROW	
		queries	error	queries	error	queries	error	queries	error	queries	error	queries	error
queries< 50000	Amazon 4	25,516	**29.59** (1.14)	37,416	29.12 (1.05)	48,347	23.54 (0.53)	-	- (-)	-	- (-)	765,424	19.48 (0.18)
	Amazon 3	25,593	8.62 (0.52)	14,069	11.12 (1.48)	49,632	5.89 (0.22)	-	- (-)	47,202	**4.32** (0.10)	765,424	3.73 (0.04)
	Amazon 1	25,515	7.56 (0.68)	18,598	7.06 (0.52)	43,549	5.35 (0.14)	43,287	**4.37** (0.12)	42,103	4.60 (0.20)	765,424	4.13 (0.05)
queries< 100000	Amazon 4	76,694	28.44 (1.13)	37,416	29.12 (1.05)	94,716	23.50 (0.48)	-	- (-)	93,071	**20.38** (0.26)	765,424	19.48 (0.18)
	Amazon 3	76,304	7.92 (0.34)	96,134	7.95 (0.59)	96,706	5.70 (0.30)	95,753	**3.94** (0.16)	87,776	4.06 (0.16)	765,424	3.73 (0.04)
	Amazon 1	76,554	6.97 (0.23)	18,598	7.06 (0.52)	88,039	5.42 (0.22)	83,840	**4.26** (0.06)	74,582	4.50 (0.04)	765,424	4.13 (0.05)
queries< 150000	Amazon 4	127,472	28.06 (0.97)	101,647	27.73 (0.75)	146,741	23.62 (0.61)	140,657	**19.96** (0.17)	136,259	20.56 (0.26)	765,424	19.48 (0.18)
	Amazon 3	127,799	7.28 (0.45)	96,134	7.95 (0.59)	149,158	5.57 (0.41)	135,749	**3.94** (0.10)	139,086	4.17 (0.10)	765,424	3.73 (0.04)
	Amazon 1	127,434	6.37 (0.38)	116,984	6.31 (0.10)	138,964	5.23 (0.28)	134,827	**4.28** (0.10)	126,939	4.42 (0.15)	765,424	4.13 (0.05)
queries< 200000	Amazon 4	179,052	27.30 (0.53)	101,647	27.73 (0.75)	180,405	24.10 (0.36)	177,854	**19.96** (0.22)	177,543	20.78 (0.15)	765,424	19.48 (0.18)
	Amazon 3	178,764	7.14 (0.53)	96,134	7.95 (0.59)	182,867	5.70 (0.19)	179,273	**3.89** (0.21)	163,613	4.26 (0.15)	765,424	3.73 (0.04)
	Amazon 1	153,328	6.54 (0.36)	116,984	6.31 (0.10)	171,912	5.52 (0.32)	159,887	**4.31** (0.17)	152,422	4.43 (0.18)	765,424	4.13 (0.05)



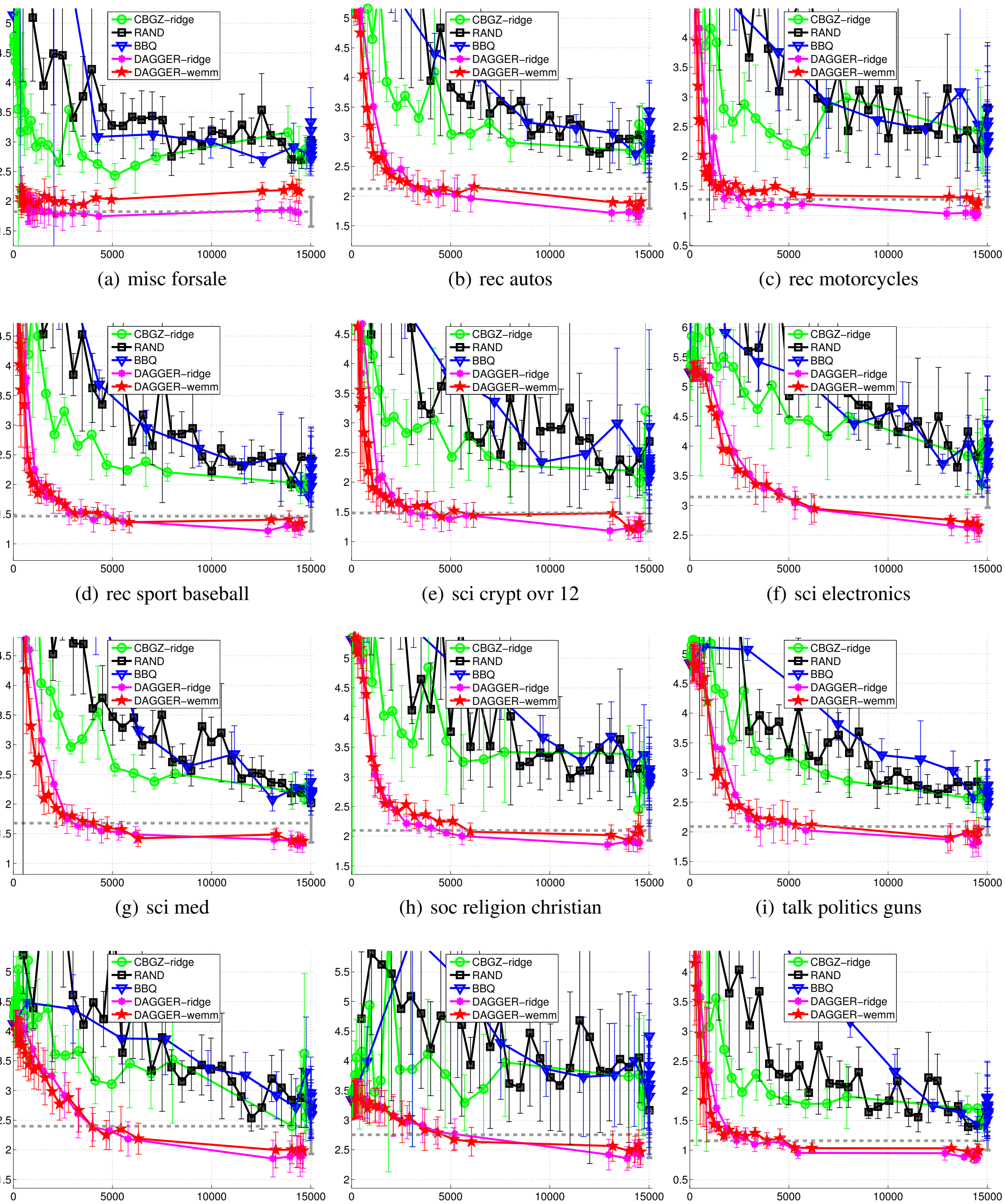
(a) ECAT (b) MCAT (c) CCAT-ECAT

(d) ECAT-GCAT (e) CCAT-GCAT (f) CCAT-MCAT

(g) ECAT-MCAT (h) alt atheism (i) comp graphics

(j) comp os ms-windows misc (k) comp sys ibm pc hardware (l) comp sys mac hardware

Figure 4: More results: error vs no of queries labels for 12 datasets: 1vs-rest RCV (2 datasets), 1vs1 RCV (4 datasets) and 20NG (5 datasets).



(a) misc forsale (b) rec autos (c) rec motorcycles

(d) rec sport baseball (e) sci crypt ovr 12 (f) sci electronics

(g) sci med (h) soc religion christian (i) talk politics guns

(j) talk politics misc (k) talk religion misc (l) talk politics mideast

Figure 5: More results: error vs no of queries labels for 12 datasets of 1vs-rest 20NG 1vs-rest.