Lossless or Quantized Boosting with Integer Arithmetic — Supplementary Material —

Richard Nock
Data61, The Australian National University & The University of Sydney
richard.nock@data61.csiro.au

Robert C. Williamson
The Australian National University & Data61
bob.williamson@anu.edu.au

Abstract

This is the Supplementary Material to Paper "Lossless or Quantized Boosting with Integer Arithmetic", appearing in the proceedings of ICML 2019. Notation "main file" indicates reference to the paper.

1 Table of contents

2 Proof of Theorem 5

2.1 Comments on properness vs the Q-loss

We explain here why we have left open the unit interval for the definition of (2) and why parameter ε in the definition of the partial losses of the Q-loss is important for its properness, *even when* the actual value of ε has absolutely no influence on RATBOOST nor the decision tree induction algorithm using $\underline{L}^Q$. A large class of partial losses is defined in Buja et al. (2005, Theorem 1)[1], from which the following,

$$\ell_1(u) \doteq \int_u^{1-\varepsilon} (1-c)w(\mathrm{d}c), \tag{1}$$

$$\ell_{-1}(u) \doteq \int_\varepsilon^u cw(\mathrm{d}c) \tag{2}$$

defines partial losses of a proper loss, where w is a positive measure require to be finite on any interval $(\varepsilon, 1-\varepsilon)$ with[2] $0 < \varepsilon \leq 1/2$. The definition of proper losses in Reid & Williamson (2010, Theorem 6) implicitly assumes that the integrals are proper so the limits of (1), (2) exist for $\varepsilon \to 0$.

In our case, it is not hard to reconstruct the partial losses of Definition 4 from (1), (2) provided we pick

$$w(\mathrm{d}c) \doteq \frac{\varrho \cdot \mathrm{d}c}{\mathrm{err}(c)^2}, \tag{3}$$

which indeed meets the requirements of Buja et al. (2005, Theorem 1) (see (9) below). So, the Q-loss implicitly constrains the domain of the pointwise Bayes risk to be $(\varepsilon, 1-\varepsilon)$ for it to fit to (1), (2). While this brings the benefit to prevent infinite values for the pointwise Bayes risk ($\lim_0 \underline{L}^Q(u) = \lim_1 \underline{L}^Q(u) = -\infty$), this also does not represent a restriction for learning:

- this restricts *in theory* the image of H_T in RATBOOST to $[\psi(\varepsilon), -\psi(\varepsilon)]$ using the canonical link, that is:

$$\mathrm{Im} H_T \subseteq \varrho \cdot \left(\frac{1}{\varepsilon} - 2\right) \cdot [-1, 1]\,, \tag{4}$$

  but all components of H_T have finite values in RATBOOST (including the images of weak hypotheses, wlog), so we can just consider that ε is implicitly fixed as small as possible for (4) to hold (again, learning H_T in RATBOOST does not depend on ε);

- this restricts *in theory* the proportion p of examples of class ± 1 at each leaf of a decision tree to be in $(\varepsilon, 1-\varepsilon)$ for the tree to be learned with $\underline{L}^Q$, *but* this happens not to be restrictive, for three reasons. First, all classical top-down induction algorithms use losses whose Bayes risk zeroes in 0, 1, so we can train those trees by discarding pure leaves in the computation of $\underline{L}$ (Section 7). Second, discarding pure leaves from the computation of the loss does not endanger the weak learning assumption. Third, in practice DTs are pruned for good generalization: classical statistical methods will in general end up with trees with pure leaves removed Kearns & Mansour (1998).

[1] And an even larger class is defined in Schervish (1989, Theorem 4.2).

[2] Buja et al. (2005, Theorem 1) is slightly more general as the integrals bounds depending on ε are replaced by variables in $(\varepsilon, 1-\varepsilon)$.

2.2 Detailed proof

We use Shuford, Jr et al. (1966, Theorem 1), Reid & Williamson (2010, Theorem 1) to show that the Q-loss is proper. For this to hold, we just need to show that $-u\ell_1^{Q\prime}(u) = (1-u)\ell_{-1}^{Q\ \prime}(u), \forall u \in (0,1)$, where ' denotes derivative. We then check that whenever $u \leq 1/2$, we have $\ell_1^{Q\prime}(u) = \varrho \cdot (-1/u^2 + 1/u)$ and $\ell_{-1}^{Q\ \prime}(u) = \varrho \cdot (1/u)$, so that

$$\begin{aligned} -u\ell_1^{Q\prime}(u) &= \varrho \cdot \left(\frac{1}{u} - 1\right) = \varrho \cdot \left(\frac{1-u}{u}\right); \\ (1-u)\ell_{-1}^{Q\ \prime}(u) &= \varrho \cdot \left(\frac{1-u}{u}\right), \end{aligned} \tag{5}$$

so the Q-loss is proper. To show that it is strictly proper is just a matter of completing three steps: (i) computing the pointwise Bayes risk $\underline{L}^Q$, (ii) computing its weight function $w^Q(u)$ and showing that it is strictly positive for any $u \in [0,1]$ Reid & Williamson (2010, Theorem 6). To achieve step (i), we remark that because ℓ^Q is proper Reid & Williamson (2010),

$$\begin{aligned} &\frac{1}{\varrho} \cdot \underline{L}^Q(u) \\ &= L^Q(u,u) \\ &= u \cdot \ell_1^Q(u) + (1-u) \cdot \ell_{-1}^Q(u) \\ &= \left\{ \begin{array}{cc} -u\log\varepsilon - 2u + 1 + u\log u - (1-u)\log\varepsilon + (1-u)\log u & \text{if} \quad u \leq 1/2 \\ -u\log\varepsilon + u\log(1-u) - (1-u)\log\varepsilon - 2(1-u) + 1 + (1-u)\log(1-u) & \text{otherwise} \end{array} \right. \quad (6) \\ &= -\log\varepsilon + \left\{ \begin{array}{cc} -2u + 1 + \log u & \text{if} \quad u \leq 1/2 \\ -2(1-u) + 1 + \log(1-u) & \text{otherwise} \end{array} \right. \quad (7) \\ &= -\log\varepsilon + \log \mathrm{err}(u) + 1 - 2\mathrm{err}(u) \\ &= \log\left(\frac{\mathrm{err}(u)}{\varepsilon}\right) + 1 - 2\mathrm{err}(u), \quad (8) \end{aligned}$$

and we retrieve (11). We then easily check that its weight function equals Buja et al. (2005)

$$\begin{aligned} w^Q(u) &\doteq -\underline{L}^{Q\prime\prime}(u) \\ &= -\varrho \cdot \left(\left\{ \begin{array}{cc} \frac{1}{u} - 2 & \text{if} \quad u \leq 1/2 \\ -\frac{1}{1-u} + 2 & \text{otherwise} \end{array} \right. \right)' \\ &= \varrho \cdot \left\{ \begin{array}{cc} \frac{1}{u^2} & \text{if} \quad u \leq 1/2 \\ \frac{1}{(1-u)^2} & \text{otherwise} \end{array} \right. \\ &= \frac{\varrho}{\mathrm{err}(u)^2}, \end{aligned} \tag{9}$$

which is indeed > 0 for any $u \in [0,1]$, and shows that the Q-loss is strictly proper. We also remark that $\underline{L}^Q$ is twice differentiable. The computation of the inverse link is then, from (5) (we recall that

$K = 0$),

$$\begin{aligned}
\psi^{Q^{-1}}(z) &\doteq \left(-\underline{L}^{Q'}\right)^{-1}(z) \\
&= \left(\varrho \cdot \left\{ \begin{array}{cc} 2 - \frac{1}{u} & \text{if} \quad u \leq 1/2 \\ -2 + \frac{1}{1-u} & \text{otherwise} \end{array} \right)^{-1} \qquad (10) \\
&= \left\{ \begin{array}{ll} \frac{1}{2 - \frac{z}{\varrho}} & \text{if} \quad z \leq 0 \\ \frac{1 + \frac{z}{\varrho}}{2 + \frac{z}{\varrho}} & \text{otherwise} \end{array} \right. \\
&= \frac{\varrho + \mathrm{H}(-z)}{2\varrho + |z|}, \qquad (11)
\end{aligned}$$

as claimed (link immediate from (10)). The convex surrogate for the Q-loss is obtained from (7), and we first search for $(-\underline{L})^\star$:

$$\begin{aligned}
&(-\underline{L}^Q)^\star(z) \\
&\doteq \sup_{z' \in \mathrm{dom}(\underline{L}^Q)} \{zz' + \underline{L}^Q(z')\} \\
&= \sup_{u \in [0,1]} \left\{ zu + \varrho \cdot \left(\log\left(\frac{\mathrm{err}(u)}{\varepsilon} \right) + 1 - 2\mathrm{err}(u) \right) \right\} \\
&= \varrho \cdot (1 - \log \varepsilon) + \max \left\{ \sup_{u \in [0,1/2]} \{(z - 2\varrho)u + \varrho \cdot \log u\}, -2\varrho + \sup_{u \in (1/2,1]} \{(z + 2\varrho)u + \varrho \cdot \log(1-u)\} \right\} \\
&= \varrho \cdot (1 - \log \varepsilon) + \max \left\{ \begin{array}{ll} \varrho \log \varrho + \varrho \cdot \frac{z - 2\varrho}{2\varrho - z} - \varrho \cdot \log(2\varrho - z) & \text{for } u = \varrho \cdot \frac{1}{2\varrho - z} \in [0, 1/2] \\ \varrho \log \varrho - 2\varrho + \frac{(z+\varrho)(z+2\varrho)}{z+2\varrho} - \varrho \cdot \log(2\varrho + z) & \text{for } u = \frac{z+\varrho}{z+2\varrho} \in (1/2, 1] \end{array} \right. \\
&= \varrho \log \varrho - \varrho \cdot \log \varepsilon + \max \left\{ \begin{array}{ll} -\varrho \cdot \log(2\varrho - z) & \text{for } u = \varrho \cdot \frac{1}{2\varrho - z} \in [0, 1/2] \\ z - \varrho \cdot \log(2\varrho + z) & \text{for } u = \frac{z+\varrho}{z+2\varrho} \in (1/2, 1] \end{array} \right. \\
&= -\varrho \log\left(\frac{\varepsilon}{\varrho}\right) + \max \left\{ \begin{array}{ll} -\varrho \cdot \log(2\varrho - z) & \text{for } z \leq 0 \\ z - \varrho \cdot \log(2\varrho + z) & \text{for } z > 0 \end{array} \right. \\
&= -\varrho \log\left(\frac{\varepsilon}{\varrho}\right) + \left\{ \begin{array}{ll} -\varrho \cdot \log(2\varrho - z) & \text{for } z \leq 0 \\ z - \varrho \cdot \log(2\varrho + z) & \text{for } z > 0 \end{array} \right. \\
&= -\varrho \cdot \log\left(2\varepsilon + \frac{\varepsilon|z|}{\varrho}\right) + \mathrm{H}(-z), \qquad (12)
\end{aligned}$$

and we get

$$\begin{aligned}
F^Q(z) &\doteq (-\underline{L}^Q)^\star(-z) \qquad (13) \\
&= -\varrho \cdot \log\left(2\varepsilon + \frac{\varepsilon|z|}{\varrho}\right) + \mathrm{H}(z), \qquad (14)
\end{aligned}$$

as claimed. This derivation also allows us to prove that the Q-loss is proper canonical using Nock & Nielsen (2008, Lemma 1). That the Q-loss is symmetric is just a consequence of its definition Reid & Williamson (2010). This ends the proof of Theorem 5.

3 Proof of Lemma 6

Denote for short

$$v \doteq z + \varrho \cdot \left(\frac{1-2u}{\mathrm{err}(u)} \right). \tag{15}$$

It is not hard to check that indeed

$$z \diamond u = \frac{\varrho + H(v)}{2\varrho + |v|} \doteq g(v), \tag{16}$$

as well as $g(-v) = 1 - g(v)$. So, we focus on the second equality. Denote for short $u \doteq n_u/d_u$, $z \doteq \varrho \cdot n_z/d_z$. We remark that the definition of z makes ϱ simplify:

$$\begin{aligned} z \diamond u &= \frac{1 + H\left(\frac{n_z}{d_z} + \frac{1-2\cdot\frac{n_u}{d_u}}{\frac{n_u}{d_u} \wedge \frac{d_u - n_u}{d_u}} \right)}{2 + \left| \frac{n_z}{d_z} + \frac{1-2\cdot\frac{n_u}{d_u}}{\frac{n_u}{d_u} \wedge \frac{d_u - n_u}{d_u}} \right|} \\ &= \frac{1 + H\left(\frac{n_z}{d_z} + \frac{d_u - 2n_u}{n_u \wedge (d_u - n_u)} \right)}{2 + \left| \frac{n_z}{d_z} + \frac{d_u - 2n_u}{n_u \wedge (d_u - n_u)} \right|} \end{aligned} \tag{17}$$

Case 1: $v \geq 0$ and $n_u \leq d_u - n_u$. We have

$$\begin{aligned} z \diamond u &= \frac{1}{2 + \frac{n_z}{d_z} + \frac{d_u - 2n_u}{n_u}} \\ &= \frac{1}{\frac{n_z}{d_z} + \frac{d_u}{n_u}} = \frac{n_u d_z}{n_u n_z + d_u d_z}. \end{aligned} \tag{18}$$

Case 2: $v \geq 0$ and $n_u > d_u - n_u$. We have

$$\begin{aligned} z \diamond u &= \frac{1}{2 + \frac{n_z}{d_z} + \frac{d_u - 2n_u}{d_u - n_u}} \\ &= \frac{1}{3 + \frac{n_z}{d_z} - \frac{n_u}{d_u - n_u}} \\ &= \frac{(d_u - n_u) d_z}{(d_u - n_u)(3d_z + n_z) - n_u d_z} \\ &= \frac{(d_u - n_u) d_z}{(d_u - n_u) n_z + d_u d_z + 2(d_u - 2n_u) d_z}. \end{aligned} \tag{19}$$

Folding both cases 1 and 2, we get

$$z \diamond u = \frac{(n_u \wedge (d_u - n_u)) d_z}{(n_u \wedge (d_u - n_u)) n_z + d_u d_z - 2\mathrm{H}(d_u - 2n_u) d_z}. \tag{20}$$

Note that this holds when $v \geq 0$, equivalent to

$$\frac{n_z}{d_z} + \frac{d_u - 2n_u}{n_u \wedge (d_u - n_u)} > 0, \tag{21}$$

that is, assuming wlog $d_z > 0$,

$$(n_u \wedge (d_u - n_u))n_z \;\; > \;\; -(d_u - 2n_u)d_z. \tag{22}$$

So, let us denote $a \doteq (n_u \wedge (d_u - n_u))d_z, b \doteq (n_u \wedge (d_u - n_u))n_z, c \doteq d_u d_z, d \doteq 2(d_u - 2n_u)d_z$. We get that if $b + (d/2) \geq 0$, then

$$z \diamond u \;\; = \;\; \frac{a}{b + c - \mathrm{H}(d)}, \tag{23}$$

and if $b + (d/2) < 0$, then we remark that $-b - (d/2) > 0$, so

$$z \diamond u \;\; = \;\; 1 - \frac{a}{-b + c - \mathrm{H}(-d)} = \frac{-b - a + c - \mathrm{H}(-d)}{-b + c - \mathrm{H}(-d)}, \tag{24}$$

as claimed.

4 Proof of Theorem 7

The proof revolves on two simple facts about F^Q: (i) since F^Q is convex and differentiable, we have $F^Q(y) - F^Q(x) - (y - x)F^{Q'}(x) \geq 0$ (the left hand side is just the Bregman divergence with generator F^Q). Also, (ii) F^Q being twice differentiable, Taylor Theorem says that for any x, y we can expand the derivative as $F^{Q'}(y) = F^{Q'}(x) + (y - x)F^{Q''}(z)$ for some $z \in [x, y]$. Using (i) and (ii) in this order, we get that fo for any $i \in \{1, 2, ..., m\}$, there exists $\alpha_i \in [0, 1]$ and

$$\beta_i \;\; = \;\; y_i H_t(\boldsymbol{x}_i) + \alpha_i \delta_t y_i h_t(\boldsymbol{x}_i) \;\; \in \;\; [y_i H_t(\boldsymbol{x}_i), y_i H_{t+1}(\boldsymbol{x}_i)] \tag{25}$$

such that:

$$\begin{aligned}
&\mathbb{E}_{i\sim D}\left[F^Q(y_i H_t(\boldsymbol{x}_i))\right] - \mathbb{E}_{i\sim D}\left[F^Q(y_i H_{t+1}(\boldsymbol{x}_i))\right] \\
&\geq \;\; \mathbb{E}_{i\sim D}\left[(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))F^{Q'}(y_i H_{t+1}(\boldsymbol{x}_i))\right] \qquad (26)\\
&= \underbrace{\mathbb{E}_{i\sim D}\left[(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))F^{Q'}(y_i H_t(\boldsymbol{x}_i))\right]}_{\doteq X} - \underbrace{\mathbb{E}_{i\sim D}\left[(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))^2 F^{Q''}(\beta_i)\right]}_{\doteq Y} \qquad (27)
\end{aligned}$$

Because F^Q is convex, $Y \geq 0$. We want to show that not just $X \geq 0$ but in fact the difference $X - Y$ is sufficiently large for the bound of the Theorem to hold. We first remark

$$\begin{aligned}
X \;\; &\doteq \;\; \mathbb{E}_{i\sim D}\left[(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))F^{Q'}(y_i H_t(\boldsymbol{x}_i))\right] \\
&= \;\; -\delta_t \mathbb{E}_{i\sim D}\left[y_i h_t(\boldsymbol{x}_i) \cdot -\psi^{Q^{-1}}(-y_i H_t(\boldsymbol{x}_i))\right] \\
&= \;\; \delta_t \mathbb{E}_{i\sim D}\left[w_{ti} y_i h_t(\boldsymbol{x}_i)\right] \\
&= \;\; \delta_t \cdot \frac{\sum_i w_{ti} y_i h_t(\boldsymbol{x}_i)}{m} \\
&= \;\; a \cdot \eta_t^2. \qquad (28)
\end{aligned}$$

We also have $F^{Q\prime\prime}(z) = \varrho/(2\varrho + |z|)^2$, so

$$\begin{aligned} Y &\doteq \mathbb{E}_{i\sim D}\left[(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))^2 F^{Q\prime\prime}(\beta_i)\right] \\ &= \varrho \cdot \mathbb{E}_{i\sim D}\left[\frac{(y_i H_t(\boldsymbol{x}_i) - y_i H_{t+1}(\boldsymbol{x}_i))^2}{(2+|\beta_i|)^2}\right] \\ &= \varrho\delta_t^2 \cdot \mathbb{E}_{i\sim D}\left[\frac{h_t^2(\boldsymbol{x}_i)}{(2\varrho+|\beta_i|)^2}\right]. \end{aligned} \tag{29}$$

Now we get because of assumption (M):

$$\begin{aligned} \mathbb{E}_{i\sim D}\left[\frac{h_t^2(\boldsymbol{x}_i)}{(2\varrho+|\beta_i|)^2}\right] &\leq \frac{1}{4\varrho^2} \cdot \mathbb{E}_{i\sim D}\left[h_t^2(\boldsymbol{x}_i)\right] \\ &\leq \frac{M^2}{4\varrho^2}. \end{aligned} \tag{30}$$

So,

$$\begin{aligned} Y &\leq \frac{\delta_t^2 M^2}{4\varrho} \\ &= \frac{a^2 \cdot \eta_t^2 M^2}{4\varrho}. \end{aligned} \tag{31}$$

We finally get

$$\begin{aligned} \mathbb{E}_{i\sim D}\left[F^Q(y_i H_t(\boldsymbol{x}_i))\right] - \mathbb{E}_{i\sim D}\left[F^Q(y_i H_{t+1}(\boldsymbol{x}_i))\right] &\geq X - Y \\ &\geq \underbrace{\left(1 - \frac{aM^2}{4\varrho}\right) \cdot a}_{\doteq Z(a)} \cdot \eta_t^2. \end{aligned} \tag{32}$$

Suppose now that we fix any $\pi \in [0,1]$ and then choose *any*

$$a \in \frac{2\varrho}{M^2} \cdot [1-\pi, 1+\pi]. \tag{33}$$

It is not hard to check that $Z(a)$ satisfies

$$Z(a) \geq (1-\pi^2) \cdot \frac{\varrho}{M^2} \cdot \eta_t^2, \tag{34}$$

so we get

$$\mathbb{E}_{i\sim D}\left[F^Q(y_i H_t(\boldsymbol{x}_i))\right] - \mathbb{E}_{i\sim D}\left[F^Q(y_i H_{t+1}(\boldsymbol{x}_i))\right] \geq \frac{(1-\pi^2)\varrho\eta_t^2}{M^2}, \forall t, \tag{35}$$

and so the final classifier H_T satisfies

$$\mathbb{E}_{i\sim D}\left[F^Q(y_i H_T(\boldsymbol{x}_i))\right] \leq F^Q(0) - \frac{(1-\pi^2)\varrho \cdot \sum_{t=1}^T \eta_t^2}{M^2}. \tag{36}$$

Remark that this holds regardless of the sequence $\{\eta_t\}_t$. If we want to guarantee that $\mathbb{E}_{i\sim D}\left[F^Q(y_i H_T(\boldsymbol{x}_i))\right] \leq F^Q(z^*)$ for some $z^* \geq 0$, then it suffices to iterate until

$$\sum_{t=1}^T \eta_t^2 \geq \frac{F^Q(0) - F^Q(z^*)}{(1-\pi^2)\varrho} \cdot M^2, \tag{37}$$

and we get the statement of the Theorem.

5 Proof of Theorem 8

The proof uses the same basic steps as the proof of Theorem 7. Denote for short

$$\tilde{w}_{ti} \doteq w_{ti} + \kappa_{ti}, \tag{38}$$

where $w_{ti} \doteq -\psi^{Q^{-1}}\left(-y_i H_t(\boldsymbol{x}_i)\right)$ is the non-quantized weights and κ_{ti} is the quantization shift in weights. Note that we do *not* have access to w_{ti}. We indicate with a tilda quantities that depend on $\tilde{w}$.

This time, we have for X the expression:

$$\begin{aligned} X &= -\tilde{\delta}_t \mathbb{E}_{i\sim D}\left[y_i h_t(\boldsymbol{x}_i) \cdot -\psi^{Q^{-1}}\left(-y_i H_t(\boldsymbol{x}_i)\right)\right] \\ &= \tilde{\delta}_t \mathbb{E}_{i\sim D}\left[w_{ti} y_i h_t(\boldsymbol{x}_i)\right] \\ &= \tilde{\delta}_t \cdot \left(\frac{\sum_i \tilde{w}_{ti} y_i h_t(\boldsymbol{x}_i)}{m} - \frac{\sum_i \kappa_{ti} y_i h_t(\boldsymbol{x}_i)}{m}\right) \\ &= a \cdot \tilde{\eta}_t^2 - a \cdot \tilde{\eta}_t \cdot \frac{\sum_i \kappa_{ti} y_i h_t(\boldsymbol{x}_i)}{m}, \end{aligned} \tag{39}$$

while the expression of Y does not change (yet including "tilda" parameters affected by the quantization of weights). Denote for short

$$\Delta_t \doteq \frac{\sum_i \kappa_{ti} y_i h_t(\boldsymbol{x}_i)}{m}. \tag{40}$$

We get in lieu of (32),

$$\begin{aligned} \mathbb{E}_{i\sim D}\left[F^Q(y_i H_t(\boldsymbol{x}_i))\right] - \mathbb{E}_{i\sim D}\left[F^Q(y_i H_{t+1}(\boldsymbol{x}_i))\right] &\geq X - Y \\ &\geq \left(1 - \frac{\Delta_t}{\tilde{\eta}_t} - \frac{aM^2}{4\varrho}\right) \cdot a\tilde{\eta}_t^2 \\ &= \underbrace{\left(\frac{4\varrho}{M^2} \cdot \frac{\tilde{\eta}_t - \Delta_t}{\tilde{\eta}_t} - a\right) \cdot a}_{\doteq Z(a)} \cdot \frac{M^2 \tilde{\eta}_t^2}{4\varrho} \end{aligned} \tag{41}$$

Choose

$$a \in \frac{2\varrho}{M^2} \cdot \left[\frac{\tilde{\eta}_t - \Delta_t}{\tilde{\eta}_t} - \pi, \frac{\tilde{\eta}_t - \Delta_t}{\tilde{\eta}_t} + \pi\right], \tag{42}$$

for any $0 \leq \pi \leq |\tilde{\eta}_t - \Delta_t|/\tilde{\eta}_t$. It follows

$$Z(a) \geq \left(\left(\frac{\tilde{\eta}_t - \Delta_t}{\tilde{\eta}_t}\right)^2 - \pi^2\right) \cdot \frac{\varrho}{M^2} \cdot \tilde{\eta}_t^2. \tag{43}$$

Suppose that the quantisation shift satisfies $|\tilde{\eta}_t - \Delta_t| \geq \zeta \cdot |\tilde{\eta}_t|$ (which holds if $|\Delta_t| \leq (1-\zeta) \cdot |\tilde{\eta}_t|$) for some $\zeta > 0$. We obtain that for any $0 \leq \pi < \zeta$,

$$Z(a) \geq \left(\zeta^2 - \pi^2\right) \cdot \frac{\varrho}{M^2} \cdot \tilde{\eta}_t^2 > 0, \tag{44}$$

which leads to the statement of the Theorem after posing $\kappa_t \doteq |\Delta_t|$.
Remark: assumption (Q) is in fact stronger than what would really be needed to get the Theorem. Under some conditions, we could indeed accept $|\Delta_t| > (1-\zeta) \cdot |\tilde{\eta}_t|$, but in the derivations above, the shift in weights due to quantisation would result in a disguised way to strenghten weak learning. Clearly, such an assumption where quantisation compensates for the weakness of the weak classifiers is unfit in a boosting setting.

6 Proof of Theorem 10

We assume basic knowledge of the proofs of Kearns & Mansour (1996). We shall briefly present the proof scheme as well as the notations, that we keep identical to Kearns & Mansour (1996) for readability.

The basic of the proof is to show that each time a leaf is replaced by a split under the weak learning assumption, there is a sufficient decrease of $\underline{L}(H)$. Denote H^+ tree H in which a leaf λ has been replaced by a split indexed with some $g : \mathbb{R} \to \{0, 1\}$ satisfying the weak learning assumption. The decrease in $\underline{L}(.)$, $\Delta \doteq \underline{L}(H) - \underline{L}(H^+)$, is lowerbounded as a function of γ and then used to lowerbound the number of iterations (each of which is the replacement of a leaf by a binary subtree) to get to a given value of $\underline{L}(.)$

It turns out that Δ can be abstracted by a better quantity to analyze, $\Delta \doteq \omega(\lambda) \cdot \Delta_{\underline{L}^Q}(q, \tau, \delta)$ with

$$\Delta_{\underline{L}^Q}(q, \tau, \delta) \doteq \underline{L}^Q(q) - (1-\tau)\underline{L}^Q(q - \tau\delta) - \tau\underline{L}^Q(q + (1-\tau)\delta) \tag{45}$$

with $q \doteq q(\lambda)$ and $\delta = \gamma q(1-q)/(\tau(1-\tau))$ with τ denoting the *relative* proportion of examples for which $g = +1$ in leaf λ, following Kearns & Mansour (1996). The following Lemma is the key to the proof of Theorem 10.

Lemma 1 *Suppose the weak hypothesis assumption is satisfied for the current split, for some constant $\gamma > 0$. For any $q, \tau \in [0, 1]$, using $\delta = \gamma q(1-q)/(\tau(1-\tau))$ yields:*

$$\Delta_{\underline{L}^Q}(q, \tau, \delta) \geq \frac{\gamma^2}{2}. \tag{46}$$

Proof Our proof follows the proof of Kearns & Mansour (1996).

Lemma 2 *Suppose $\tau \leq 1/2, q > 1/2$ or $\tau \geq 1/2, q < 1/2$. If $\gamma \leq 1/25$, $\Delta_{\underline{L}^Q}(q, \tau, \delta)$ is minimized by some $\tau \in [0.4, 0.6]$.*

Proof To prove the Lemma we use the trick of Kearns & Mansour (1996, Lemma 4), which consists of studying function

$$\begin{aligned} U(q, X) &\doteq \underline{L}^Q(q - X) + X\underline{L}^{Q\prime}(q - X) \\ &= \begin{cases} \log(q - X) + \frac{X}{q-X} + 1 - 2q & \text{if} \quad q - X \leq \frac{1}{2} \\ \log(1 - q + X) - \frac{X}{1-q+X} - 1 + 2q & \text{if} \quad q - X > \frac{1}{2} \end{cases} \end{aligned} \tag{47}$$

and show

$$U(q,\tau\delta) \leq U(q,-(1-\tau)\delta), \forall \tau \leq 0.4, \quad (48)$$
$$U(q,\tau\delta) \geq U(q,-(1-\tau)\delta), \forall \tau \geq 0.6, \quad (49)$$

Case 1: $\tau \leq 0.4$ (and therefore $q < 1/2$). We have two subcases to show (48).

Case 1.1: $q + (1-\tau)\delta < 1/2$. In this case, $q - X < 1/2$ for both instantiations of X in (48). We then have

$$U(q,\tau\delta) = \log\left(1 - \frac{\gamma(1-q)}{1-\tau}\right) + \frac{\frac{\gamma(1-q)}{1-\tau}}{1-\frac{\gamma(1-q)}{1-\tau}} + 1 - 2q + \log q \quad (50)$$
$$= \log\left(\frac{\tau - 1 + \gamma(1-q)}{\tau - 1}\right) - \frac{\gamma(1-q)}{\tau - 1 + \gamma(1-q)} + 1 - 2q + \log q \quad (51)$$
$$U(q,-(1-\tau)\delta) = \log\left(1 + \frac{\gamma(1-q)}{\tau}\right) - \frac{\frac{\gamma(1-q)}{\tau}}{1+\frac{\gamma(1-q)}{\tau}} + 1 - 2q + \log q \quad (52)$$
$$= \log\left(\frac{\tau + \gamma(1-q)}{\tau}\right) - \frac{\gamma(1-q)}{\tau + \gamma(1-q)} + 1 - 2q + \log q, \quad (53)$$

so (48) is equivalent to showing

$$\log\left(\frac{\tau - 1 + \gamma(1-q)}{\tau - 1}\right) - \frac{\gamma(1-q)}{\tau - 1 + \gamma(1-q)} \leq \log\left(\frac{\tau + \gamma(1-q)}{\tau}\right) - \frac{\gamma(1-q)}{\tau + \gamma(1-q)} \quad (54)$$

which after reorganising and simplification amounts to showing

$$\log\left(1 - \frac{\gamma(1-q)}{(\tau + \gamma(1-q))(1-\tau)}\right) \leq -\frac{\gamma(1-q)}{(\tau + \gamma(1-q))(1-\tau-\gamma(1-q))}. \quad (55)$$

We remark that for the $\log$ to be defined in (51), we must have $\tau < 1 - \gamma(1-q)$, which implies that the RHS of (55) is negative. To show (55), we use the fact that $\log(1-X) \leq -X - X^2/2$ when $X \geq 0$, so fixing $X \doteq \gamma(1-q)/((\tau + \gamma(1-q))(1-\tau))$ we obtain

$$\log\left(1 - \frac{\gamma(1-q)}{(\tau + \gamma(1-q))(1-\tau)}\right) \leq -\frac{\gamma(1-q)}{\tau + \gamma(1-q)} \cdot \left(\frac{1}{1-\tau} + \frac{\gamma(1-q)}{2(\tau + \gamma(1-q))(1-\tau)^2}\right) \quad (56)$$

To show (55), we can then show

$$\frac{1}{1-\tau-\gamma(1-q)} \leq \frac{1}{1-\tau} + \frac{\gamma(1-q)}{2(\tau + \gamma(1-q))(1-\tau)^2}, \quad (57)$$

which, after simplification, is equivalent to

$$\frac{1-\tau-\gamma(1-q)}{2(\tau + \gamma(1-q))(1-\tau)} \geq 1, \quad (58)$$

or equivalently $3\tau - 2\tau^2 + 3\gamma(1-q) - 2\tau\gamma(1-q) \leq 1$. Since $\tau \leq 2/5$, $3\tau - 2\tau^2 \leq 22/25$. If we pick $\gamma \leq 1/25$, then $3\gamma(1-q) - 2\tau\gamma(1-q) \leq 3\gamma(1-q) \leq 3\gamma = 3/25$, so that

$3\tau - 2\tau^2 + 3\gamma(1-q) - 2\tau\gamma(1-q) \leq 1$, as claimed (end of Case 1.1).

Case 1.2: $q + (1-\tau)\delta > 1/2$. In this case,

$$U(q, -(1-\tau)\delta) = \log\left(1 - \frac{\gamma q}{\tau}\right) + \frac{\frac{\gamma q}{\tau}}{1 - \frac{\gamma q}{\tau}} + 1 - 2(1-q) + \log(1-q) \quad (59)$$

$$= \log\left(\frac{\tau - \gamma q}{\tau}\right) + \frac{\gamma q}{\tau - \gamma q} + 2q - 1 + \log(1-q). \quad (60)$$

We also remark that $1 - 2q + \log q \leq 2q - 1 + \log(1-q)$ for $q < 1/2$, so to prove (48), it is sufficient to show

$$\log\left(\frac{\tau - 1 + \gamma(1-q)}{\tau - 1}\right) - \frac{\gamma(1-q)}{\tau - 1 + \gamma(1-q)} \leq \log\left(\frac{\tau - \gamma q}{\tau}\right) + \frac{\gamma q}{\tau - \gamma q}, \quad (61)$$

which reduces after simplification to showing that

$$\log\left(1 + \frac{\gamma(q-\tau)}{(\tau - \gamma q)(1-\tau)}\right) \leq \frac{\gamma(q-\tau)}{(\tau - \gamma q)(1 - \tau - \gamma(1-q))}. \quad (62)$$

Because $q + (1-\tau)\delta > 1/2$, if $\tau \geq 10\gamma q(1-q)$, then $q > 0.4$ and therefore $q > \tau$. If, on the other hand $\tau \leq 10\gamma q(1-q)$, then if $\gamma \leq 1/10$, it follows also $\tau \leq q$. To summarize, $q + (1-\tau)\delta > 1/2$ and $\gamma \leq 1/10$ imply $q \geq \tau$.

Using the fact that $\log(1+X) \leq X$ and $\gamma(1-q) \geq 0$, we easily obtain the proof of (62) via the chain of inequalities

$$\log\left(1 + \frac{\gamma(q-\tau)}{(\tau - \gamma q)(1-\tau)}\right) \leq \frac{\gamma(q-\tau)}{(\tau - \gamma q)(1-\tau)} \leq \frac{\gamma(q-\tau)}{(\tau - \gamma q)(1 - \tau - \gamma(1-q))}. \quad (63)$$

This ends up the proof for Case 1.

Case 2: $\tau \geq 0.6$ (and therefore $q > 1/2$). We have two cases again, this time to show (49).

Case 2.1: $q - \tau\delta > 1/2$. In this case, $q - X > 1/2$ for both instantiations of X in (49). We then have

$$U(q, \tau\delta) = \log\left(1 + \frac{\gamma q}{1-\tau}\right) - \frac{\gamma q}{1 - \tau + \gamma q} - 1 + 2q + \log(1-q) \quad (64)$$

$$U(q, -(1-\tau)\delta) = \log\left(1 - \frac{\gamma q}{\tau}\right) + \frac{\gamma q}{\tau - \gamma q} - 1 + 2q + \log(1-q), \quad (65)$$

To show (49), it is thus sufficient to show that

$$\log\left(1 + \frac{\gamma q}{1-\tau}\right) - \frac{\gamma q}{1 - \tau + \gamma q} \geq \log\left(1 - \frac{\gamma q}{\tau}\right) + \frac{\gamma q}{\tau - \gamma q}, \quad (66)$$

or equivalently, after reordering and simplifying,

$$\log\left(1 - \frac{\gamma q}{\tau(1 - \tau + \gamma q)}\right) \leq -\frac{\gamma q}{(\tau - \gamma q)(1 - \tau + \gamma q)}, \quad (67)$$

which is (55) with the substitution $\tau \mapsto 1-\tau$ and $q \mapsto 1-q$. Since then $1-\tau \leq 0.4$, we can directly apply the proof of (55), which ends the proof of Case 2.1.

Case 2.2: $q - \tau\delta < 1/2$. In this case,

$$U(q,\tau\delta) \;=\; \log\left(1-\frac{\gamma(1-q)}{1-\tau}\right)+\frac{\gamma(1-q)}{1-\tau-\gamma(1-q)}+1-2q+\log q, \tag{68}$$

while we still have

$$U(q,-(1-\tau)\delta) \;=\; \log\left(1-\frac{\gamma q}{\tau}\right)+\frac{\gamma q}{\tau-\gamma q}-1+2q+\log(1-q), \tag{69}$$

and so we want to show

$$\begin{aligned}&\log\left(1-\frac{\gamma q}{\tau}\right)+\frac{\gamma q}{\tau-\gamma q}-1+2q+\log(1-q)\\ &\leq\; \log\left(1-\frac{\gamma(1-q)}{1-\tau}\right)+\frac{\gamma(1-q)}{1-\tau-\gamma(1-q)}+1-2q+\log q,\end{aligned} \tag{70}$$

We also remark that $-1+2q+\log(1-q) \leq 1-2q+\log q$ for $q > 1/2$, so to prove (70), it is sufficient to show

$$\log\left(1-\frac{\gamma q}{\tau}\right)+\frac{\gamma q}{\tau-\gamma q} \;\leq\; \log\left(1-\frac{\gamma(1-q)}{1-\tau}\right)+\frac{\gamma(1-q)}{1-\tau-\gamma(1-q)}, \tag{71}$$

which reduces after simplification to showing that

$$\log\left(1+\frac{\gamma(\tau-q)}{(1-\tau-\gamma(1-q))\tau}\right) \;\leq\; \frac{\gamma(\tau-q)}{(\tau-\gamma q)(1-\tau-\gamma(1-q))}, \tag{72}$$

wich turns out to be (62) with the substitution $\tau \mapsto 1-\tau$ and $q \mapsto 1-q$. Since then $1-\tau \leq 0.4$, we can directly apply the proof of (62), which ends the proof of Case 2.2, and the proof of Lemma 2 as well. (end of the proof of Lemma 2) ■

Following Kearns & Mansour (1996), we define

$$F_{\underline{L}^Q}(q,\tau,\delta) \;\doteq\; -\frac{\tau(1-\tau)\delta^2}{2}\underline{L}^{Q\prime\prime}(q)-\frac{\tau(1-\tau)(1-2\tau)\delta^3}{6}\underline{L}^{Q(3)}(q). \tag{73}$$

We now state and prove the equivalent of (Kearns & Mansour, 1996, Lemma 3).

Lemma 3 *For any $q,\tau,\delta \in [0,1]$,*

$$\Delta_{\underline{L}^Q}(q,\tau,\delta) \;\geq\; F_{\underline{L}^Q}(q,\tau,\delta). \tag{74}$$

Proof We have

$$\underline{L}^{Q(k)}(q) \;=\; \varrho\cdot\left\{\begin{array}{lll}\frac{(-1)^{k-1}(k-1)!}{q^k}-2\cdot[\![k=1]\!] & \text{if} & q<1/2\\ -\frac{(k-1)!}{(1-q)^k}+2\cdot[\![k=1]\!] & \text{if} & q>1/2\end{array}\right., \tag{75}$$

and we check that only the first and second order derivatives are defined in $q = 1/2$. Since $\underline{L}^Q$ is symmetric around $1/2$, $\Delta_{\underline{L}^Q}$ satisfies

$$\begin{aligned}\Delta_{\underline{L}^Q}\left(\frac{1}{2}-q,1-\tau,\delta\right) &= \underline{L}^Q\left(\frac{1}{2}-q\right)-\tau\underline{L}^Q\left(\frac{1}{2}-q-(1-\tau)\delta\right)-(1-\tau)\underline{L}^Q\left(\frac{1}{2}-q+\tau\delta\right)\\ &= \underline{L}^Q(q)-\tau\underline{L}^Q\left(\frac{1}{2}-(q+(1-\tau)\delta)\right)-(1-\tau)\underline{L}^Q\left(\frac{1}{2}-(q-\tau\delta)\right)\\ &= \underline{L}^Q(q)-\tau\underline{L}^Q(q+(1-\tau)\delta)-(1-\tau)\underline{L}^Q(q-\tau\delta)=\Delta_{\underline{L}^Q}(q,\tau,\delta), \quad (76)\end{aligned}$$

so we study $\Delta_{\underline{L}^Q}$ for $q > 1/2$ without loss of generality. In this case, all derivatives $\underline{L}^Q$ at order $k \geq 4$ are all negative, which from (Kearns & Mansour, 1996, Lemma 3) guarantees that

$$\Delta_{\underline{L}^Q}(q,\tau,\delta) \;\geq\; F_{\underline{L}^Q}(q,\tau,\delta), \quad (77)$$

as claimed. (end of the proof of Lemma 3) ■

We now lowerbound $F_{\underline{L}^Q}(q,\tau,\delta)$, which, from Lemma 3, will also provide a lowerbound for the decrease in $\Delta_{\underline{L}^Q}(q,\tau,\delta)$ and in fact will show Lemma 1. From now on, let us fix $\delta = \gamma q(1-q)/(\tau(1-\tau))$, if we denote $V(\tau,q) \doteq (1-2\tau)\left(q - [\![q < 1/2]\!]\right)$, then

$$F_{\underline{L}^Q}(q,\tau,\delta) \;=\; \max\{q,1-q\}^2\gamma^2\cdot\left(\frac{1}{2\tau(1-\tau)}+\frac{\gamma}{3\tau^2(1-\tau)^2}\cdot V(\tau,q)\right). \quad (78)$$

We immediately obtain

Lemma 4 *Let $\delta = \gamma q(1-q)/(\tau(1-\tau))$. Then for any τ, q such that $V(\tau,q) \geq 0$,*

$$F_{\underline{L}^Q}(q,\tau,\delta) \;\geq\; \frac{\gamma^2}{2}. \quad (79)$$

Proof For any τ, q such that $V(\tau,q) \geq 0$, we have

$$F_{\underline{L}^Q}(q,\tau,\delta) \;\geq\; \max\{q,1-q\}^2\gamma^2\cdot\frac{1}{2\tau(1-\tau)} \geq \frac{1}{4}\cdot\gamma^2\cdot 2 = \frac{\gamma^2}{2}, \quad (80)$$

as claimed (end of the proof of Lemma 4). ■

Lemma 4 means that when $\tau \leq 1/2, q < 1/2$ or $\tau \geq 1/2, q > 1/2$, the drop $\Delta_{\underline{L}^Q}(q,\tau,\delta)$ is guaranteed to be "big". If this does not happen, we make use of Lemma 2. In this case, if we pick wlog $\tau \leq 1/2, q > 1/2$ and get:

$$\begin{aligned}F_{\underline{L}^Q}(q,\tau,\delta) &= \max\{q,1-q\}^2\gamma^2\cdot\left(\frac{1}{2\tau(1-\tau)}-\frac{\gamma(1-2\tau)(1-q)}{3\tau^2(1-\tau)^2}\right)\\ &\geq \frac{\gamma^2}{2}\cdot\left(2-\frac{\gamma(1-2\cdot 0.4)}{3\cdot 0.4^2(1-0.4)^2}\right)=\gamma^2\cdot\left(1-\frac{625\gamma}{216}\right)\geq\gamma^2\cdot\left(1-\frac{25}{216}\right)\geq\frac{\gamma^2}{2},\end{aligned}$$

which therefore implies that $F_{\underline{L}^Q}(q,\tau,\delta) \geq \gamma^2/2$ in all cases. We just have to use Lemma 3 to finish the proof of Lemma 1 (end of the proof of Lemma 1). ■

We can now finish the proof of Theorem 10. Suppose the current tree H has t leaves. There must be a leaf with $\omega(\lambda) \geq 1/t$, so

$$\begin{aligned} \Delta &\doteq \underline{L}^Q(H) - \underline{L}^Q(H^+) \\ &= \omega(\lambda)\Delta_{\underline{L}^Q}(q,\tau,\delta) \geq \frac{\gamma^2}{2t} \\ &\geq \frac{\gamma^2}{2t} \cdot \frac{\underline{L}^Q(H)}{\underline{L}^Q(H_0)}, \end{aligned} \tag{81}$$

where the last inequality follows from the concavity of $\underline{L}^Q$, letting H_0 the single-root node tree for which $\underline{L}^Q(H_0) = \underline{L}^Q(q(\mathcal{S}))$, and more generally H_t a tree with $t+1$ leaves (thus we have made t iterations of the boosting procedure). It therefore comes the recurrence relationship

$$\underline{L}^Q(H_{t+1}) \leq \left(1 - \frac{\gamma^2}{2\underline{L}^Q(q(\mathcal{S})) \cdot t}\right) \cdot \underline{L}^Q(H_t), \tag{82}$$

and we get (see (Kearns & Mansour, 1996, proof of Theorem 10))

$$\underline{L}^Q(H_t) \leq \exp\left(-\frac{\gamma^2 \log t}{4\underline{L}^Q(q(\mathcal{S}))}\right) \cdot \underline{L}^Q(q(\mathcal{S})), \tag{83}$$

to obtain $\underline{L}^Q(H_t) \leq \rho \cdot \underline{L}^Q(q(\mathcal{S}))$ for $\rho \in (0,1]$, it therefore suffices that

$$t \geq \left(\frac{1}{\rho}\right)^{\frac{4 \cdot \underline{L}^Q(q(\mathcal{S}))}{\gamma^2}}. \tag{84}$$

We finally remark that $\underline{L}^Q(q(\mathcal{S})) \leq \varrho \cdot \log 1/(2\varepsilon)$ and conclude that (84) holds when

$$t \geq \left(\frac{1}{\rho}\right)^{\frac{4\varrho}{\gamma^2 \log \frac{1}{2\varepsilon}}}, \tag{85}$$

as claimed.

Remark: we can compare at this stage our guarantees to those of Kearns & Mansour (1996). The knowledge of their proofs immediately sheds light on the fact that our lowerbound on $\Delta_{\underline{L}^Q}(q,\tau,\delta)$ in Lemma 10 does not depend on q whereas all of theirs do (Kearns & Mansour, 1996, Lemmata 5, 6, 7), and in fact vanish as $q \to 0, 1$. A closer look at the weak learning assumption shows that it in fact precludes this extreme regime for q as it enforces $q \in [\tau\delta, 1-(1-\tau)\delta]$ when $\delta \leq 1$; as a consequence their bounds can also be reformulated to exclude q and their convergence rate for their best splitting criterion is within the same order as ours.

7 Experiments *in extenso*

7.1 Implementation

We give here a few details on the implementation. The Java implementation of the algorithms, available separately, implements the version of Nock & Nielsen (2006); Schapire & Singer (1999) respectively for ADABOOST$_{\mathbb{R}}$ and AdaBoost.

The implementation of RATBOOSTE uses methods from class `Math` that allow to throw an `ArithmeticException` when a `long` overflow happens – in which case we catch the exception and redo the corresponding method after quantization. To make the code faster, we have also included the possibility to trigger quantization when the `long`s encoding length exceeds a user-fixed threshold.

The implementation of RATBOOSTA$_b$ uses a regular k-means with Forgy initialization. If you want to optimize this with your best hard clustering algorithm, you just have to rewrite a few methods from class `KMeans_R` in file `Misc.java`. Note that the implementation also allows to use stochastic weight assignation with adaptive quantization (a combination of RATBOOSTA$_b$ and RATBOOSTQ$_b$), but it is not reported (see `README`).

Domain summary Table

Table 1 details the UCI domains we have used Blake et al. (1998). We now detail the per-domain training curves *when there is no stopping criterion* (other than to boost for 10 000 iterations). In the results reported in Tables 1 (main file) and 2 (this), we keep the classifier which minimizes the empirical risk among all iterations, which amounts to a cutoff point for boosting around the minimal values of each curve (because of the statistical uncertainty, we are not guarantee that this may be minimal on testing). Results of ADABOOST$_{\mathbb{R}}$ are omitted to not clutter the plots but they are included in the full Table 2.

Domain	m	d
Fertility	100	9
Haberman	306	3
Transfusion	748	4
Banknote	1 372	4
Breast wisc	699	9
Ionosphere	351	33
Sonar	208	60
Yeast	1 484	7
Wine-red	1 599	11
Cardiotocography (*)	2 126	9
CreditCardSmall (**)	1 000	23
Abalone	4 177	8
Qsar	1 055	41
Wine-white	4 898	11
Page	5 473	10
Mice	1 080	77
Hill+noise	1 212	100
Hill+nonoise	1 212	100
Firmteacher	10 800	16
Magic	19 020	10
EEG	14 980	14
Skin	245 057	3
Musk	6 598	166
Hardware	28 179	95
Twitter (***)	583 250	77

Table 1: UCI domains considered in our experiments (m = total number of examples, d = number of features), ordered in increasing $m \times d$. (*) we used features 13-21 as descriptors; (**) we used the first 1 000 examples of the UCI domain; (***) due to the size of the domain, only AdaBoost and AdaBoost$_{\mathbb{R}}$ were run for $T = 5000$ iterations, the other algorithms were rum for a smaller $T' = 1000$ iterations.

UCI `fertility`

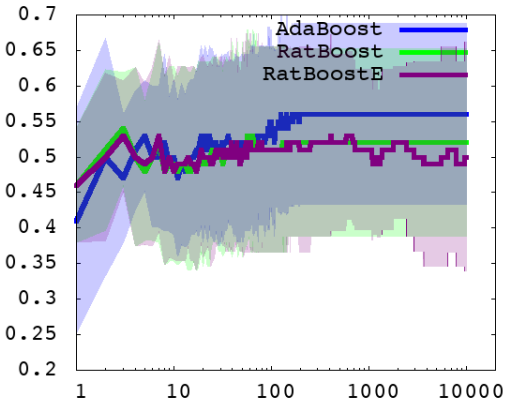

Figure 1: UCI domain `fertility`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

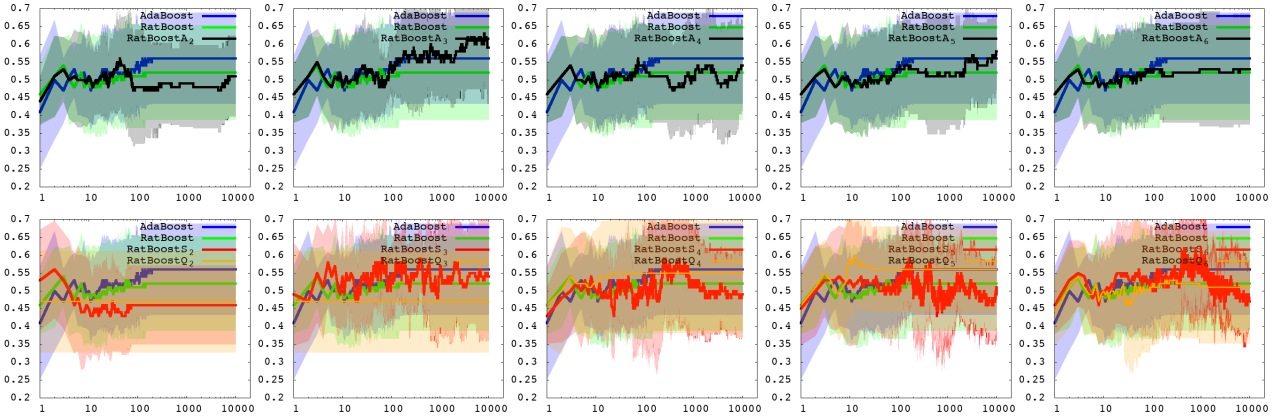

Figure 2: UCI domain `fertility`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `haberman`

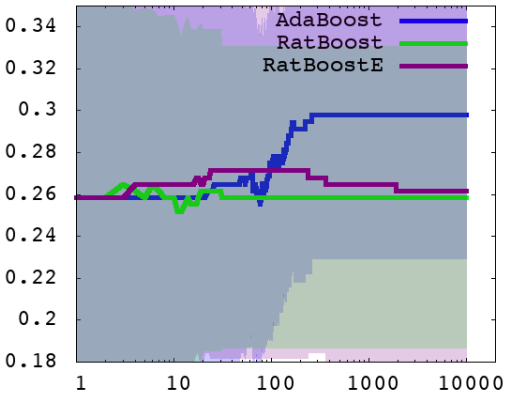

Figure 3: UCI domain `haberman`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

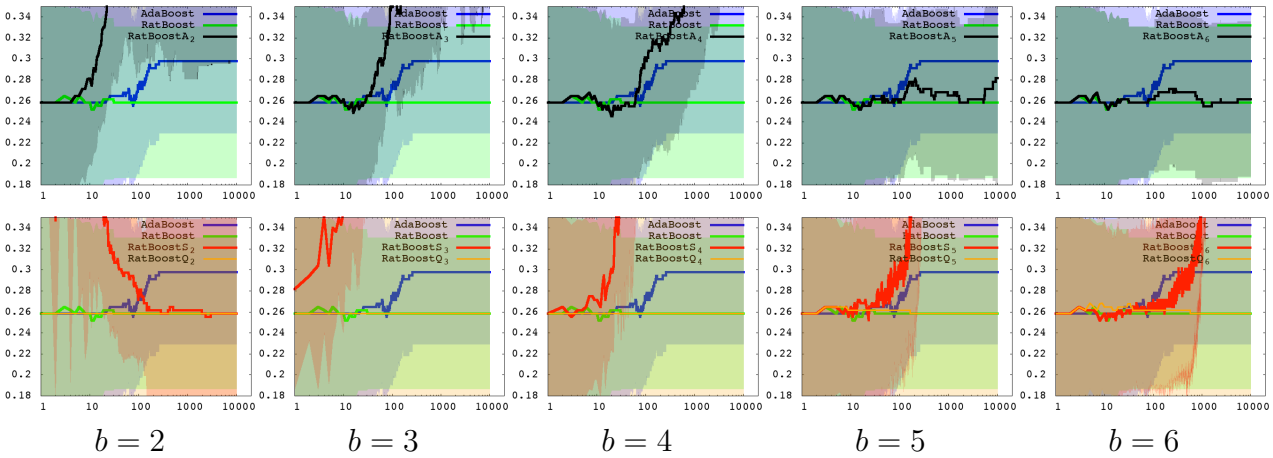

Figure 4: UCI domain `haberman`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `transfusion`

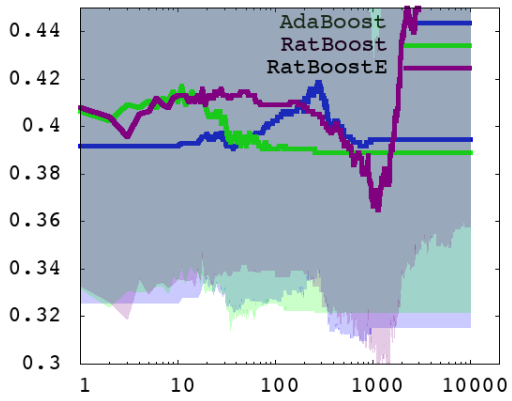

Figure 5: UCI domain `transfusion`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

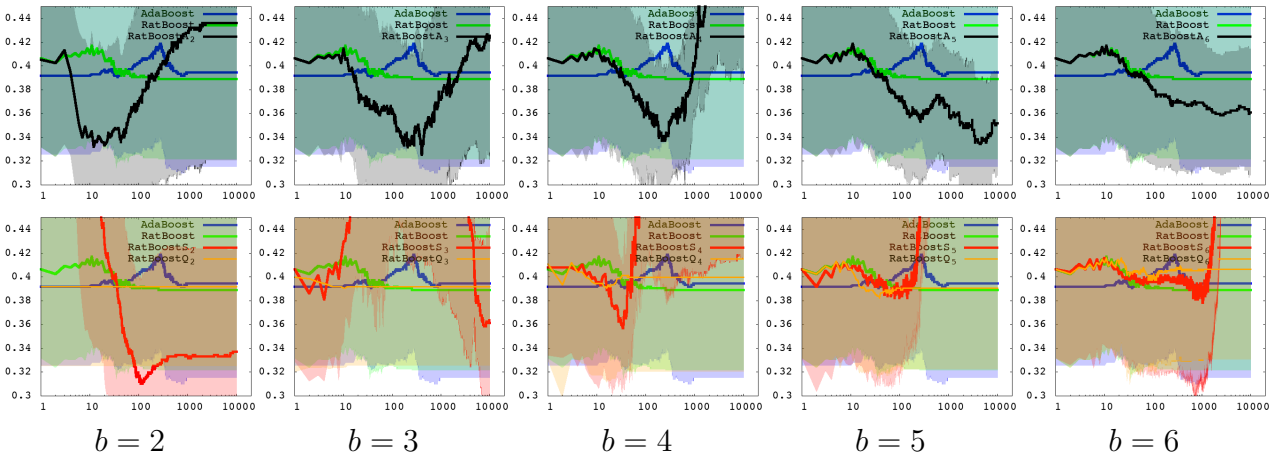

Figure 6: UCI domain `transfusion`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `banknote`

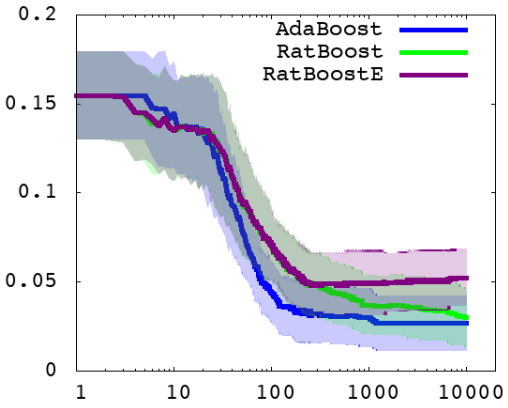

Figure 7: UCI domain `banknote`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

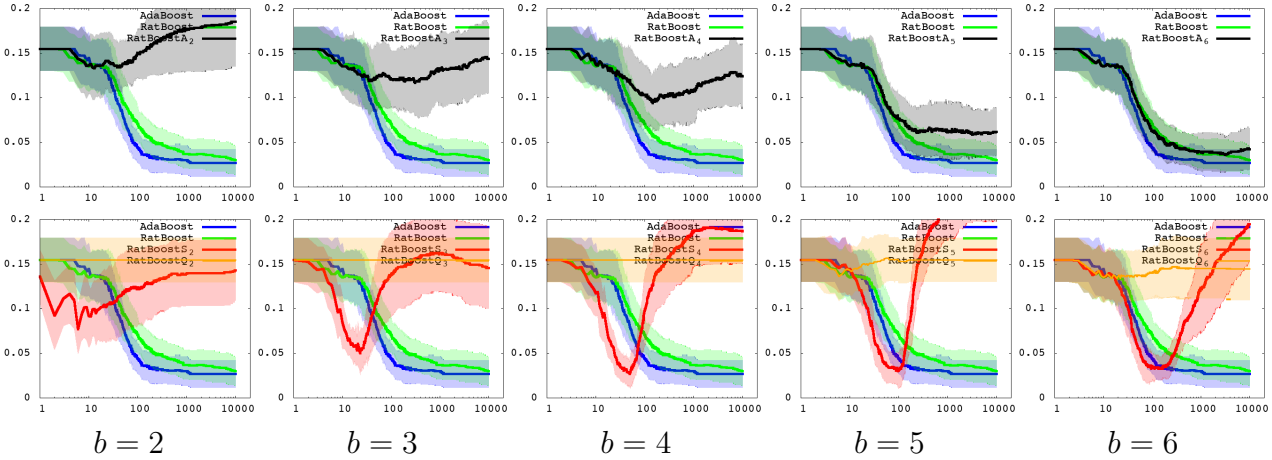

Figure 8: UCI domain `banknote`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `breastwisc`

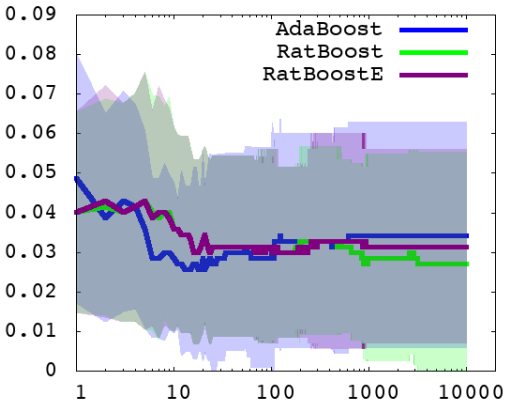

Figure 9: UCI domain `breastwisc`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

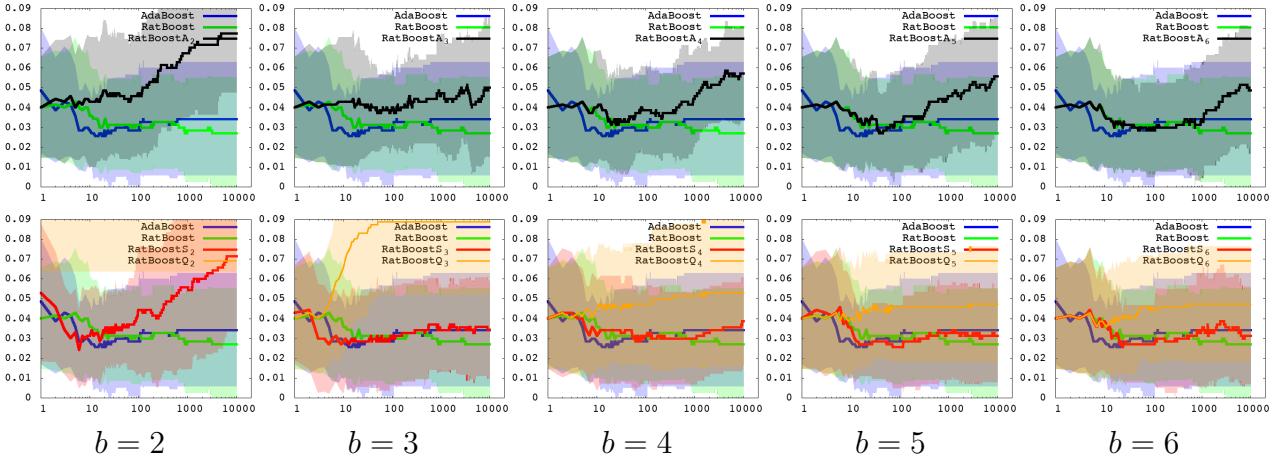

Figure 10: UCI domain `breastwisc`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `ionosphere`

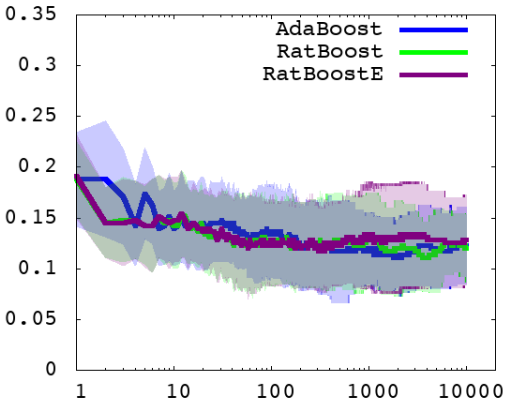

Figure 11: UCI domain `ionosphere`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

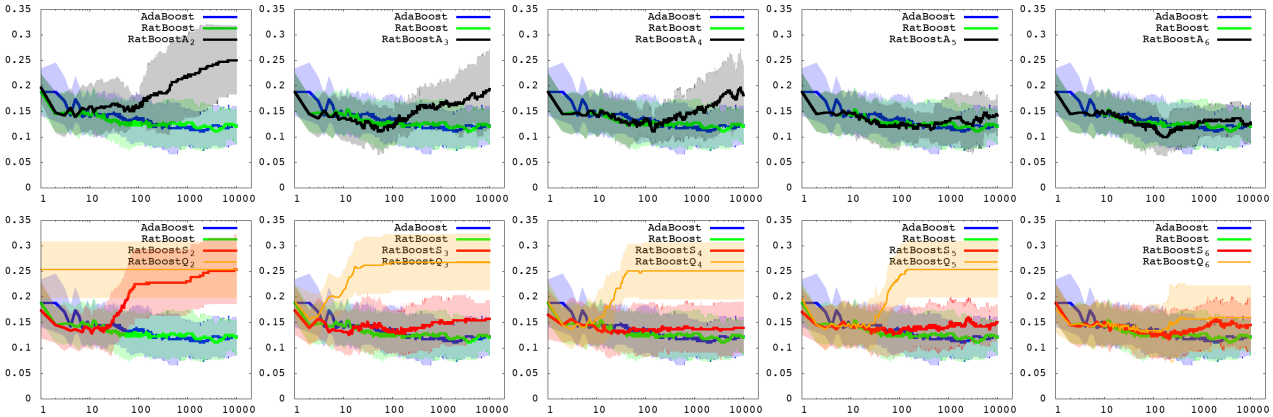

Figure 12: UCI domain `ionosphere`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `sonar`

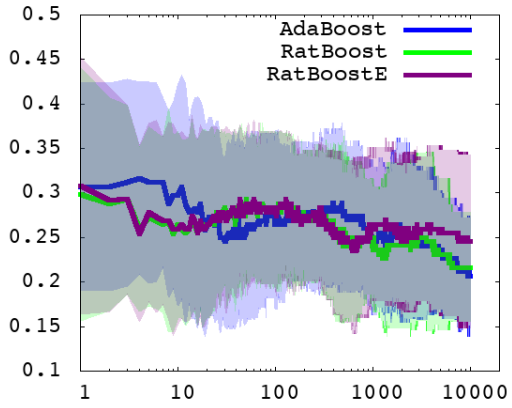

Figure 13: UCI domain `sonar`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

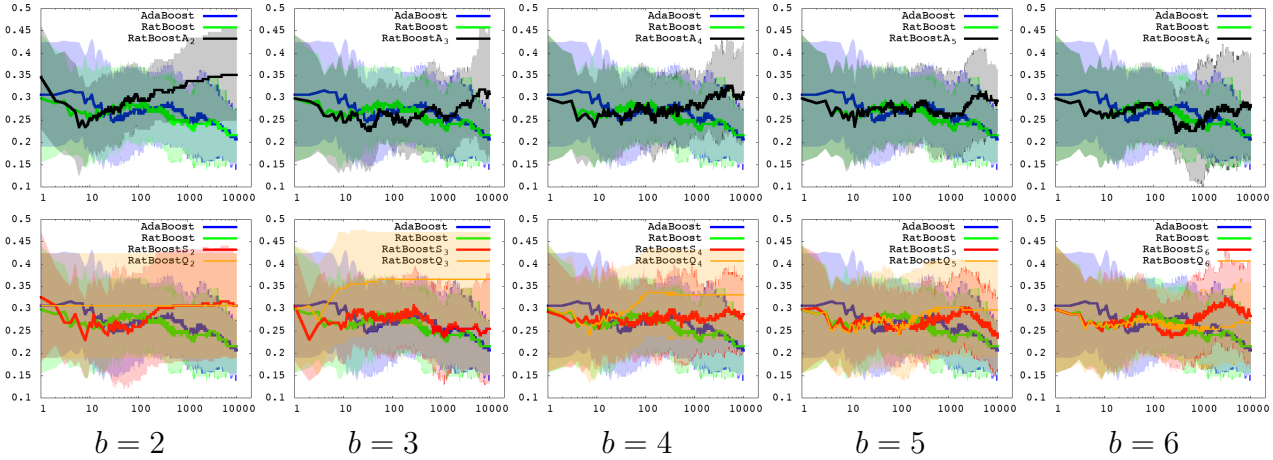

Figure 14: UCI domain `sonar`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `yeast`

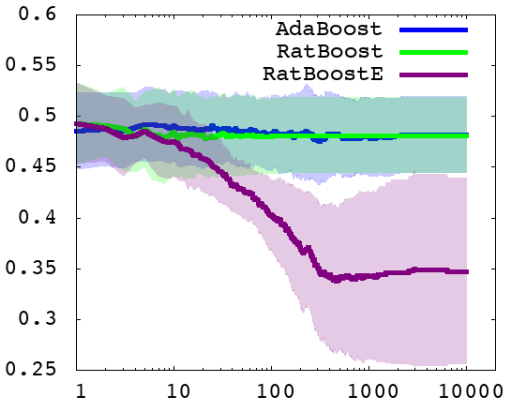

Figure 15: UCI domain `yeast`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

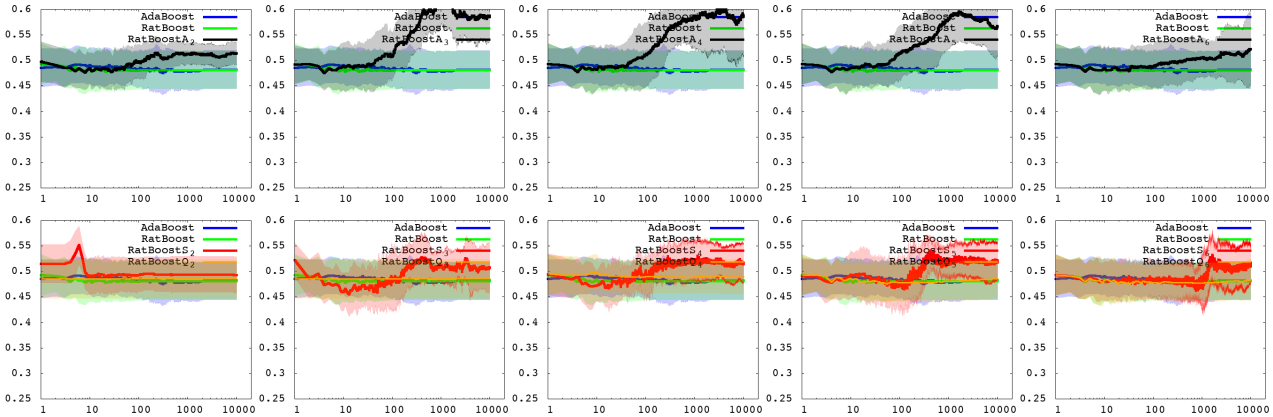

Figure 16: UCI domain `yeast`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `winered`

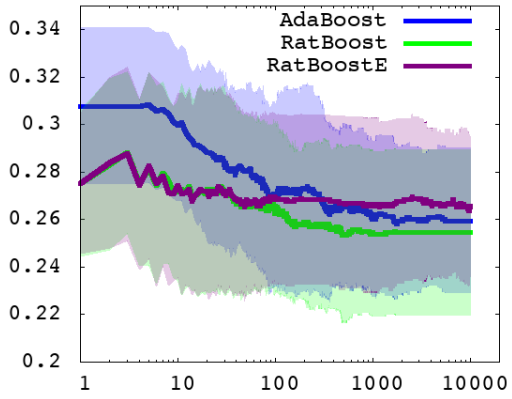

Figure 17: UCI domain `winered`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

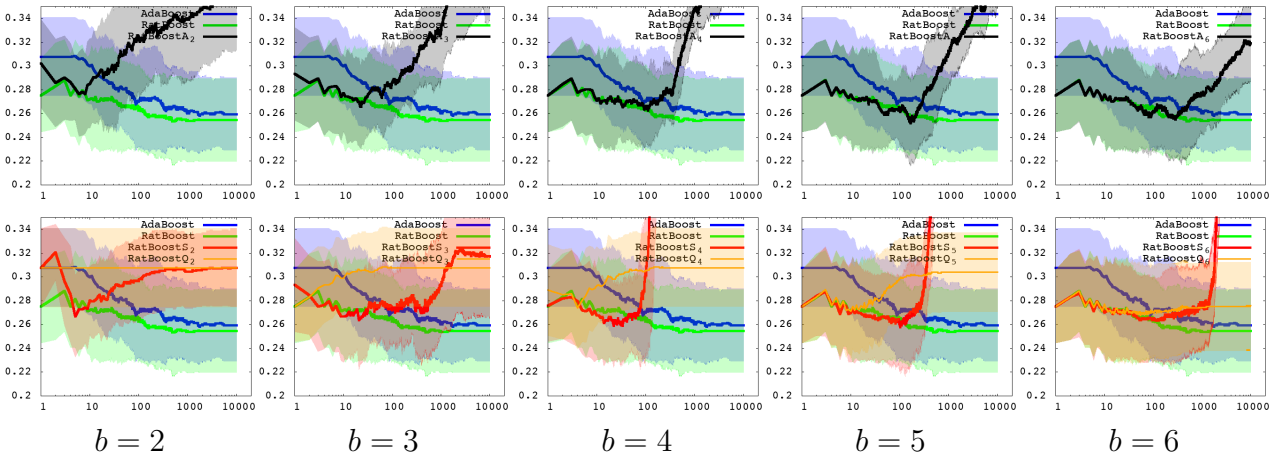

Figure 18: UCI domain `winered`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `cardiotocography`

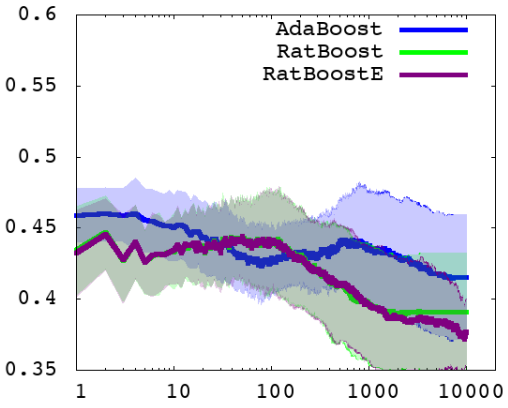

Figure 19: UCI domain `cardiotocography`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

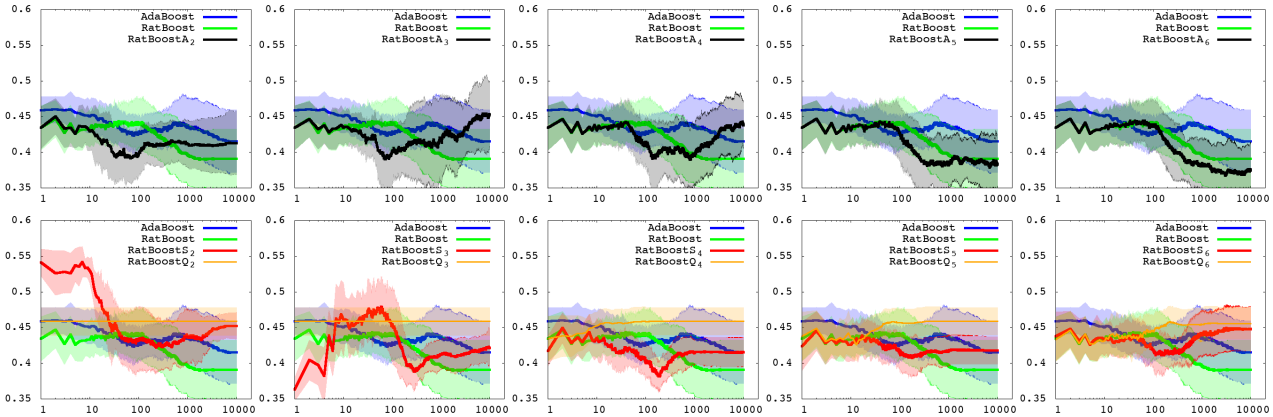

Figure 20: UCI domain `cardiotocography`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `CreditCardSmall`

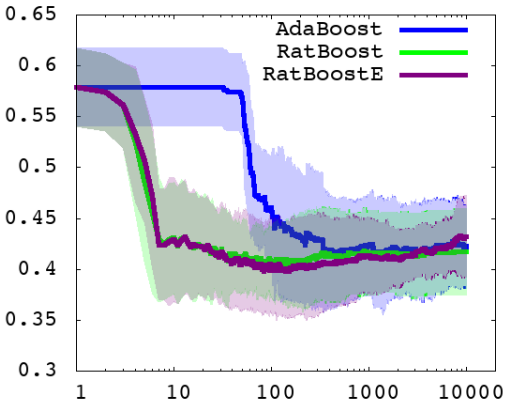

Figure 21: UCI domain `creditcardsmall`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

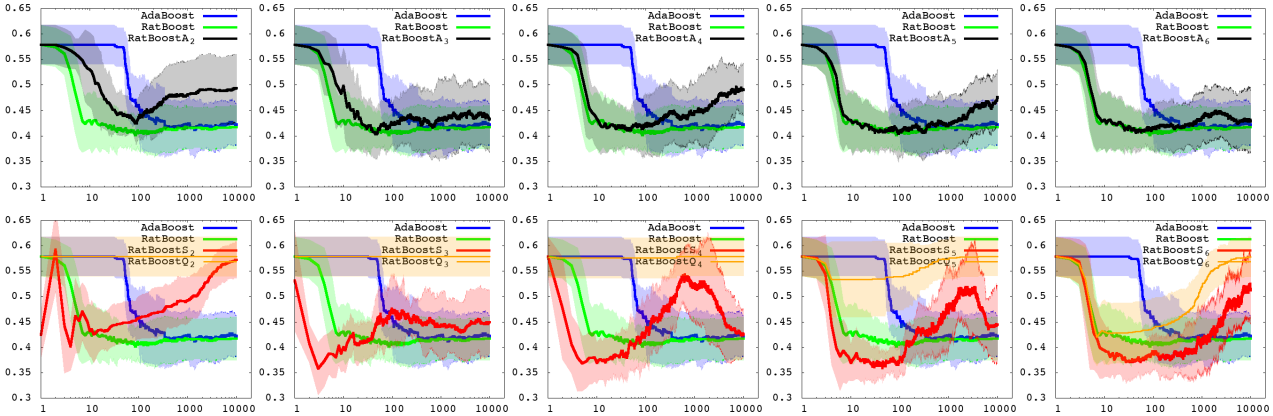

Figure 22: UCI domain `creditcardsmall`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `abalone`

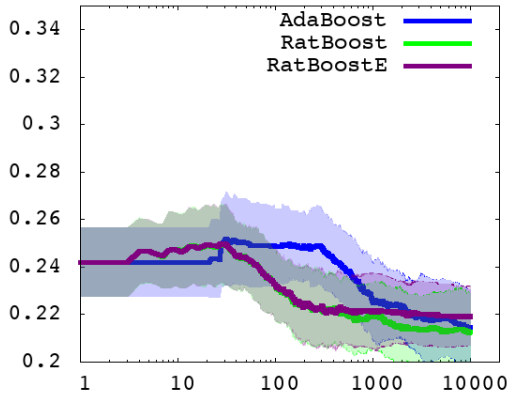

Figure 23: UCI domain `abalone`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

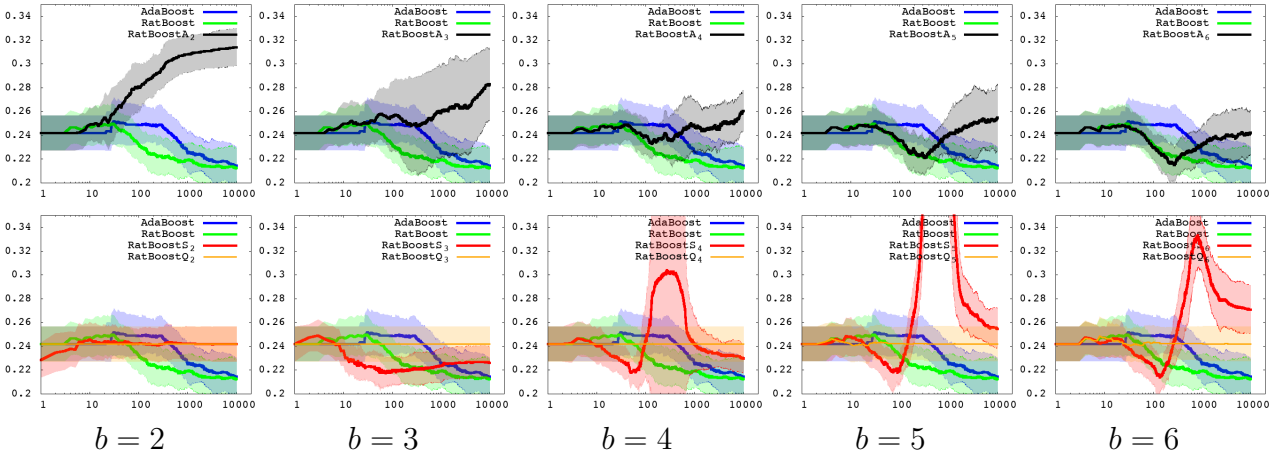

Figure 24: UCI domain `abalone`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `qsar`

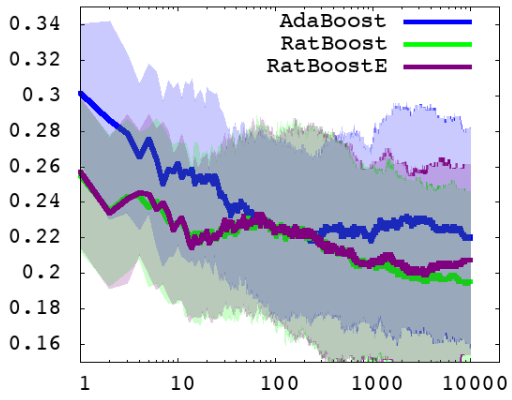

Figure 25: UCI domain `qsar`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

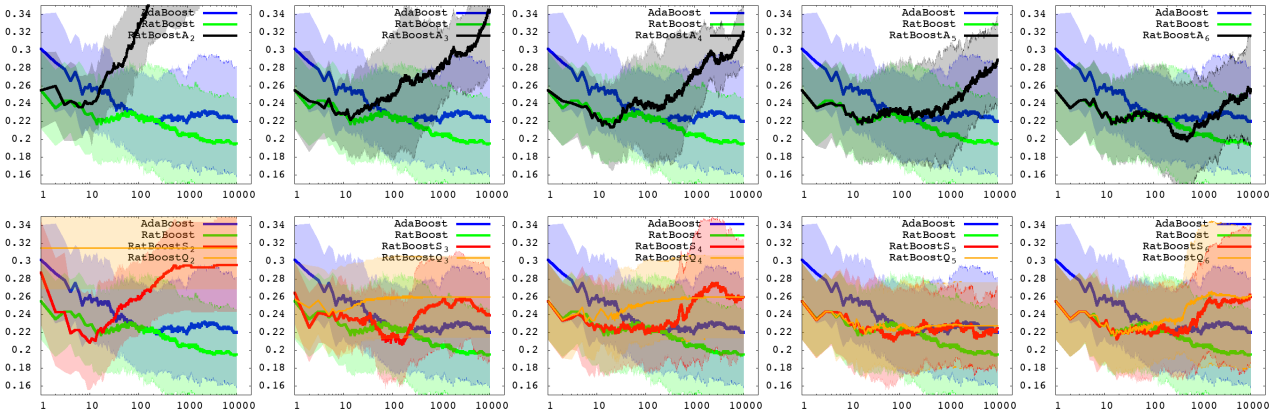

Figure 26: UCI domain `qsar`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `winewhite`

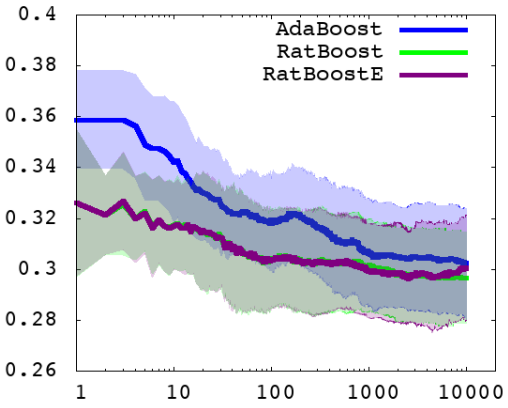

Figure 27: UCI domain `winewhite`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

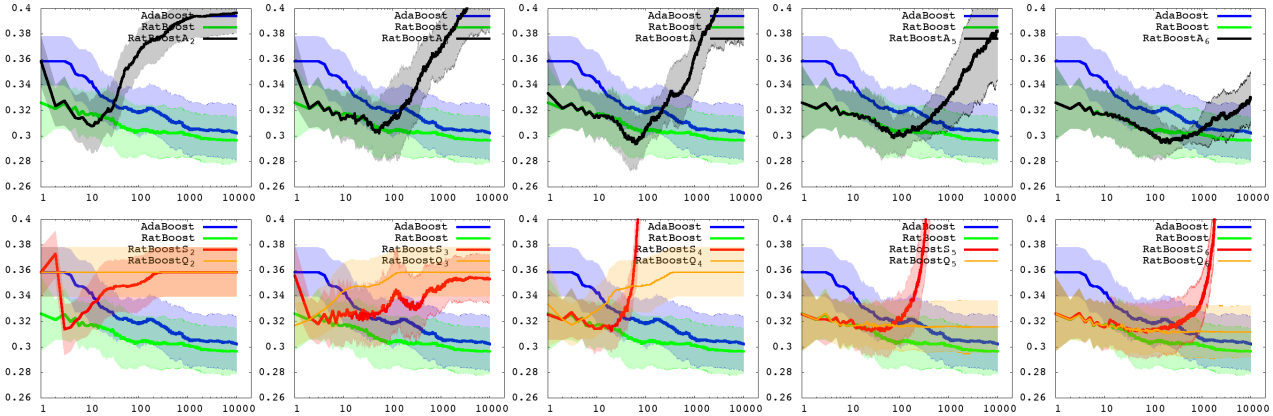

Figure 28: UCI domain `winewhite`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI page

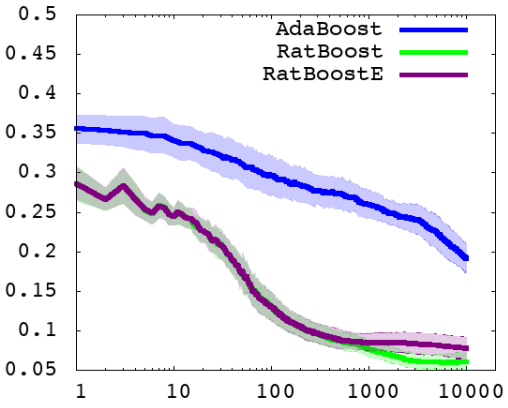

Figure 29: UCI domain `page`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

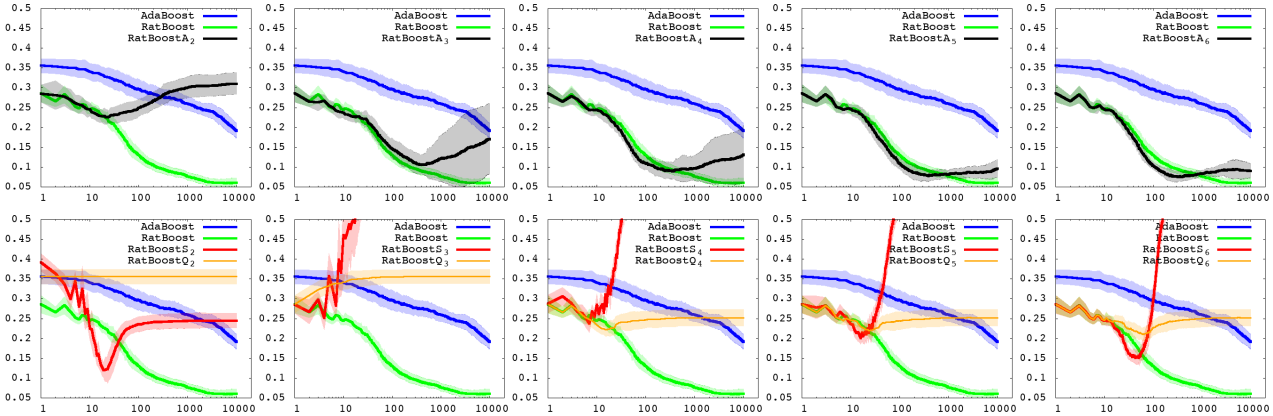

Figure 30: UCI domain `page`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `mice`

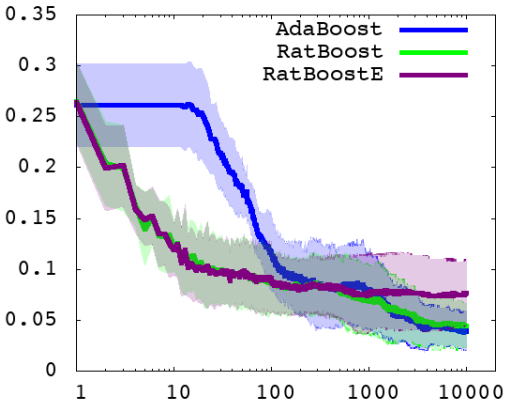

Figure 31: UCI domain `mice`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

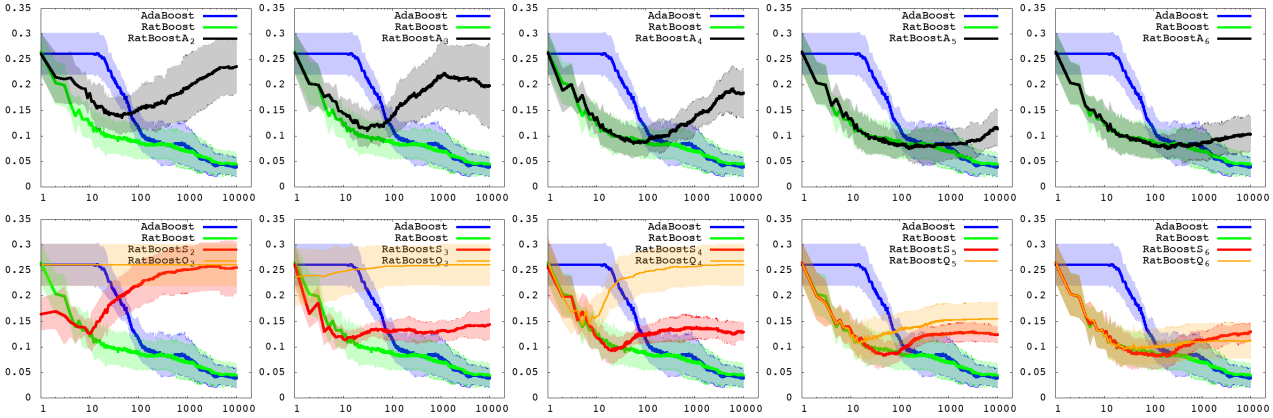

Figure 32: UCI domain `mice`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `hill+noise`

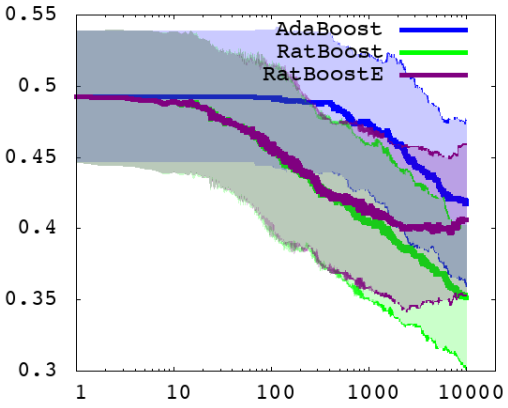

Figure 33: UCI domain `hill+noise`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

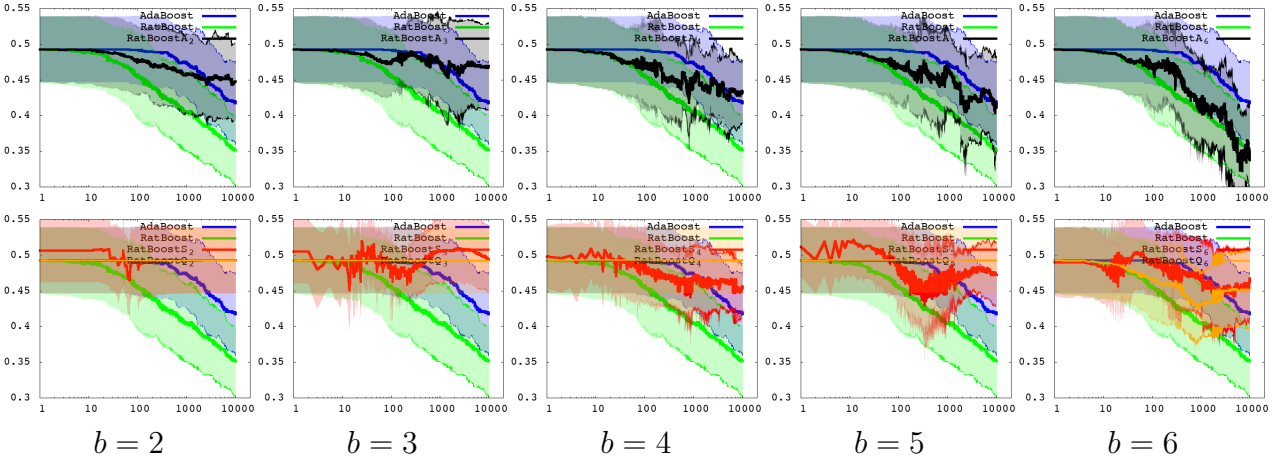

Figure 34: UCI domain `hill+noise`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `hill+nonoise`

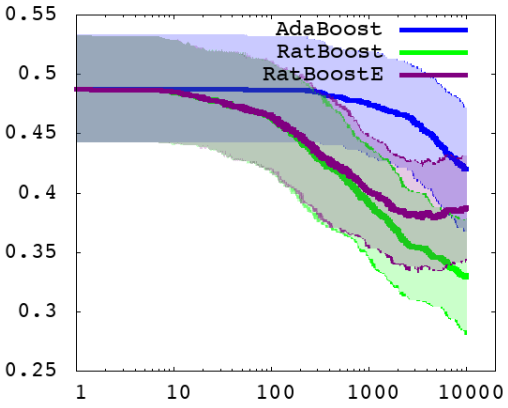

Figure 35: UCI domain `hill+nonoise`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

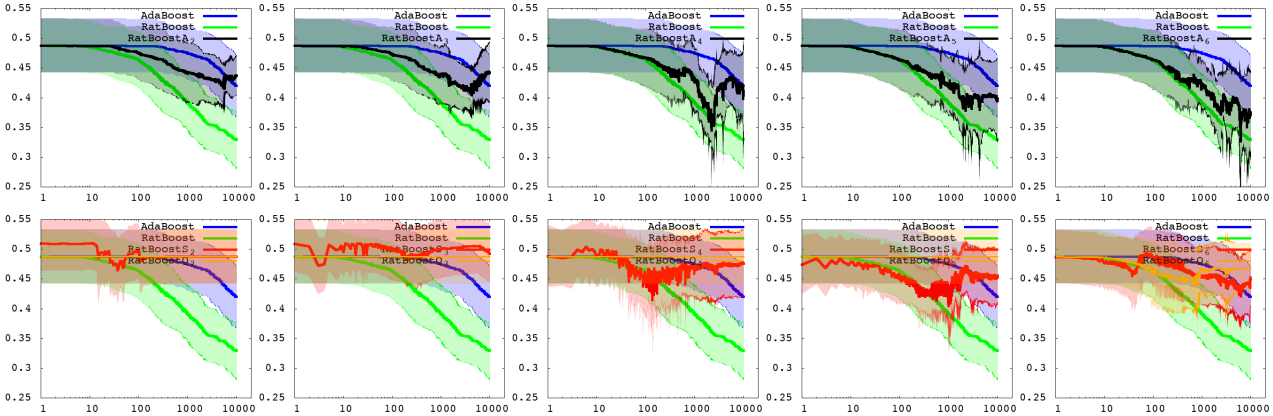

Figure 36: UCI domain `hill+nonoise`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `firmteacher`

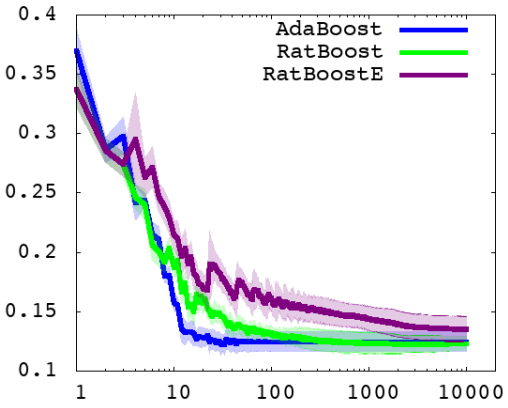

Figure 37: UCI domain `firmteacher`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

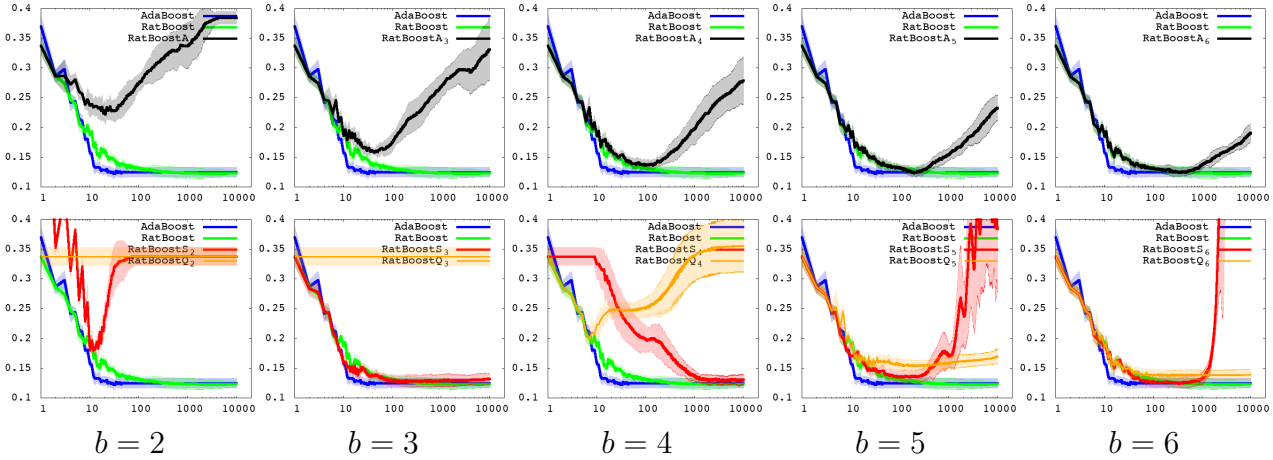

Figure 38: UCI domain `firmteacher`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `magic`

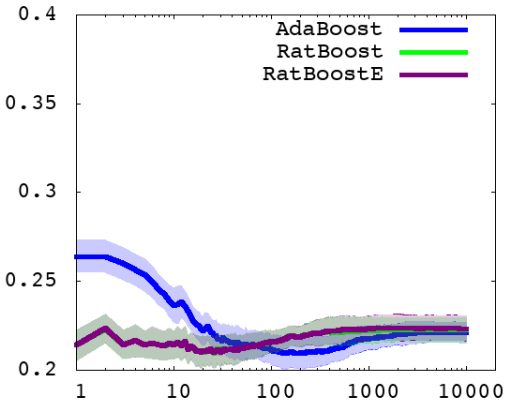

Figure 39: UCI domain `magic`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

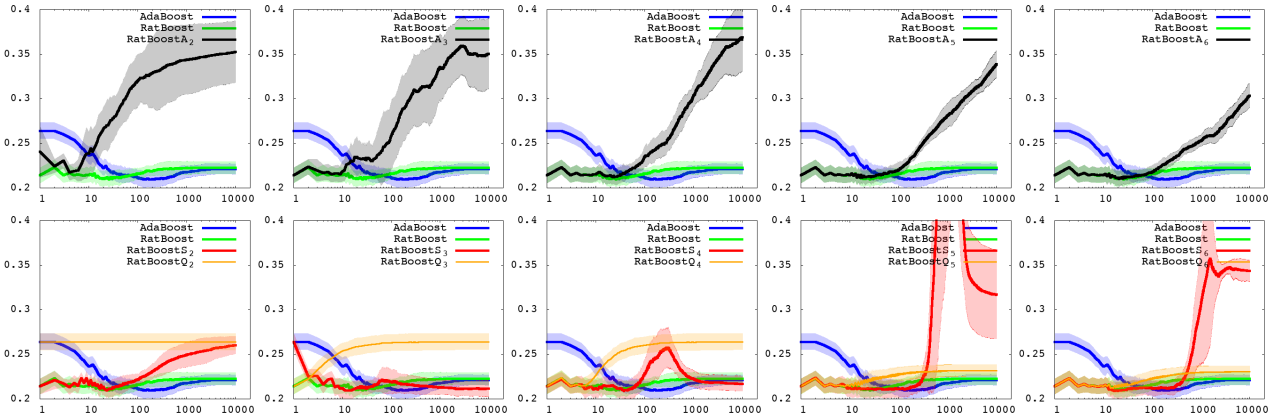

Figure 40: UCI domain `magic`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI eeg

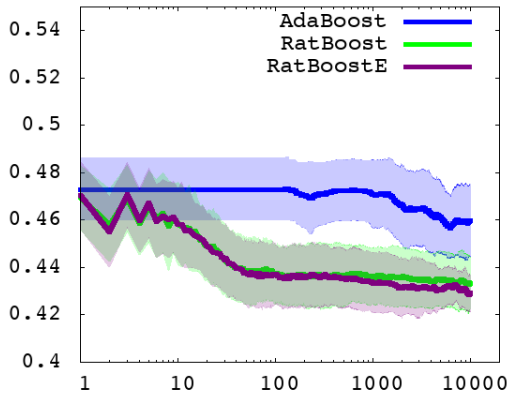

Figure 41: UCI domain `eeg`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

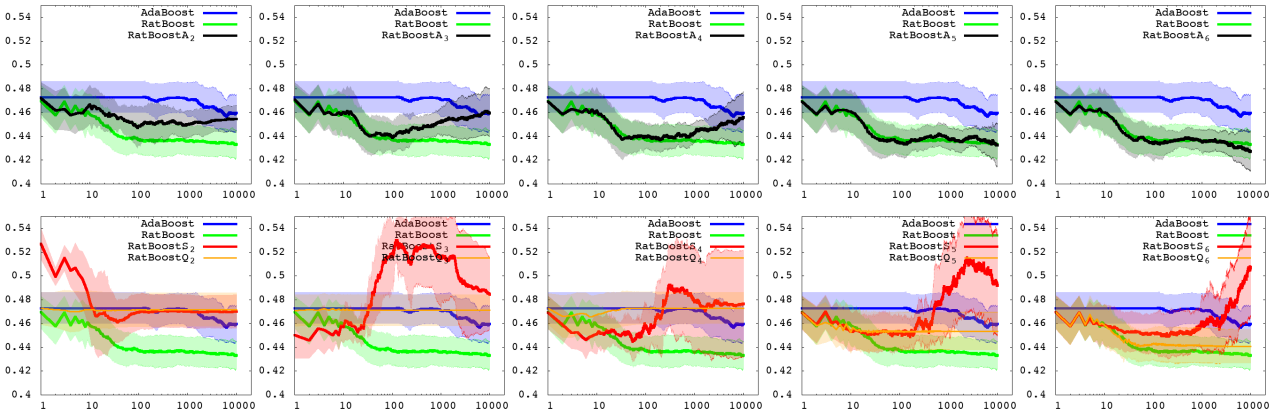

Figure 42: UCI domain `eeg`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `skin`

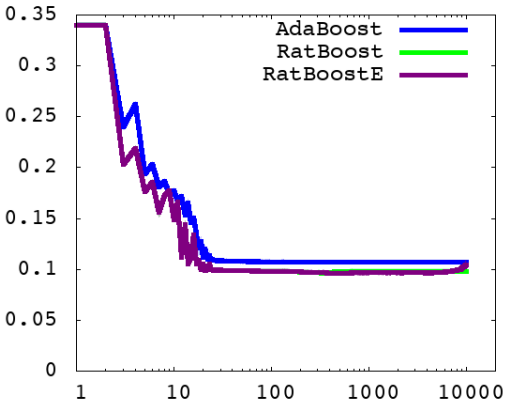

Figure 43: UCI domain `skin`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

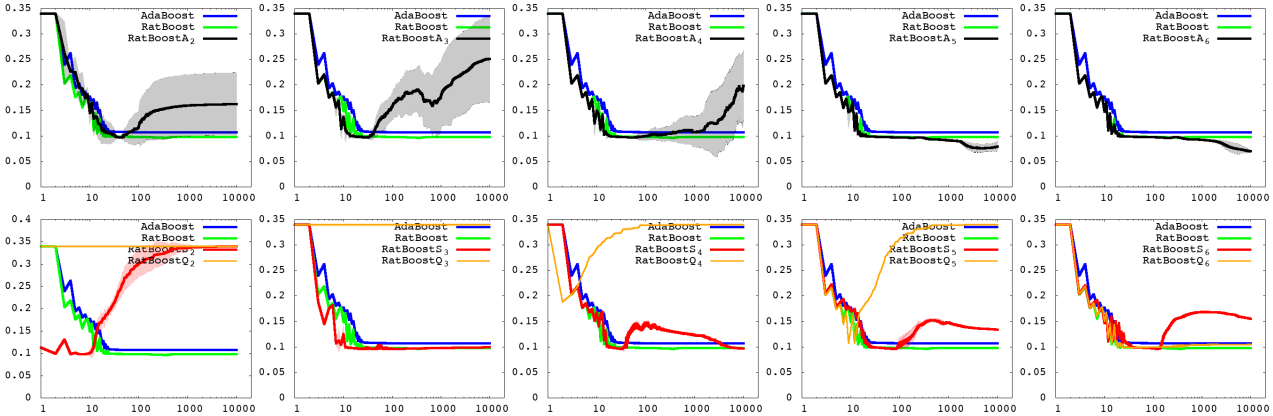

Figure 44: UCI domain `skin`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `musk`

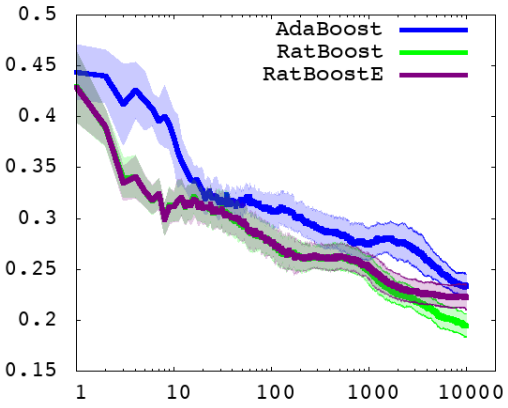

Figure 45: UCI domain `musk`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

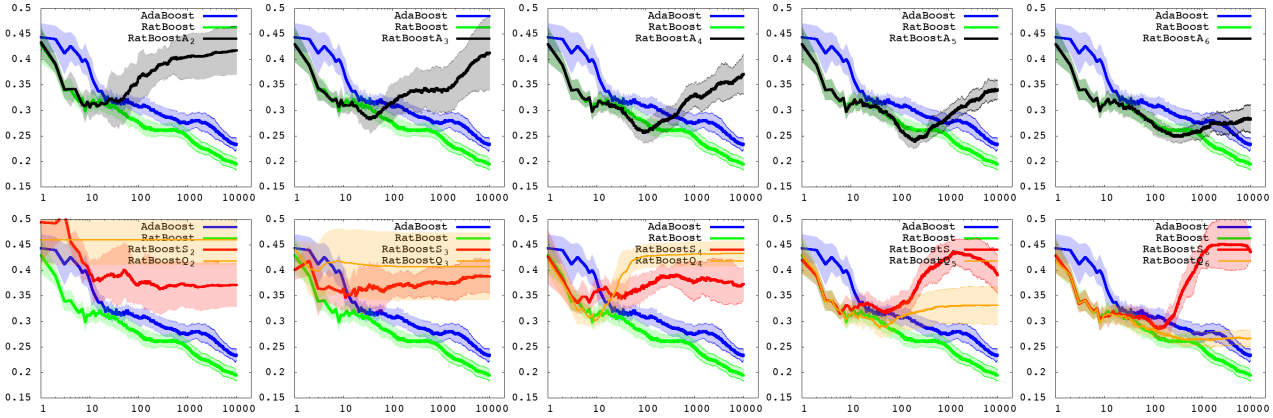

Figure 46: UCI domain `musk`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `hardware`

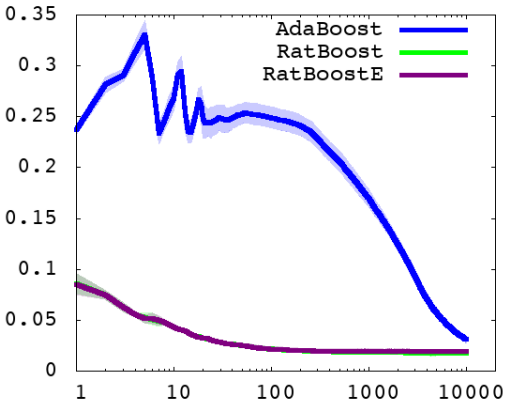

Figure 47: UCI domain `hardware`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

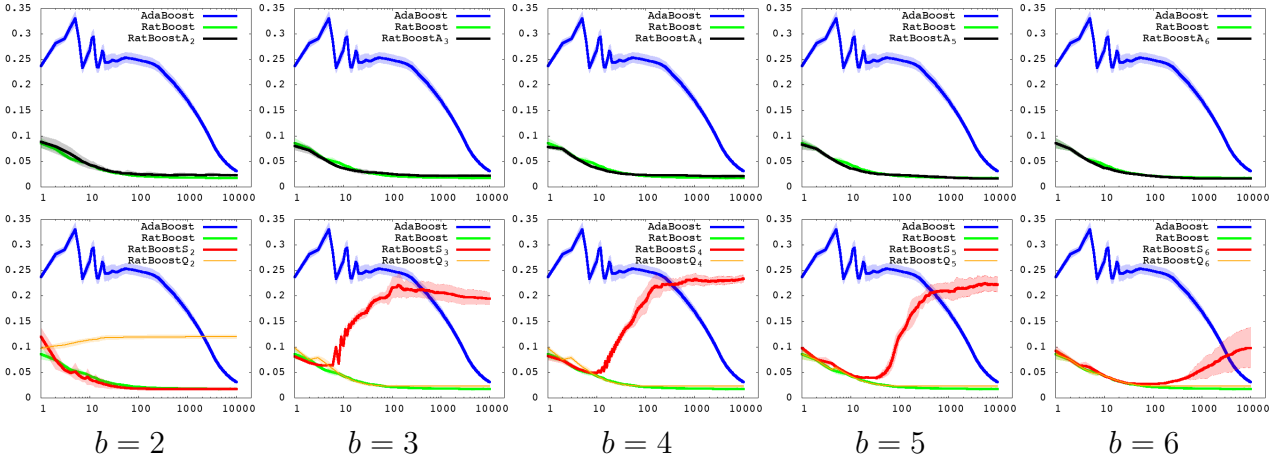

Figure 48: UCI domain `hardware`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 10000$ iterations.

UCI `twitter`

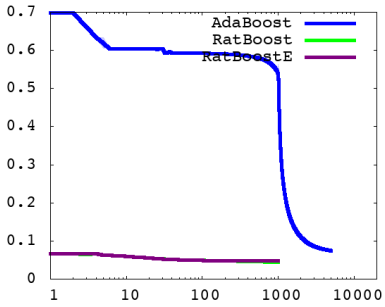

Figure 49: UCI domain `twitter`. Results comparing AdaBoost (blue), RATBOOST (green) and RATBOOSTE (purple). *Note*: there is *no* other stopping criterion apart from running for $T = 5000$ iterations (AdaBoost) and $T' = 1000$ iterations (RATBOOST, RATBOOSTE).

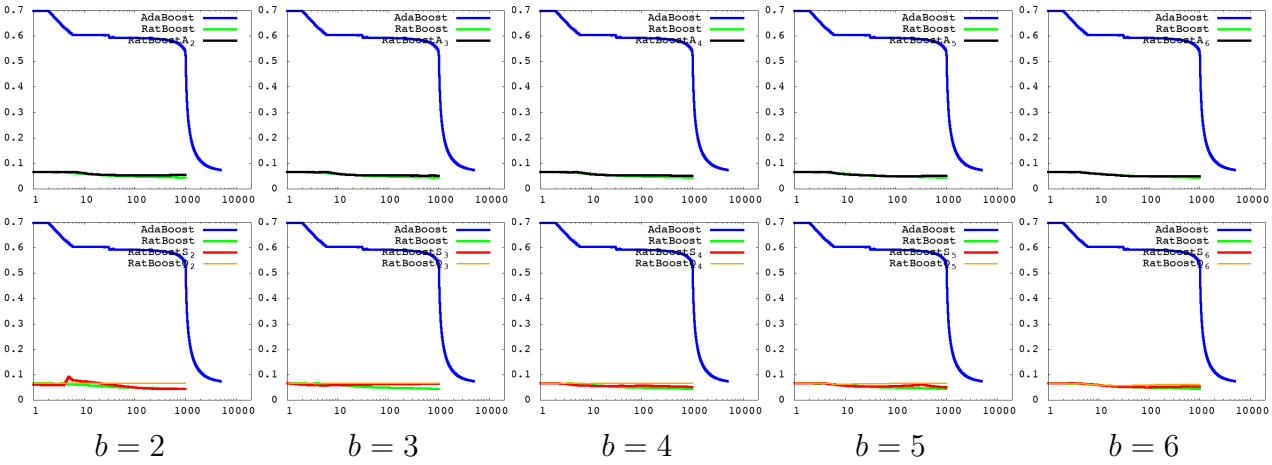

Figure 50: UCI domain `twitter`. Results comparing AdaBoost (blue), RATBOOST (green) and the quantized versions RATBOOSTA$_b$ (black) / RATBOOSTQ$_b$ (thin orange) / RATBOOSTS$_b$ (red), for various values of the quantization index bit-size b. *Note*: there is *no* other stopping criterion apart from running for $T = 5000$ iterations (AdaBoost) and $T' = 1000$ iterations (RATBOOST, RATBOOSTA$_b$, RATBOOSTQ$_b$, RATBOOSTS$_b$).

Summary of Results

References

Blake, C. L., Keogh, E., and Merz, C. UCI repository of machine learning databases, 1998. http://www.ics.uci.edu/∼mlearn/MLRepository.html.

Buja, A., Stuetzle, W., and Shen, Y. Loss functions for binary class probability estimation and classification: structure and applications, 2005. Technical Report, University of Pennsylvania.

Kearns, M. and Mansour, Y. On the boosting ability of top-down decision tree learning algorithms. In *Proc. of the 28th ACM STOC*, pp. 459–468, 1996.

Kearns, M. J. and Mansour, Y. A Fast, Bottom-up Decision Tree Pruning algorithm with Near-Optimal generalization. In *Proc. of the 15th International Conference on Machine Learning*, pp. 269–277, 1998.

Nock, R. and Nielsen, F. A $\mathbb{R}$eal Generalization of discrete AdaBoost. In *Proc. of the 17th European Conference on Artificial Intelligence*, pp. 509–515, 2006.

Nock, R. and Nielsen, F. On the efficient minimization of classification-calibrated surrogates. In *NIPS*21*, pp. 1201–1208, 2008.

Reid, M.-D. and Williamson, R.-C. Composite binary losses. *JMLR*, 11:2387–2422, 2010.

Schapire, R. E. and Singer, Y. Improved boosting algorithms using confidence-rated predictions. *MLJ*, 37:297–336, 1999.

Schervish, M.-J. A general method for comparing probability assessors. *Ann. of Stat.*, 17(4): 1856–1879, 1989.

Shuford, Jr, E.-H., Albert, A., and Massengil, H.-E. Admissible probability measurement procedures. *Psychometrika*, 31:125–145, 1966.

	AdaBoost	AdaBoost$_{\mathrm{E}}$	RatBoost	RatBoost$_{\mathrm{E}}$	RatBoostQ$_b$, $b=$					RatBoostS$_b$, $b=$					RatBoostA$_b$, $b=$				
					2	3	4	5	6	2	3	4	5	6	2	3	4	5	6
F	38.00±10.33	37.00±9.49	40.00±9.43	40.00±11.55	47.00±14.94	47.00±14.94	39.00±15.24	42.00±11.35	42.00±11.35	38.00±7.89	46.00±18.38	39.00±7.38	47.00±14.18	52.00±16.19	41.00±7.38	46.00±12.65	43.00±9.49	47.00±14.94	39.00±8.76
H	25.53±8.79	25.53±8.79	25.85±8.32	26.81±9.71	25.84±9.83	25.84±9.83	25.84±9.83	26.48±9.07	26.81±9.71	25.84±9.34	26.51±8.86	26.49±8.96	25.52±9.65	26.17±10.01	25.52±9.65	25.52±8.90	29.80±11.78	28.18±11.36	25.84±8.31
T	38.78±6.86	39.05±6.68	38.92±7.15	34.91±7.25	39.18±7.02	39.18±7.22	38.53±9.35	38.52±7.20	40.66±7.98	30.24±8.27	34.65±8.25	35.31±6.62	37.58±5.44	38.24±6.92	33.59±8.06	30.91±8.44	32.37±6.14	32.78±5.57	35.71±5.39
B	2.70±1.62	2.63±1.55	2.99±1.70	4.89±1.89	15.46±2.59	15.46±2.59	15.46±2.59	13.64±2.96	13.93±2.97	7.73±3.12	4.45±2.32	2.77±1.78	2.99±2.08	3.43±1.33	12.54±3.22	11.01±4.25	8.53±2.85	5.32±2.07	3.57±2.16
BW	2.86±2.78	2.86±2.78	3.29±2.52	3.14±2.59	10.02±3.83	4.01±3.07	4.72±2.94	4.15±2.56	3.58±2.72	2.86±2.13	3.14±2.68	3.00±2.73	2.86±2.43	3.00±2.65	4.29±3.09	4.29±2.85	3.29±2.34	3.29±1.91	3.00±2.37
I	11.39±4.01	11.11±3.91	11.68±3.92	12.54±5.26	25.37±5.82	15.10±4.28	13.67±4.82	14.82±5.85	13.69±4.25	12.83±3.64	13.97±4.58	12.82±4.89	13.97±4.38	13.40±5.41	14.53±4.94	11.40±2.72	11.96±2.21	11.41±4.49	13.40±3.60
S	20.67±7.12	20.67±7.12	21.64±6.47	25.48±9.88	30.69±12.30	25.88±15.02	25.50±11.50	27.40±11.23	27.38±9.72	24.02±8.71	28.40±8.74	28.86±9.11	28.38±9.25	26.07±9.29	23.10±9.62	22.62±9.88	26.90±7.79	28.83±9.37	24.05±11.76
Y	48.18±4.43	48.18±4.43	48.59±4.59	34.04±7.09	48.52±4.00	48.52±4.00	48.79±3.47	49.06±4.15	48.45±4.60	49.33±3.83	47.17±5.18	46.23±4.47	47.11±4.47	46.77±3.67	48.79±2.64	49.53±3.18	47.71±4.02	48.72±3.69	49.33±3.97
WR	26.14±3.02	26.14±3.15	25.45±3.70	26.27±3.18	30.77±3.48	27.52±3.13	27.46±3.66	26.83±4.30	27.02±4.23	27.08±3.36	26.71±4.02	26.20±3.94	26.20±4.04	26.89±4.08	27.64±3.45	27.14±4.02	27.14±2.67	26.21±3.51	25.96±3.39
Ca	41.63±4.62	41.58±4.55	39.23±4.46	37.91±2.88	45.86±2.04	45.86±2.04	42.43±3.06	42.80±3.09	42.47±2.80	42.06±4.33	36.08±2.38	38.10±1.93	40.59±3.16	42.00±2.27	38.05±3.89	37.77±4.40	38.62±4.98	36.60±4.20	37.11±3.54
CCS	40.00±4.62	39.90±4.70	40.90±3.31	39.90±4.56	57.90±4.12	57.90±4.12	57.60±4.22	53.40±7.82	42.10±5.95	39.90±3.87	35.30±5.36	37.00±3.40	36.50±2.88	36.40±4.40	42.60±4.58	39.60±5.27	43.30±3.65	41.40±4.90	40.90±4.33
Ab	21.64±1.81	21.62±1.86	21.35±1.60	22.10±1.41	24.18±1.55	24.18±1.55	24.18±1.55	24.37±1.47	24.40±1.42	22.86±1.54	21.81±1.17	22.07±1.48	21.83±1.54	21.52±1.46	24.28±1.40	23.01±1.98	22.82±1.02	22.60±1.61	21.74±1.45
Q	22.47±6.54	22.37±6.50	19.81±5.14	20.48±5.55	31.47±4.82	23.99±4.87	24.65±4.45	22.56±4.95	22.65±4.44	22.47±5.49	21.33±4.53	20.47±4.82	20.28±2.64	22.37±5.72	24.55±4.60	22.00±4.42	23.22±4.56	22.47±6.05	21.24±5.74
WW	30.36±2.18	30.32±2.09	29.77±1.95	29.64±2.03	35.87±2.05	31.69±1.82	31.73±1.76	31.56±2.04	31.69±1.68	31.46±2.12	31.93±1.95	31.63±1.89	31.28±1.89	31.58±2.23	31.30±2.30	29.87±2.44	29.44±2.13	30.38±1.64	29.75±1.23
P	19.26±1.91	19.24±1.84	6.01±1.18	7.80±1.45	35.61±1.93	28.69±2.38	22.33±1.87	21.98±1.47	21.14±1.77	11.04±2.20	23.55±2.38	23.61±2.31	19.99±1.75	14.80±2.04	22.35±1.68	10.12±2.47	8.04±1.91	7.23±1.19	7.22±1.07
Mi	4.07±2.15	3.89±2.04	4.44±2.30	7.41±3.55	26.11±4.32	23.15±6.19	13.89±4.05	11.02±3.74	10.09±3.77	11.94±3.61	11.48±2.23	9.54±2.73	8.80±2.32	8.33±2.90	13.70±2.72	11.02±2.67	8.70±3.27	7.87±2.66	7.31±2.97
H+n	41.91±5.96	41.91±5.96	35.15±5.32	39.93±5.56	49.25±4.85	49.25±4.85	49.34±4.87	49.34±4.87	44.06±6.79	45.72±6.26	42.74±6.75	40.26±4.07	39.20±6.40	40.18±5.64	42.99±5.54	40.84±6.88	36.80±7.16	35.23±8.04	28.96±9.20
H+nn	41.99±5.45	41.99±5.45	32.91±5.07	37.95±4.98	48.76±4.78	48.76±4.78	48.76±4.78	48.76±4.78	41.99±8.40	42.58±6.41	46.28±5.53	32.93±8.25	35.98±7.08	37.47±5.10	41.66±4.26	37.87±5.86	20.46±8.14	28.22±12.40	19.63±8.76
Ft	12.23±0.93	12.39±0.90	12.33±0.85	13.56±1.07	33.78±1.62	33.78±1.62	19.81±1.48	15.31±1.14	13.74±0.75	17.13±1.13	12.83±0.81	13.04±0.74	13.37±1.11	12.56±0.82	21.12±1.42	15.46±0.93	13.54±0.80	12.45±0.67	12.57±0.90
Ma	21.00±1.00	21.01±0.93	20.91±0.97	20.94±0.98	26.41±0.97	21.41±0.88	21.45±0.93	21.46±0.92	21.11±0.88	20.95±0.93	20.93±1.04	20.94±1.08	21.01±1.06	20.94±0.96	21.45±0.88	21.38±0.91	21.03±0.91	21.08±0.94	21.01±0.96
E	45.55±1.48	45.55±1.49	43.48±1.36	42.92±0.81	47.26±1.43	47.12±1.39	46.46±1.88	45.23±1.93	44.07±1.45	45.63±1.71	44.64±1.63	44.28±1.36	44.75±1.06	44.51±1.80	44.83±1.50	44.06±1.82	43.40±1.01	42.60±1.19	42.25±0.72
Sk	9.62±0.22	10.18±0.29	10.74±0.21	9.65±0.23	33.97±0.29	33.97±0.29	18.79±0.31	10.68±0.24	9.87±0.23	9.74±0.24	9.61±0.23	9.62±0.23	9.61±0.23	9.61±0.23	9.62±0.23	9.61±0.23	7.99±1.46	7.19±0.75	6.77±0.34
Mu	23.36±1.19	23.28±1.24	19.48±1.12	22.26±1.19	46.07±5.17	39.18±6.37	29.46±2.12	28.22±1.87	26.20±1.46	32.92±2.97	32.11±2.72	32.80±2.15	30.96±1.96	28.13±2.28	28.54±2.03	27.25±2.59	25.40±2.77	23.79±1.66	24.72±1.04
Ha	1.94±0.23	3.11±0.31	3.11±0.33	1.76±0.25	9.73±0.44	2.28±0.21	2.28±0.21	2.29±0.21	2.28±0.20	1.80±0.32	6.28±0.28	4.85±0.21	3.85±0.30	2.72±0.19	2.29±0.23	2.26±0.23	2.14±0.34	1.66±0.19	1.66±0.17
Tw	7.45±0.08	7.45±0.08	4.42±0.11	4.72±0.10	6.63±0.07	6.55±0.08	6.55±0.08	6.25±0.09	5.65±0.14	4.34±0.11	5.80±0.09	5.21±0.15	5.07±0.12	5.07±0.11	5.21±0.19	5.13±0.23	5.10±0.12	4.97±0.12	4.90±0.15

Table 2: Complete results for Table 1 (in main file). Domains ordered following Table 1 (in this SM). Each result is the average + stddev of the classifiers retained at each CV fold. The classifier retained at each fold is the one minimizing the empirical risk among the T, T' boosting iterations.