Lossless or Quantized Boosting with Integer Arithmetic

Richard Nock

CSIRO DATA 61

Australian National University

THE UNIVERSITY OF SYDNEY

Robert C. Williamson

Australian National University

CSIRO DATA 61

The big picture

- Many constraints in today-ML (e.g. for privacy, at-the-edge, distributed or deep)

  — **generic**: integer encoding, small set of operations, quantisation (+ accuracy)

- The shortest path to solutions: **hammering** existing SOTA for new constraints

  — does not go without uncertainty or loss in SOTA guarantees

(may need a **big** hammer)

- Alternative: **“replace current ML algorithms with [*constraint-friendly*] ones”**

  — some great stories in supervised ML start from the same ground, ***a “nice” loss function…***

  (SVM, Boosting, etc.)

cryptographic

(NeurIPS’18 PPML workshop)

…so we created a new loss that fits to the constraints

Statistical properties of a good loss

- Loss for *class probability estimation*: $\ell : \mathcal{Y} \times (0,1) \rightarrow \mathbb{R}$, where $\mathcal{Y} \doteq \{-1,1\}$
- Can be defined from *partial losses* $\ell(y,u) \doteq [\![y=1]\!] \cdot \ell_1(u) + [\![y=-1]\!] \cdot \ell_{-1}(u)$
- Ex: square loss has $\ell_1^{\mathrm{SQ}}(u) \doteq (1/2)\cdot(1-u)^2$ & $\ell_{-1}^{\mathrm{SQ}}(u) \doteq (1/2)\cdot u^2$

- (pointwise) Bayes risk: $\underline{L}(\pi) \doteq \inf_c \mathsf{E}_{\mathsf{Y}\sim\pi}\ell(\mathsf{Y},c)$, **proper** if π in inf
- Ex: square loss has $\underline{L}^{\mathrm{SQ}}(\pi) = (1/2)\cdot\pi(1-\pi)$ (*Gini entropy*), *proper*

(concave)

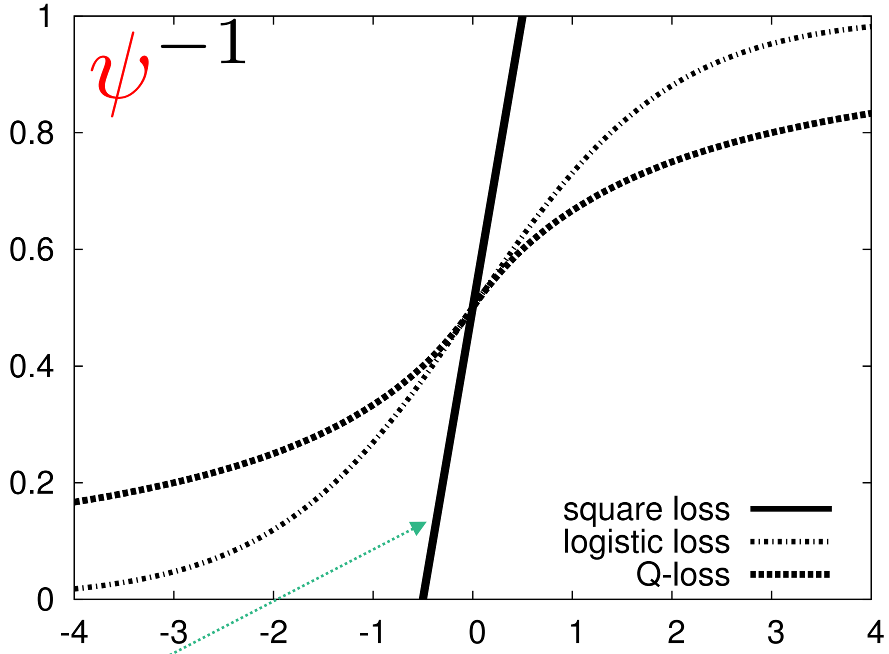
- Real valued classification via a link $\psi : [0,1] \rightarrow \mathbb{R}$ giving $\ell_\psi(y,z) \doteq \ell(y,\psi^{-1}(z))$
- Proper loss **canonical** if link *implicit*, given by $\psi \doteq -\underline{L}'$
- Ex: square loss has $\ell_\psi^{\mathrm{SQ}}(y,z) = (1/4)\cdot(1-2yz)^2$, *proper canonical*

$F(yz)$ (convex)

"surrogate"

Desiderata for our loss, summarised

- 1- Statistics: strictly proper canonical,
- 2- Statistics: link with image spanning the full $\mathbb{R}$,
- 3- Optimisation: F strictly convex 2x differentiable,
- 4- Learning: mirror update $\diamond$ entails +, -, /, *, |.|



$$z \diamond u \leftarrow \psi^{-1}(-z + \psi(u))$$

- Ex: 1 rules out exp loss, 2 rules out square loss, 4 rules out log loss, etc., so no popular loss fits…

The Q-loss

- *partial losses*: The Q-loss is defined from the following partial losses,

$\varepsilon \in (0, 1/2), \varrho > 0$

$$\ell_y^Q(u) \doteq \varrho \cdot \left(\log\left(\frac{u}{\varepsilon}\right) + [\![y=1]\!] \cdot \left(-2 + \frac{1}{u}\right)\right) \qquad u \leq 1/2$$

$$\ell_y^Q(u) \doteq \ell_{-y}^Q(1-u) \qquad u > 1/2$$

Pointwise Bayes risk:

$\mathrm{err}(u) \doteq u \wedge (1-u)$

$$\underline{L}^Q(u) = \varrho \cdot \left(\log\left(\frac{\mathrm{err}(u)}{\varepsilon}\right) + 1 - 2 \cdot \mathrm{err}(u)\right)$$

Surrogate:

$\mathrm{H}(z) \doteq 0 \vee -z$

$$F^Q(z) = -\varrho \cdot \log \varepsilon - \varrho \cdot \log\left(2 + \frac{|z|}{\varrho}\right) + \mathrm{H}(z)$$

Mirror update:

(can be simplified)

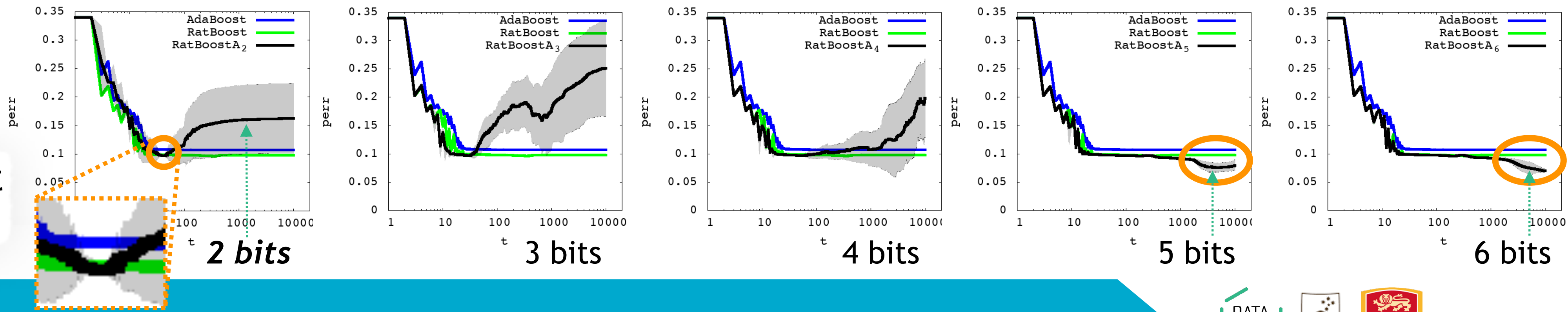
$$z \diamond u = \frac{\varrho \cdot \mathrm{err}(u) + H\left(z \cdot \mathrm{err}(u) + \varrho \cdot (1-2u)\right)}{2\varrho \cdot \mathrm{err}(u) + |z \cdot \mathrm{err}(u) + \varrho \cdot (1-2u)|}$$

Theorem:

The Q-loss fits all four constraints

Learning with the Q-loss

- Boosting 1: ◇ yields a **boosting-compliant** algorithm for F^Q(weak/strong learning)
  - If inputs are rational numbers, the *exact solution* can be computed using *integer arithmetic* (**lossless solution**)
  - Sufficient condition on weight **quantization** to *keep* boosting convergence
- Boosting 2: pointwise Bayes risk $\underline{L}^Q$ yields **optimal** boosting rate for decision trees
- Experiments:



Efficient adaptive weight quantization scheme

Thank you !

(more on achieving lossless boosting by choosing a loss ? poster # 194)