Supplemental Material for "Online Linear Models for Edge Computing"

Hadar Sivan, Moshe Gabel, Assaf Schuster

This file contains proofs and figures that are described in Sections 4 and 5 but are not included there due to lack of space.

Proof of Bounding Prediction of New Model

This is a proof of Lemma 1 from Section 4.2:

Lemma 1. *Let β_1^*, β_2^* and r be as in Theorem 1, and let x be a sample. Then the upper and lower bounds on the prediction of β_2^* for x are:*

$$L(x^T\beta_2^*) := \min_{\beta \in \Omega} x^T\beta = x^T\beta_1^* - x^T r - \|x\|\|r\| \tag{7a}$$

$$U(x^T\beta_2^*) := \max_{\beta \in \Omega} x^T\beta = x^T\beta_1^* - x^T r + \|x\|\|r\|. \tag{7b}$$

Proof. Every vector β in the sphere Ω could be represented as the sum of two vectors: the vector m, which is the center of the sphere, and vector u that starts from the center of the sphere and whose magnitude is bounded by the sphere radius vector ($\|u\| \leq \|r\|$). Therefore, the dot product between β and a given x is

$$x^T\beta = x^T(m+u) = x^T m + x^T u = x^T m + \|x\|\|u\| \cos\left(\angle(x,u)\right).$$

The minimum of the dot product $x^T\beta$, with respect to u, is obtained when $\|u\| = \|r\|$ and $\cos\left(\angle(x,u)\right) = -1$, i.e., u is a vector in the opposite direction of x and with the maximum magnitude under the constraint that u is on the sphere. In this case the lower bound is obtained,

$$L(x^T\beta_{new}^*) = x^T m - \|x\|\|r\|. \tag{8}$$

Using similar arguments, the maximum of the dot product $x^T\beta$ is obtained when $\|u\| = \|r\|$ and $\cos\left(\angle(x,u)\right) = 1$. This time u is in the same direction as x. In this case the upper bound is obtained,

$$U(x^T\beta_{new}^*) = x^T m + \|x\|\|r\|. \tag{9}$$

By substituting $m = \beta_1^* - r$ (from definition of Ω in Section 4.1) in the above expressions of the lower and upper bounds, we obtain (7). □

Figure 4 shows these vectors in two dimensions, the sphere Ω, and the vectors on its surface that yield the maximum and minimum dot product with x.

See Okumura et al. [21] for an alternative derivation of these bounds in a different form.

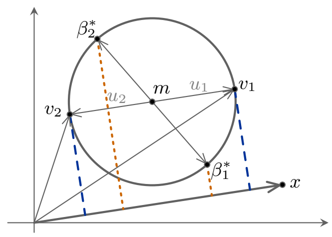

Fig. 4. Illustration of Lemma 1. Vector v_1 is the vector in the circle that maximizes the projection on vector x, while v_2 minimizes the projection on x. The projections of β_1^* and β_2^* on x are always between the projection of v_1, and v_2.

Reanalysis of Okumura et al. [21] Bound to $\|\boldsymbol{\beta_1^*} - \boldsymbol{\beta_2^*}\|$

Okumura et al. suggest in their paper [21] an upper bound to the distance between models $\|\beta_1^* - \beta_2^*\|$. By reanalysis of their bound we show that the new bound we describe in Theorem 1 is tighter.

In [21, Section 2.2], a one-hot vector e_j, $j \in [d]$ where d is the dimension of x, is used to compute the upper and lower bounds of the j^{th} element of the new classifier – $\beta_{2,j}^*$. Then, by [21, Corollary 2]:

$$\|\beta_1^* - \beta_2^*\|_q \leq \left(\sum_{j\in[d]} \max\{\beta_{1,j}^* - L(\beta_{2,j}^*), U(\beta_{2,j}^*) - \beta_{1,j}^*\}^q \right)^{\frac{1}{q}} \tag{10}$$

where $\|\cdot\|_q$ is the L_q norm. The lower and upper bounds, $L(\beta_{2,j}^*)$ and $U(\beta_{2,j}^*)$, are as in (7) for $x = e_j$. Assignment of $x = e_j$ in (7) gives:

$$L(\beta_{2,j}^*) = \beta_{1,j}^* - r_j - \|r\|$$
$$U(\beta_{2,j}^*) = \beta_{1,j}^* - r_j + \|r\|.$$

Therefore:

$$\beta_{1,j}^* - L(\beta_{2,j}^*) = r_j + \|r\|$$
$$U(\beta_{2,j}^*) - \beta_{1,j}^* = -r_j + \|r\|.$$

If $r_j \geq 0$ then $\beta_{1,j}^* - L(\beta_{2,j}^*) \geq \|r\|$, otherwise $U(\beta_{2,j}^*) - \beta_{1,j}^* \geq \|r\|$. Therefore:

$$\max\{\beta_{1,j}^* - L(\beta_{2,j}^*), U(\beta_{2,j}^*) - \beta_{1,j}^*\} \geq \|r\|. \tag{11}$$

Using (11) with (10) gives:

$$\left(\sum_{j\in[d]} \max\{\beta_{1,j}^* - L(\beta_{2,j}^*), U(\beta_{2,j}^*) - \beta_{1,j}^*\}^q \right)^{\frac{1}{q}} \geq \left(\sum_{j\in[d]} \|r\|^q \right)^{\frac{1}{q}} = d^{\frac{1}{q}} \|r\|.$$

In general, for every $d > 2^q$ the bound $\|\beta_1^* - \beta_2^*\| \leq 2\|r\|$ is tighter than (10). Specifically, for L_2 norm, for any $d > 4$ the bound is tighter.

Evaluation Figures

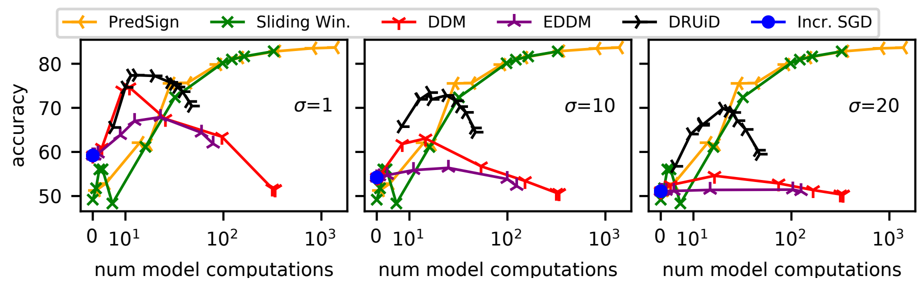

Fig. 5. Sine1+ dataset with 50 attributes and different scale (σ) values. As with 2 attributes, the incremental based algorithms' performance is affected by σ.

Figure 5 shows that the effect of σ does not depend on the number of attributes. See Section 5.4 for description and analysis.