In order to attenuate overfitting, we consider the margin-adjusted loss.

Optimizing Combinatorial and Non-decomposable Metrics with ExactBoost

Let \( \mathcal{L} \) and \( \mathcal{L}^* \) denote the conditional distributions of \( X \) and \( X^* \), respectively. The score estimates \( \hat{s}(x) \) and \( \hat{y} \) are such that

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

Theorem 2. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 2 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 3. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 3 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 4. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 4 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 5. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 5 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 6. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 6 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 7. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 7 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 8. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 8 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.

Theorem 9. Let \( \mathcal{S} \) be the set of weak learners. The KS loss is lowerbounded by

\[
\mathbb{E}[\mathcal{L}(x, \hat{s}(x), \hat{y})] \leq \mathbb{E}[\mathcal{L}(x, s(x), y)]
\]

The result above generalizes to any set of base learners. Theorem 9 states that the empirical risk for any learner \( \mathcal{S} \) is bounded by the expected risk of the optimal learner.