**Abstract**

- Stochastic Gradient Descent (SGD) has been widely studied with classification accuracy as a performance measure.
- These algorithms are not applicable when non-decomposable pairwise performance measures are used, such as Area under the ROC curve (AUC).
- We propose a Variance Reduced Stochastic Proximal algorithm for AUC Maximization (VRSPAM) which converges faster than existing methods.

**Introduction**

- Class imbalance poses a challenge in several domains for instance, medical diagnosis of rare diseases. [1]
- AUC is commonly used to evaluate the performance of a binary classifier in [2] reformulated the pairwise squared loss surrogate of AUC and gave an algorithm with a convergence rate of $O\left(\frac{1}{n}\right)$, under strong convexity.
- This rate is sub-optimal to the linear rate SGD achieves with classification VRSPAM.
- Variance of the gradient in SPAM [2] does not go to zero as it is a stochastic gradient descent based algorithm.

**Algorithm**

Let-

- $C(w; z) = \partial_w F(w, a(w), b(w), \zeta(w); z)$
- $\mu = \frac{1}{n} \sum_{i=1}^{n} C(w; z_i)$
- $v_t = C(w_t; z_{i_t}) - C(w_t; z_{i_t}) + \mu$

**Bounded Variance**

Lemma 1. Consider VRSPAM (Algorithm 1), then the variance of the $v_t$ is upper bounded as:

$$\mathbb{E}[v_t - f(w_t)]^2 \leq 4M^2\beta^2\left[|w_t - w| + (\beta t)^2\left|w - w\right|^2\right]^2$$

- At the convergence, $w = w^*$ and $w_t = w^*$
- Variance of the updates are bounded and go to zero as the algorithm converges
- Variance of the gradient in SPAM [2] does not go to zero as it is a stochastic gradient descent based algorithm

**Convergence Analysis**

**Theorem 1.** Consider VRSPAM (Algorithm 1) and let $w^* = \arg\min_w f(w) + \Omega(w)$; if $\theta > 1$, then there exists $a < 1$ and we have the geometric convergence in expectation:

$$\mathbb{E}[w_t - w^*] \leq a^t\mathbb{E}[w_0 - w^*]$$

- We get a geometric convergence rate of $\alpha^*$ which is much stronger than the $O\left(\frac{1}{n}\right)$ convergence rate obtained in [2].

**Complexity Analysis**

- For any $0 < \theta < 1$ and $E = \frac{1}{1 - \theta}$, if we choose $m = \frac{n}{\ln\left(1 + \frac{1}{\theta - 1}\right)}$, then $\alpha = \frac{2E^2}{\theta}$
- Thus, the time complexity of the algorithm is $O\left(n + \frac{\ln\left(\frac{1}{\theta - 1}\right)}{\theta - 1}\right)$.
- As the order has inverse dependency on $E = -\log_{\theta}\left(\frac{1}{\theta - 1}\right)$, increase in $M$ will result in increase in number of iterations i.e. as the maximum norm of training samples is increased, larger $m$ is required to reach $\alpha$ accuracy.
- SPAM algorithm takes $O\left(\frac{1}{\theta^2}\right)$ iterations to achieve $\alpha$ accuracy. Thus, SPAM has lower per iteration complexity but slower convergence rate as compared to VRSPAM. Therefore, VRSPAM will take less time to get a good approximation of the solution.

**Results**

- Proposed variance reduced stochastic proximal algorithm for AUC maximization (VRSPAM).
- Obtained convergence rate of $O\left(\frac{1}{\theta^2}\right)$, where $\theta > 1$, improving upon state-of-the-art methods [2] which have a convergence rate of $O\left(\frac{1}{n}\right)$.
- Showed theoretically and empirically VRSPAM converges faster than existing methods for AUC maximization.

**Conclusion**

References