

カーネル法における random feature を用いた 確率的勾配法の期待識別誤差の線形収束性

八嶋晋吾¹ 二反田篤史^{1,2} 鈴木大慈^{1,2}

¹ 東京大学大学院情報理工学系研究科

² 理研 AIP

1 はじめに

本研究では、仮説空間を再生核ヒルベルト空間とするカーネル法を用いて、二値識別問題を解く問題設定を考える。二値識別問題では期待識別誤差を最小化する識別器を求める事が究極の目標だが、識別誤差は $\{0, 1\}$ -値関数であり直接的な最適化が困難である。そこで、ロジスティック損失のような一致性を備えた凸損失関数で近似し、代わりにこの損失関数が定める期待損失関数を最小化することで期待識別誤差を小さくすることを試みる。すると、確率的勾配法といった凸最適化手法により効率的に識別機を学習することができる。

一方、実用上においてカーネル法はサンプルサイズに対して計算量がスケールしないという問題を持ち、大規模な問題では近似して計算量を減らす方法がよく用いられる。特に random feature [3] は確率的なサンプリングによりカーネル関数を低ランクな正定値対称関数で近似する手法であり、その実装の簡便さと強固な理論的保証を有することから広く用いられている。本研究ではこの random feature と、同じく大規模問題で広く使われる最適化手法である確率的勾配法を用い、二値識別問題を解くという設定で期待識別誤差の収束について解析を行う。

2 強低ノイズ条件の下での期待識別誤差の線形収束性

確率的勾配法において、期待損失関数の収束率はサンプルサイズに対して一般には劣線形であることが知られており、そのため損失関数の収束を経由して導出した期待識別誤差の収束率も一般には劣線形となる。しかし近年、サンプルの分布が強低ノイズ条件という性質を満たす場合、期待識別誤差については確率的勾配法によって線形収束させられることが示された（二乗損失を用いた場合は [2] により、一般の損失を用いた場合は [1] による）。

本研究ではそれらの結果を random feature による近似を用いて解く場合に拡張する。まず一般の損失関数について、random feature による近似が期待損失最小解に与える影響を誤差解析した結果を紹介する。この結果により、feature 数 M に対して $M^{-1/4}$ のレートで誤差の L^2 ノルムが収束することが示された。さらにその結果を用い、強低ノイズ条件の下での二値識別問題において、十分な数の feature をサンプリングすることで期待識別誤差が線形収束することを示す。また、数値実験において feature 数を増やすことにより実際に期待識別誤差の収束が高速になることを紹介する。

References

- [1] Atsushi Nitanda and Taiji Suzuki. “Stochastic Gradient Descent with Exponential Convergence Rates of Expected Classification Errors”. In: *Artificial Intelligence and Statistics*. 2019, pp. 1417–1426.
- [2] Loucas Pillaud-Vivien, Alessandro Rudi, and Francis Bach. “Exponential Convergence of Testing Error for Stochastic Gradient Methods”. In: *Conference On Learning Theory*. 2018, pp. 250–296.
- [3] Ali Rahimi and Benjamin Recht. “Random features for large-scale kernel machines”. In: *Advances in Neural Information Processing Systems*. 2008, pp. 1177–1184.