NON-SPARSE HIGH-DIMENSIONAL STATISTICS AND ITS APPLICATIONS

MASAAKI IMAIZUMI[1,2]

[1]*The University of Tokyo,* [2] *RIKEN Advanced Intelligence Project*

Abstract. In this talk, we present several results in non-sparse high-dimensional statistics. Specifically, the generalization and Bayesian estimation of high-dimensional linear regression models, statistical inference for high-dimensional generalized linear models, and the regret analysis on contextual bandit problems applying high-dimensional linear models. The analysis in these studies uses the theory of benign overfitting using spectrum, the risk analysis using the convex Gaussian minimax theorem, and the statistical inference using approximate message propagation methods.

1. Outline

1.1. **Linear Regression.** We consider a linear regression problem with p-dimensional covariates and a parameter. Suppose that we observe i.i.d. n pairs $\{(X_i, Y_i)\}_{i=1}^n$ of a covariate $X_i \in \mathbb{R}^p$ and a target variable $Y_i \in \mathbb{R}$ generated from the following linear model with the true parameter $\theta_0 \in \mathbb{R}^p$:

$$Y_i = \langle X_i, \theta_0 \rangle + \xi_i, \ i = 1, ..., n,$$

where ξ_i is a centered noise variable. Let $\Sigma = \mathbb{E}[X_i X_i^\top]$ be a covariance matrix of the covariate.

The goal of this problem is to estimate the parameter θ_0 from the observations. Here, we consider the high-dimensional setting, specifically, we consider $p \gg n$ or where $p = \infty$ regardless of n. Also, we do not impose the assumption of sparsity on the true parameter θ_0. In this setting, the notion of benign overfitting is actively studied.

We investigate whether benign overfitting-like phenomena occur in situations in which we relax key assumptions in this foundational model. [NI22] considers the case where the covariates are dependent in the sample direction and shows that similar benign overfitting occurs depending on the strength of the dependence. [WI23] develops an informative prior distribution that allows for a Bayesian estimator and its distribution approximation that is valid even in non-sparse high dimensions. [TI23] considers that the noise ξ_i is dependent on the covariance and gives conditions for the estimation error convergence, by utilizing the Gaussian minimax inequality.

1.2. **Generalized Linear Regression.** We next consider the generalized linear model (GLM): for a pair (X, Y) of p-dimensional random features X and random responses Y, we consider the following model

$$\mathbb{E}[Y \mid X = x] = g(x^\top \beta), \ \forall x \in \mathbb{R}^p, \tag{1}$$

where $g : \mathbb{R} \to \mathbb{R}$ is an inverse link function that monotonically increases, and $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is an unknown deterministic coefficient vector. Suppose that we observe i.i.d. n pairs $\{(X_i, Y_i)\}_{i=1}^n$ of a feature vector $X_i \in \mathbb{R}^p$ and a target variable $Y_i \in \mathcal{Y}$ that follow the GLM (1), where $\mathcal{Y}$ is a response space, such as $\mathbb{R}, \mathbb{R}_+, \{0, 1\}, \{0, 1, 2, \ldots\}$, and so on.

We consider the proportional high-dimensional regime: we are particularly interested in the proportional limit of the coefficient dimension p and sample size n:

$$n, p \to \infty \text{ and } p/n \to \exists \kappa \in (0, \infty).$$

In this regime, for the logistic regression as the special case, statistical inference on β has been actively studied without the sparsity of β.

[SUI23] develops a methodology for statistical inference for a broad class of GLMs, by deriving the asymptotic normality of an estimator. This method is based on the analysis of a state evolution equation by a vector approximate massage passiong (VAMP) and its application to statistical models.

1.3. **Bandit with Linear Context.** We study a bandit problem with K arms associated with a linear model for its rewards with p-dimensional context vectors. For each round $t \in [T] := \{1, 2, ..., T\}$ and arm $i \in [K]$, we define a context $X_t^{(i)}$ which is a p-dimensional zero-mean sub-Gaussian vector, which is independent among rounds t. An agent chooses an arm $I(t) \in [K]$ based on $X_t^{(i)}$ of all the arms $k \in [K]$, and then observes a reward that follows a linear model as shown in

$$Y^{(I(t))} = \langle X^{(I(t))}, \theta^{I(t)} \rangle + \xi(t).$$

The unknown true parameters $\theta^{(i)}$ for each arm $i \in [K]$ lie in a parameter space $\mathbb{R}^p$, and the independent sub-Gaussian noise $\xi(t)$ with zero mean and variance $\sigma^2 > 0$. We define $i^*(t) := \mathrm{argmax}_{i \in [K]} \langle X_t^{(i)}, \theta^{(i)} \rangle$ as the (ex ante) optimal arm at round t.

Our goal is to design an algorithm that maximizes the total reward, which is equivalent to minimizing the following expected regret. [KI23] considers the high-dimensional setting $p \gg n$ or $p = \infty$ without the sparsity, then derive a novel explore-then-commit strategy to achieve minimize the expected regret.

References

[KI23] Junpei Komiyama and Masaaki Imaizumi. High-dimensional contextual bandit problem without sparsity. *Advances in Neural Information and Processing Systems*, 2023.

[NI22] Shogo Nakakita and Masaaki Imaizumi. Benign overfitting in time series linear model with over-parameterization. *arXiv preprint arXiv:2204.08369*, 2022.

[SUI23] Kazuma Sawaya, Yoshimasa Uematsu, and Masaaki Imaizumi. Feasible adjustments of statistical inference in high-dimensional generalized linear models. 2023.

[TI23] Toshiki Tsuda and Masaaki Imaizumi. Benign overfitting of non-sparse high-dimensional linear regression with correlated noise. *arXiv preprint arXiv:2304.04037*, 2023.

[WI23] Tomoya Wakayama and Masaaki Imaizumi. Bayesian analysis for over-parameterized linear model without sparsity. *arXiv preprint arXiv:2305.15754*, 2023.