STAT 499: Undergraduate Research

Week 2: Recovery in the noiseless setting

Facheng Yu Compiled on: 2024-01-21, 18:00:01

In this lecture, we consider the case where no noise w exists. Suppose that there is some vector $\theta^* \in \mathbb{R}^d$ with at most $s << d$ non-zero entries such that $y = X\theta^*$. Our goal is to solve this system with a sparse vector.

2.1 Basis pursuit linear program

The l_0-norm is defined as

$$\|\theta\|_0 = \sum_{j=1}^{d} \mathbf{1}\{\theta_j \neq 0\}$$

The following problem forms the solution with the fewest non-zero entries:

$$\min_{\theta \in \mathbb{R}^d} \|\theta\|_0 \text{ such that } X\theta = y. \tag{2.1}$$

This problem can be solved by searching all the possible sparsity patterns to have the constraints hold, but it requires at most $\sum_{j=1}^{s} \binom{d}{j}$ to find.

Basis pursuit linear program. To avoid the computational difficulties, a strategy is to replace l_0-norm with a l_1-norm:

$$\min_{\theta \in \mathbb{R}^d} \|\theta\|_1 \text{ such that } X\theta = y. \tag{2.2}$$

This is now a convex optimization, more precisely, a linear program. Now our question is, when is solving (2.2) equivalent to solving (2.1)?

Assume that there is a vector $\theta^* \in \mathbb{R}^d$ whose support is $S \subset \{1, \ldots, d\}$ such that $y = X\theta^*$.

Nullspace. $\mathsf{null}(X) = \{\Delta \in \mathbb{R}^d : X\Delta = 0\}$.

It is natural to see that $\theta^* + \mathsf{null}(X)$ is the feasible space for (2.2).

Tangent cone. $\mathbb{T}(\theta^*) = \left\{\Delta \in \mathbb{R}^d : \|\theta^* + t\Delta\|_1 \leq \|\theta^*\|_1 \text{ for some } t > 0\right\}$.

So if we want the solution of (2.2) to be unique and exactly θ^*, it is equivalent to require that

$$\mathsf{null}(X) \cap \mathbb{T}(\theta^*) = \{0\}. \tag{2.3}$$

Proposition 2.1 *Define* $\mathbb{C}(S) = \{\Delta \in \mathbb{R}^d : \|\Delta_{S^c}\|_1 \leq \|\Delta_S\|_1\}$. *Then, it holds that*

$$\mathbb{T}(\theta^*) \subset \mathbb{C}(S)\}.$$

Proof: For any $\Delta \in \mathbb{T}(\theta^*)$, there exists $t > 0$ such that

$$\|\theta^* + t\Delta\|_1 \leq \|\theta^*\|_1.$$

Thus,

$$0 \leq \|\theta^*\|_1 - \|\theta^* + t\Delta\|_1 = \|\theta^*_S\|_1 - t\|\Delta_{S^c}\|_1 - \|\theta^*_S + t\Delta_S\|_1 \leq t\|\Delta_S\|_1 - t\|\Delta_{S^c}\|_1.$$

That is, $\|\Delta_{S^c}\|_1 \leq \|\Delta_S\|_1$. So $\mathbb{T}(\theta^*) \subset \mathbb{C}(S)\}$. ■

Restricted nullspace property. Based on (2.3), we give the following definition: The matrix X satisfies the restricted nullspace property with respect to S if

$$\mathsf{null}(X) \cap \mathbb{C}(S) = \{0\}. \tag{2.4}$$

If X satisfies the restricted nullspace property, we have the following theorem for the solution uniqueness

Theorem 2.2 *The following two properties are equivalent:*

(a) For any vector $\theta^* \in \mathbb{R}^d$ with support S, the basis pursuit program (2.2) applied with $y = \mathbf{X}\theta^*$ has unique solution $\widehat{\theta} = \theta^*$.

(b) The matrix $\mathbf{X}$ satisfies the restricted nullspace property with respect to S.

Now the question becomes when the restricted nullspace property holds? There are some sufficient conditions for the property being true. Check sec. 7.2.3. to see more details, but it is more complex. Usually we do not consider the noiseless setting.