Report 3

Part I. Sparse Subspace Clustering via l1 regression(lasso)

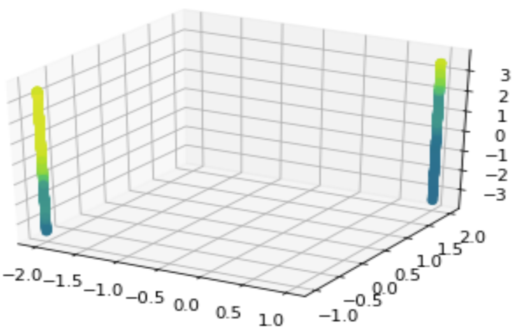
First, do SSC with l1 penalty on original dataset. Simulate 2000 data points as the simplest condition we did before(two parallel lines),

1.do N l1 regression (N=2000)

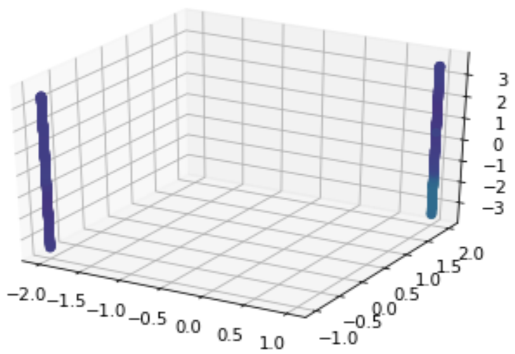
2.formulate adjacency matrix

3.graph laplacian and clustering

In this part, I tried few different α for lasso' s regularization part: $\alpha = 0.1, 0.05$, and first got affinity matrix. I found the matrix really sparse. Only 6 to 8 elements in a column are nonzero. And obtained graph Laplacian matrix for the regression results and plot the original data points colored with the second smallest eigenvector of Laplacian. But the plot was quite bad.



(lasso with 0.1)



(lasso with 0.05)

I am confused with this result. Under this ideal condition, this method should classify data points perfectly. I am still checking the code, could you explain this situation? I am wondering if the parameter does not work well.

I find some reasons for this results. First, the two datasets are not orthogonal so the ideal conditions are actually not satisfied. Second, Lasso regression is a special version of OWL regression and the gap between weights is 0, which means Lasso regression does not satisfy conditions in theorem 1. Are these understandings right?

Part II. OWL norm

I finished reading the paper *Scalable Sparse Subspace Clustering via Ordered Weighted `1 Regression*, and understood the main idea: consider OWL regularizer instead of l1 regularizer to improve the performance of SSC. I understood the basic theory about this method but there exist some details that confuse me.

1. In section 3.3 of this paper, it said *For higher values of affinity, OWL with bigger weight gaps does not provide any benefit over Lasso hence we must set the weight ratio to one*. But according to theorem 1 in this part, larger gap leads to smaller value of right side of inequality in theorem 1. Thus, I think we need to obtain larger gap but not smaller one. So how to understand this sentence?

2. Also in section 3.3 of this paper, I cannot understand the following part:

It is easy to see that $\widehat{\beta} = \underline{0}$, if $\Omega^*_w(X^T y) < 1$ where $\Omega^*_w(\beta)$ is the dual norm of OWL defined later in the section. In order to ensure that the solution is non-trivial we need at least $\bar{w} \leq \|X^T y\|_\infty$ for $\bar{w} = \sum_{j=1}^N w_j/N$. The $\|X^T y\|_\infty$ term scales at most like $\sqrt{(\log N)/d}$ and for OWL-Ramp, $\bar{w} \approx \lambda$. Intuitively, the ℓ_1 component needs to be made small enough to achieve a non-trivial solution.

3. This paper put forward an algorithm

Algorithm 1: OSC with k random seeds

Input : A set of data points $X \in \mathbb{R}^{n \times N}$, $k \in \{1, \dots, N\}$.
Initialize $B = 0_{N \times N}$.
For $i \in \{1, \dots, k\}$,
  Randomly select an index j_i from $[N]$.
  Obtain $\widehat{\beta}$ by regressing $y = X_{\cdot j_i}$ onto the remaining columns of X using the OWL minimization (4).
  Store $B_{\cdot j_i} = \widehat{\beta}$ with $B_{j_i, j_i} = 0$.
Form affinity matrix $W = |B| + |B|^T$.
Apply spectral clustering to the Laplacian of W to obtain a partition.
Output : Subspaces $\{\mathcal{S}_\ell\}_1^L$, Cluster labels.

I have some problems with step6, why affinity is calculated through this way? And are there any recommendation for *k* here? Theoretically, larger *k* performs better but increases the complexity. Small *k* obtains super sparse matrix which may leads to really bad performance.

Part III. Realization of OWL regression

I am very interesting in the results of this paper, but it does not show me details of realization. So I find out another paper *The Ordered Weighted `1 Norm: Atomic Formulation, Projections, and Algorithms,* and am still working on the code.