EQ2415 – Machine Learning and Data Science
HT22

Tutorial 5

A. Honoré, A. Ghosh

1 Sparse Representation

1.1 Norms

Question 1. A norm is used for quantifying a measure of distance between two vectors. But, it should obey certain axioms in order to be called a valid norm. List all such axioms that must be followed by a valid norm.

Question 2. You have encountered the p-norm in the slides $\|\mathbf{x}\|_p$. For $p < 1$, is this still a norm. If not, which property does this norm violate?

1.2 The P_0 problem

Reducing the l_0 norm of a vector arises in problems where we wish to reconstruct a vector $\mathbf{b}$, from a linear combination of a minimum number of columns of a matrix $\mathbf{A}$. This is called a P_0 problem and is one of our principal problem of interest.

$$P_0 : \quad \mathbf{x}^\star = \arg\min_{\mathbf{x}\in\mathbb{R}^m} \|\mathbf{x}\|_0 \quad \text{s.t.} \;\; \mathbf{A}\mathbf{x} = \mathbf{b} \tag{1}$$

where we have an under-determined problem setup with $\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{b} \in \mathbb{R}^n$ and $m > n$. This problem is in fact NP-hard.

Question 3. What is the computational complexity for solving such a problem ? Use a simple numerical reasoning to illustrate your point.

Question 4. Propose two approximate formulations for the problem P_0.

1.3 The spark of a matrix

A special case of P_0 allow us to find a quantity called the *spark*.
Question 5. (a) Define the rank, nullity and spark of a matrix.

Question 5. (b) Consider a matrix that is constructed as $\mathbf{I}_n - \mathbf{S}_n$, where $\mathbf{S}_n$ is a real, skew-symmetric matrix. Calculate the rank, nullity and spark of $\mathbf{I}_n - \mathbf{S}_n$. *Hint:* Use Schur's determinant identity for a block matrix, and the relation that $\det(A + c^T r) = \det(A) + r^T \mathrm{adj}(A)c$

Question 5. (c) Assume that you have an algorithm which is known for giving you the sparsest solution $\mathbf{x}$ for a P_0 problem. It is also assumed that the matrix $\mathbf{A}$ in P_0 is square and full rank. One of your friend also comes and shows you a solution $\mathbf{y}$ for the same problem and claims as well that it is the sparsest. How can you resolve this conflict?

1.3.1 Some useful quantities

1.4 Greedy algorithms for P_0

Finding the spark of a matrix is a combinatorial problem. There exists Greedy algorithms to solve the P_0 problem approximately.

1.4.1 Orthogonal Matching Pursuit

Here, we consider a pre-specified sparsity level of $\mathbf{x}$ is known before the start of the algorithm. We specify this as $\|\mathbf{x}\|_0 = k$. The **support** of $\mathbf{x}$ defined as $\mathcal{S}_{\mathbf{x}} = \{i : x_i \neq 0\}$, The **non-zero elements** of $\mathbf{x}$ can be referred to as $\mathbf{x}_{\mathcal{S}_{\mathbf{x}}}$. Also, recall that the ℓ_0 norm of $\mathbf{x}$ is equal to the cardinality of the support set $\mathcal{S}_{\mathbf{x}}$. We know that the support set is characterized as $\mathcal{S}_{\mathbf{x}} = \{n : x_n \neq 0\}$. So, now the problem is find the k elements of this support set $\mathcal{S}_{\mathbf{x}}$, as we have pre-specified cardinality. A greedy solution avoids a brute-force $\binom{m}{k}$ search and instead tries to find an iterative solution. One such greedy algorithm is the Orthogonal matching pursuit (OMP).

OMP is a greedy solution to the support-finding problem. We are assumed to be given as inputs: $\mathbf{A}$, $\mathbf{b}$ and k_0 (sparsity level of $\mathbf{x}$). OMP consists of the following parts:

1. Initialization
   - Set $k = 0$ (iteration counter)
   - Set initial support set $\mathcal{S}_x^{(0)} = \phi$
   - Set initial residual to be $\mathbf{r}^{(0)} = \mathbf{b}$
   - Set error threshold ε
2. Repeat until either $\|\mathbf{r}^{(k)}\|_2 < \varepsilon$ or max no. of iterations is completed or $\|\mathbf{r}^{(k)}\|_2 > \|\mathbf{r}^{(k-1)}\|_2$
   - Sweep stage: Compute errors $\epsilon(j) = \min_{z_j} \|\mathbf{a}_j z_j - \mathbf{r}^{(k-1)}\|_2^2$ (find the optimal choice) and then finding $i_k^\star = \operatorname{argmin}_j \epsilon(j)$. This can be also done in one single step.
   - Update support $\mathcal{S}_x^{(k)} = \mathcal{S}_x^{(k-1)} \cup i_k^\star$
   - Update residual $\mathbf{r}^{(k)} = \mathbf{b} - \mathbf{A}_{\mathcal{S}_x^{(k)}} \mathbf{A}^\dagger_{\mathcal{S}_x^{(k)}} \mathbf{b}$
   - Update counter $k = k + 1$
3. Finally get $\hat{x} \in \mathbb{R}^N$ with $\hat{\mathbf{x}}_{\mathcal{S}_x^{(k)}} = \mathbf{A}^\dagger_{\mathcal{S}_x^{(k)}} \mathbf{b}$ and remaining part as zeros.

Question 6. Show that the sweep stage is equivalent to finding $i_k^\star = \operatorname{argmax}_j \mathbf{A}^\top \mathbf{r}^{(k-1)}$?