Submodularity of a Set Label Disagreement Function

Toufiq Parag
Janelia Farm Research Campus- HHMI
Ashburn, VA 20147
paragt@janelia.hhmi.org

Abstract

A set label disagreement function is defined over the number of variables that deviates from the dominant label. The dominant label is the value assumed by the largest number of variables within a set of binary variables. The submodularity of a certain family of set label disagreement function is discussed in this manuscript. Such disagreement function could be utilized as a cost function in combinatorial optimization approaches for problems defined over hypergraphs.

1 Introduction

Let $\mathbf{x} = \{x_1, \ldots, x_k\}$ be the binary labels (i.e., $x_i \in \{0, 1\}$) of a set of datapoints $\mathbf{v} = \{v_1, \ldots, v_k\}$ of size k. The dominant label $\rho(\mathbf{x})$ among the labels of $\mathbf{x}$ is defined as the label that largest number of vertices are assigned to. For example, $\rho(\{1, 0, 1, 1\}) = 1$ and $\rho(\{0, 0, 0, 1\}) = 0$. Let us also denote the number of variables to assume a value $c \in \{0, 1\}$ by n_c.

In what follows, we analyze what label disagreement functon g, defined on $k - n_{\rho(\mathbf{x})}$, are submodular. Such function can be utilized in combinatorial approach of hypergraph clustering [1] where the disagreement function g acts as a penalty for number of hyperedge nodes that deviates from the dominant label $\rho(\mathbf{x})$. It may also be exploited as a cost function for Markov Random Fields (MRF) with higher order potentials [2].

Unlike the method presented in Kolmogorov and Zabih [3] for proving submodularity of functions defined over subsets larger than 2, *we do not project* the sets on to pairs of variables and show them to be submodular. The following analysis provides the researchers an alternate approach to do the same exploiting the label arrangement of $\mathbf{v}$ directly.

2 Proposition

For nondecreasing concave g, the label disagreement function $d(\mathbf{x}) = g(k - n_{\rho(\mathbf{x})})$ is submodular.

Proof:

Let $\mathbf{a} = [a_1, \ldots, a_k]^T$ and $\mathbf{b} = [b_1, \ldots, b_k]^T$ be two instantiation of the labels $\mathbf{x}$. Denoting $\vee$ and $\wedge$ as element-wise logical 'or'and 'and' respectively, we need to prove the following for the submodularity of d.

$$d(\mathbf{a}) + d(\mathbf{b}) \geq d(\mathbf{a} \vee \mathbf{b}) + d(\mathbf{a} \wedge \mathbf{b}) \tag{1}$$

Table 1 describes the possible configuration of values in $\mathbf{a}$ and $\mathbf{b}$. The first two rows of Table 1 states that, there are κ_1 zeros among the values of both $\mathbf{a}$ and $\mathbf{b}$; and there are κ_4 ones in both of them. But, values of $\mathbf{a}$ and $\mathbf{b}$ differs in $\kappa_2 + \kappa_3$ places, i.e., there are $\kappa_2 + \kappa_3$ places where $a_i = 1 - b_i$. The rows of $a \vee b$ and $a \wedge b$ in Table 1 show the resulting configuration due to the values in $\mathbf{a}$ and $\mathbf{b}$.

config	κ_1	κ_2	κ_3	κ_4
$\mathbf{a}$	0	0	1	1
$\mathbf{b}$	0	1	0	1
$\mathbf{a} \vee \mathbf{b}$	0	1	1	1
$\mathbf{a} \wedge \mathbf{b}$	0	0	0	1

Table 1: Possible combinations of the values in $\mathbf{a}$ and $\mathbf{b}$. The first row of the table implies that $\kappa_1 + \kappa_2$ values of $\mathbf{a}$ are zeros and $\kappa_3 + \kappa_4$ of them are ones.

Let us examine all possible cases of $\rho(\mathbf{a} \vee \mathbf{b})$ and $\rho(\mathbf{a} \wedge \mathbf{b})$ values using Table 1 and prove that the condition in (1) holds for them.

- **Case** $\rho(\mathbf{a} \vee \mathbf{b}) = 0$ **:** According to Table 1, this case enforces that $\rho(\mathbf{a} \wedge \mathbf{b}) = 0$. Therefore, we need to show the following for (1) to hold.

$$\begin{aligned} & g(\kappa_3 + \kappa_4) + g(\kappa_2 + \kappa_4) \geq g(\kappa_2 + \kappa_3 + \kappa_4) + g(\kappa_4) \\ \Rightarrow \quad & \frac{g(\kappa_2 + \kappa_4) - g(\kappa_4)}{\kappa_2} \geq \frac{g(\kappa_2 + \kappa_3 + \kappa_4) - g(\kappa_3 + \kappa_4)}{\kappa_2}. \end{aligned} \tag{2}$$

  This condition holds only as g is a nondecreasing concave function (i.e., nonincreasing slope).

- **Case** $\rho(\mathbf{a} \wedge \mathbf{b}) = 1$ **:** According to Table 1, this case enforces that $\rho(\mathbf{a} \vee \mathbf{b}) = 1$ and the proof is similar to that of above case.

- **Case** $\rho(\mathbf{a} \vee \mathbf{b}) = 1$ **and** $\rho(\mathbf{a} \wedge \mathbf{b}) = 0$ **:** From Table 1, we can write $d(\mathbf{a} \vee \mathbf{b}) = g(\kappa_1)$ and $d(\mathbf{a} \wedge \mathbf{b}) = g(\kappa_4)$. It is straightforward to show that if $\rho(\mathbf{a}) \neq \rho(\mathbf{b})$, the inequality (1) holds due to the nondecreasing nature of g.

  If we have $\rho(\mathbf{a}) = \rho(\mathbf{b}) = 0$, the condition we need to satisfy is as follows.

$$g(\kappa_3 + \kappa_4) + g(\kappa_2 + \kappa_4) \geq g(\kappa_1) + g(\kappa_4). \tag{3}$$

  The concavity of g gives us,

$$g(\kappa_2 + \kappa_4) + g(\kappa_3 + \kappa_4) \geq g(\kappa_2 + \kappa_3 + \kappa_4) + g(\kappa_2 + \kappa_4) \tag{4}$$

  We know that $\rho(\mathbf{a} \vee \mathbf{b}) = 1$ implies $\kappa_2 + \kappa_3 + \kappa_4 \geq \kappa_1$. Furthermore, due to nondecreasing nature of g, we have $g(\kappa_2 + \kappa_3 + \kappa_1) \geq g(\kappa_1)$. Therefore inequality in (3) holds and d is submodular. Similar proof can be reproduced for $\rho(\mathbf{a}) = \rho(\mathbf{b}) = 1$.

- **Case** $\rho(\mathbf{a} \vee \mathbf{b}) = 0$ **and** $\rho(\mathbf{a} \wedge \mathbf{b}) = 1$ **:** For this case to occur we need $\kappa_1 \geq \kappa_2 + \kappa_3 + \kappa_4$ and $\kappa_4 \geq \kappa_1 + \kappa_2 + \kappa_3$. These conditions will only be true when $\kappa_2 + \kappa_3 = 0$ which implies $\kappa_1 = \kappa_4$. All the possible scenarios can be proved trivially with κ_1 being equal to κ_4. $\square$

References

[1] Kiyohito Nagano, Yoshinobu Kawahara, and Satoru Iwata. Minimum average cost clustering. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1759–1767, 2010.

[2] Daniel Freedman and Petros Drineas. Energy minimization via graph cuts: Settling what is possible. In *CVPR*, 2005.

[3] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, 2004.