Assignment 2 - Optimization & Submodularity

Nikhil Manjrekar

Spring 2023

Contents

1 Question 1

In this question, we need to prove the modularity/submodularity/super-modularity of the given functions. In each of the cases we assume two subsets A, B such that $A \subseteq B$, and use the following inequality for submodularity,

$$f(A \cup \{x\}) - f(A) \geq f(B \cup \{x\}) - f(B) \text{ if } A \subseteq B$$

and the following inqueality for super-modularity,

$$f(A \cup \{x\}) - f(A) < f(B \cup \{x\}) - f(B) \text{ if } A \subseteq B$$

Notations : $E(X, Y)$ = set of edges from set X to set Y.

1. $|I(V_1)|$, where $I(V_1)$ = set of edges with atleast one endpoint in V_1, $V_1 \in V(G)$. - $\boxed{\text{Submodular}}$

   Let there be two sets A and B such that $A \subseteq B \subseteq V(G)$. Since, $A \subseteq B$, we can also say that $I(A) \subseteq I(B)$, since if an edge e is part of $I(A)$, it must have a vertex in A which must also be present in B. We add a vertex $x \in V - B$ to both A and B. Observe that on adding x, some of the edges might already be present in the sets $I(A)$ and $I(B)$.

   Consider the differences, $|I(A \cup \{x\}) - |I(A)|$ and $|I(B \cup \{x\}) - |I(B)|$.

   $$|I(A \cup \{x\}) - |I(A)| = |E(x, V - A)|$$

   $$|I(B \cup \{x\}) - |I(B)| = |E(x, V - B)|$$

   And since, $(V - B) \subseteq (V - A)$, therefore $E(x, V - B) \subseteq E(x, V - A)$. Hence we can write,

   $$|E(x, V - A)| \geq |E(x, V - B)|$$

   $$\boxed{|I(A \cup \{x\}) - |I(A)| \geq |I(B \cup \{x\}) - |I(B)|}$$

   Therefore, $I(V_1)$ is a **submodular** function.

2. $|cut(V_1)|$, where $cut(V_1)$ = set of all edges with only one endpoint in $V_1, V_1 \in V(G)$ - $\boxed{\text{Submodular}}$

   Let there be two sets A and B such that $A \subseteq B \subseteq V(G)$. When the vertex $x \in V(G) - B$ in included in A, we remove all the edges $cut(A)$ with one endpoint on x and add new edges with one endpoint x and the other endpoint in $V - A - \{x\}$. Thus,

   $$|cut(A \cup \{x\})| - |cut(A)| = |E(x, V - A - \{x\})| - |E(x, A)|$$

   Same can be written for subset B. Since, $(A \subseteq B) \implies (V - B - \{x\} \subseteq V - A - \{x\})$. Therefore, $E(x, V - B - \{x\}) \subseteq E(x, V - A - \{x\})$ and $E(x, A) \subseteq E(x, B)$. Thus, we can write the following inequalities,

   $$|E(x, V - A - \{x\})| \geq |E(x, V - B - \{x\})|$$

   $$|E(x, A)| \leq |E(x, B)|$$

   Using the above two inequalities we can finally write

   $$|E(x, V - A - \{x\})| - |E(x, A)| \geq |E(x, V - B - \{x\})| - |E(x, B)|$$

   $$\boxed{|cut(A \cup \{x\})| - |cut(A)| \geq |cut(B \cup \{x\})| - |cut(B)|}$$

   And hence, the function $|cut(.)|$ is **submodular**.

3. $|I(V_1)| + |cut(V_1)|$ - $\boxed{\text{Submodular}}$

   claim: Sum of two submodular function is also submodular, i.e.,

   $$(f+g)(P \cup Q) + (f+g)(P \cap Q) \leq (f+g)(P) + (f+g)(Q)$$

   proof for claim: consider two submodular functions f and g. Therefore by the definition of submodularity, we have,

   $$f(P \cup Q) + f(P \cap Q) \leq f(P) + f(Q)$$
   $$g(P \cup Q) + g(P \cap Q) \leq g(P) + g(Q)$$

   for some sets P and Q. Also we can write,

   $$\begin{aligned}(f+g)(P \cup Q) + (f+g)(P \cap Q) &= (f(P \cup Q) + f(P \cap Q)) + (g(P \cup Q) + g(P \cap Q)) \\ &\leq (f(P) + f(Q)) + (g(P) + g(Q)) \\ &\leq (f+g)(P) + (f+g)(Q)\end{aligned} \tag{1}$$

   hence claim is proved. We can use the above claim to state that, since the function $|I(V_1)|$ and $|cut(V_1)|$ are both submodular, therefore their sum is also going to be **submodular**.

4. $|E_L|, |E_R|$ where $E_L(x) =$ set of all the vertices in V_R adjacent only to vertices in X, $X \in V_L$ $\boxed{\text{Super-modular}}$

   Consider two sets A and B such that $A \subseteq B \subseteq V_L$. Also denote $N(X)$ as the neighborhood of vertices in X such that $N(X) \subseteq V_R$. Consider a vertex $x \in V_L - B$. We can observe that any vertex $y \in N(x)$ will not be part of set $E_L(A)$ and $E_L(B)$, otherwise it would have been against the definition of E_L as $x \notin A, B$. Thus, new elements in $E_L(A \cup x)$ will include vertices that are only connected to vertex x and the vertices that are connected to only vertices in A. Same argument for B. Thus,

   $$|E_L(A \cup x)| - |E(A)| = |E_L(x)| + |Z(x, A)|$$
   $$|E_L(B \cup x)| - |E(B)| = |E_L(x)| + |Z(x, B)|$$

   Here, the function $Z(x, A) \subseteq V_R$ defines the set of vertices in the neighborhood of x such that all the vertices have neigborhood in $A \cup x$. We can also observe that, $Z(x, A) \subseteq Z(x, B)$ because set B has more number of vertices with which vertex in V_R in $N(x$ can have edge to. Thus,

   $$|Z(x, A)| \leq |Z(x, B)|$$
   $$\implies |E_L(x)| + |Z(x, A)| \leq |E_L(x)| + |Z(x, B)|$$
   $$\implies \boxed{|E_L(A \cup x)| - |E(A)| \leq |E_L(B \cup x)| - |E(B)|}$$

   Hence the above function E_L is **supermodular**. Similar is the proof for E_R

2 Question 2

Given that $A, B \subseteq \Omega$, where Ω is the universal set. Also for the monotone submodular function f, submodular mutual information is given as,

$$I_f(A; B) = f(A) + f(B) - f(A \cup B)$$

1. Expression for $I_f(A_1, A_2, \ldots A_k) =?$

   Observe that $I_f(A; B)$ is essentially similar to applying function f on intersection of two sets (set analogy). By using inclusion exclusion principle for intersection (similar to union), we can write

   $$I_f(A_1, A_2 \ldots A_k) = \sum_{i=1}^{n} f(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} f(A_{i_1} \cup A_{i_2}) + \ldots + (-1)^{k+1} f(A_1 \cup A_2 \cup A_3 \ldots A_k)$$

   $$\boxed{I_f(A_1, A_2 \ldots A_k) = \sum_{\Phi \neq J \subseteq \{1,2,\ldots n\}} (-1)^{|J|+1} f(\bigcup_{j \in J} A_j)}$$

2. $I_f(A; B) \geq 0$ and $I_f(A; B|C) \geq 0$

   proof Since f is a submodular function, we can write,

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$$
$$\implies f(A) + f(B) \geq f(A \cup B)$$
$$\implies f(A) + f(B) - f(A \cup B) \geq 0$$
$$\implies \boxed{I_f(A; B) \geq 0}$$

   Hence first part is proved. Using the conditional gain for monotone submodular function we have,

$$f(A|B) = f(A \cup B) - f(B)$$

   Therefore, we can write $I_f(A; B|C)$ as

$$I_f(A; B|C) = f(A|C) + f(B|C) - f(A \cup B|C) \tag{2}$$

   Using the conditional gain we can rewrite the above equation as,

$$\begin{aligned} I_f(A; B|C) &= (f(A \cup C) - f(C)) + (f(B \cup C) - f(C)) - (f(A \cup B \cup C) - f(C)) \\ &= f(A \cup C) + f(B \cup C) - f(A \cup B \cup C) - f(C) \end{aligned} \tag{3}$$

   Again using the submodularity of function f, we can write,

$$\begin{aligned} f(A \cup C) + f(B \cup C) &\geq f(A \cup B \cup C) + f((A \cup C) \cap (B \cup C)) \\ &\geq f(A \cup B \cup C) + f((A \cap B) \cup C) \end{aligned} \tag{4}$$

   Since f is also monotone, therefore $f(Y) \geq f(X)$, if $X \subseteq Y$. Since, $C \subseteq (A \cap B) \cup C$, we ca write,

$$f(A \cup C) + f(B \cup C) \geq f(A \cup B \cup C) + f(C)$$
$$\implies f(A \cup C) + f(B \cup C) - f(A \cup B \cup C) - f(C) \geq 0$$
$$\implies \boxed{I_f(A; B|C) \geq 0}$$

   Hence proved.

3. $min(f(A), f(B)) \geq I_f(A; B) \geq f(A \cap B)$

   proof: Using the submodularity we have,

$$f(A) + f(B) - f(A \cup B) \geq f(A \cap B) \tag{5}$$
$$\implies \boxed{I_f(A; B) \geq f(A \cap B)} \tag{6}$$

   Using the conditional gain $f(A|B) = f(A \cup B) - f(B) \implies f(A \cup B) = f(B) + f(A|B)$. Thus we can write,

$$\begin{aligned} I_f(A; B) &= f(A) + f(B) - (f(B) + f(A|B)) \\ &= f(A) - f(A|B) \end{aligned} \tag{7}$$

   Similarly, $f(A \cup B) = f(A) + f(B|A)$ and therefore above equation can also be written as $I_f(A; B) = f(B) - f(B|A)$. Since, f is monotone and submodular, therefore, $f(B|A)$ and $f(A|B) \geq 0$. Therefore, we have $I_f(A; B) \leq f(A)$ and $I_f(A; B) \leq f(B)$. Combining both the results, we have,

$$\boxed{I_f(A; B) \leq min(f(A), f(B))} \tag{8}$$

   Using inequality 6 and 8 we can write the required results.
   Clarification: $f(A|B) \geq 0$ because, $f(A \cup B) \geq f(B) \implies f(A|B) \geq 0$ by monotonicity.

4. $min(f(A|C), f(B|C)) \geq I_f(A; B|C) \geq f(A \cap B|C)$

   proof: In order to prove this, we can simply use the result from previous part, but in order to do that we to show that the function $g(A) = f(A|C)$ is also submodular for fixed set C. By monotonicty of f, we can write
   $$f(A \cup C) \geq f(C) \implies f(A|C) \geq 0 \implies g(A) \geq 0$$
   Further note that, in $g(A) = f(A \cup C) - f(C)$, $f(A \cup C)$ is submodular because of the conditioning property that union with fixed set is also submodular. Also since, C is fixed, $f(C)$ will also be fixed. Hence, the function $g(A)$ is also submodular. Therefore, $g(A)$ has similar properties has $f(A)$, therefore, results for previous part can be rewritten as,
   $$min(g(A), g(B)) \geq I_f(A; B|C) \geq g(A \cap B)$$

5. $f(A) - \sum_{j \in A-B} f(A; B) \leq I_f(A; B) \leq f(A) - \sum_{j \in A-B} f(j|\Omega - j) \leq f(A)$

   Proof: Let's start with the lower bound first. Using **second order partial derivatives** (class slides) we have,
   $$f(i|X \cup j) \leq f(i|X)$$
   for some sets X,i,j. We can use this in our proof. So, let's assume that $A - B = \{a_1, a_2, \ldots, a_k\}$. Then, we can have,
   $$\begin{aligned} &f(a_1|B) \geq f(\{a_1\}|B) = f(\{a_1\} \cup B) - f(B) \\ &f(a_2|B) \geq f(\{a_2\}|B \cup \{a_1\}) = f(\{a_2\} \cup \{a_1\} \cup B) - f(\{a_1\} \cup B) \\ &\qquad\vdots \\ &f(a_k|B) \geq f(\{a_k\}|B \cup \{a_{k-1}\} \cup \ldots \{a_1\}) = f(\{a_k\} \cup \{a_{k-1}\} \ldots \{a_1\} \cup B) - f(\{a_{k-1}\} \ldots \{a_1\} \cup B) \end{aligned} \tag{9}$$
   Summing up on both the sides of the inequality we will have,
   $$\sum_{i=1}^{k} f(\{a_i\}|B) \geq f(\{a_k\} \cup \{a_{k-1}\} \ldots \{a_1\} \cup B) - f(B) \tag{10}$$
   Observe that the quantity $\{a_k\} \cup \{a_{k-1}\} \ldots \{a_1\} \cup B = A \cup B$, Hence,
   $$\begin{aligned} \sum_{i=1}^{k} f(\{a_i\}|B) &\geq f(A \cup B) - f(B) \\ &\geq f(A|B) \end{aligned} \tag{11}$$
   Also it is known that $I_f(A; B) = f(A) - f(A|B) \implies f(A|B) = f(A) - I_f(A; B)$. we can use this in the above inequality as,
   $$\begin{aligned} &\sum_{i=1}^{k} f(\{a_i\}|B) \geq f(A) - I_f(A; B) \\ &\implies I_f(A; B) \geq f(A) - \sum_{i=1}^{k} f(\{a_i\}|B) \end{aligned} \tag{12}$$
   Now, for the upper bound, observe that $B \cup \{a_{k-1}\} \cup \{a_{k-2}\} \ldots \cup \{a_1\} \subseteq \Omega \setminus \{a_k\}$ (Taking advantage of notation in the upper bound to be proved since a_k is present in neither of the sets). From sets on inequalities in 13 we can have
   $$\begin{aligned} &f(a_1|B) \geq f(\{a_1\}|B) \geq f(\{a_1\}|\Omega \setminus \{a_1\}) \\ &f(a_2|B) \geq f(\{a_2\}|B \cup \{a_1\}) \geq f(\{a_2\}|\Omega \setminus \{a_2\}) \\ &\qquad\vdots \\ &f(a_k|B) \geq f(\{a_k\}|B \cup \{a_{k-1}\} \cup \ldots \{a_1\}) \geq f(\{a_k\}|\Omega \setminus \{a_k\}) \end{aligned} \tag{13}$$

Summing up all the inequalities we will have,

$$\sum_{i=1}^{k} f(\{a_i\}|B) \geq f(A|B) \geq \sum_{i=1}^{k} f(\{a_i\}|\Omega \setminus \{a_i\}) \tag{14}$$

Using the right inequality and $f(A|B) = f(A) - I_f(A;B)$, we will have,

$$\boxed{f(A) - \sum_{i=1}^{k} f(\{a_i\}|\Omega \setminus \{a_i\} \geq I_f(A;B)}$$

3 Question 3

1. Given that $y = [y_1\ y_2\ \ldots y_n]^T, a = [\beta_0, \beta_1]^T, B = [1\ x_1; 1\ x_2; \ldots; 1\ x_n]$ and $g(x) = diag(w_1(x), w_2(x), \ldots, w_n(x))$ Thus,

$$(\mathbf{y} - \mathbf{Ba}) = \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \right) = \left(\begin{bmatrix} y_1 - \beta_0 - \beta_1 x_1 \\ y_2 - \beta_0 - \beta_1 x_2 \\ \vdots \\ y_n - \beta_0 - \beta_1 x_n \end{bmatrix} \right)$$

also $g(x)(y - Ba)$ will be given as,

$$g(x)(\mathbf{y} - \mathbf{Ba}) = diag(w_1(x), w_2(x) \ldots w_n(x)) \begin{bmatrix} y_1 - \beta_0 - \beta_1 x_1 \\ y_2 - \beta_0 - \beta_1 x_2 \\ \vdots \\ y_n - \beta_0 - \beta_1 x_n \end{bmatrix} = \begin{bmatrix} w_1(x)(y_1 - \beta_0 - \beta_1 x_1) \\ w_2(x)(y_2 - \beta_0 - \beta_1 x_2) \\ \vdots \\ w_n(x)(y_n - \beta_0 - \beta_1 x_n) \end{bmatrix}$$

hence we can observe that the product $(\mathbf{y} - \mathbf{Ba})^T g(x)(\mathbf{y} - \mathbf{Ba})$ gives the equation for the minimization as given in question

2. In order to find $\hat{\beta}_0$ and $\hat{\beta}_1$ we need to differentiate the formulation we obtained above wrt $\mathbf{a} = [\beta_0\ \beta_1]^T$. We will use the following property,

$$\boxed{\frac{\partial \mathbf{x^T A x}}{\partial \mathbf{x}} = \mathbf{x^T}(\mathbf{A^T} + \mathbf{A})}$$

Therefore, we have,

$$\begin{aligned} \frac{\partial (\mathbf{y} - \mathbf{Ba})^T g(x)(\mathbf{y} - \mathbf{Ba})}{\partial \mathbf{a}} &= (\mathbf{y} - \mathbf{Ba})^T (2 * g(x)) \frac{\partial (\mathbf{y} - \mathbf{Ba})}{\partial \mathbf{a}} \\ &= -2(\mathbf{y} - \mathbf{Ba})^T g(x)\mathbf{B} \end{aligned}$$

for minima, above differential will be 0, thus,

$$\mathbf{y^T} g(x)\mathbf{B} - \mathbf{a^T B^T} g(x)\mathbf{B} = 0$$

$$\implies \boxed{\mathbf{a^T} = \mathbf{y^T} g(x)\mathbf{B}(\mathbf{B^T} g(x)\mathbf{B})^{-1}}$$

Observe that $\mathbf{B^T} g(x)\mathbf{B}$ will be square invertible matrix. Also we can write $\hat{f}(x) = \mathbf{\hat{a}^T}[1\ x]^T$, thus,

$$\hat{f}(x) = \mathbf{y^T} g(x)\mathbf{B}(\mathbf{B^T} g(x)\mathbf{B})^{-1} \begin{bmatrix} 1 \\ x \end{bmatrix}$$

$$\implies \boxed{\hat{f}(x) = \mathbf{y^T h(x)}}$$

where $\mathbf{h}(\mathbf{x}) = g(x)\mathbf{B}(\mathbf{B^T}g(x)\mathbf{B})^{-1}\begin{bmatrix}1\\x\end{bmatrix}$ which is $n \times 1$ dimensional column vector and its each element is $h_i(x)$ which is a scalar. Therefore, we can write the above equation as

$$\boxed{\hat{f}(x) = \sum_{i=1}^{n} h_i(x)y_i}$$

4 Question 4

1. Part1

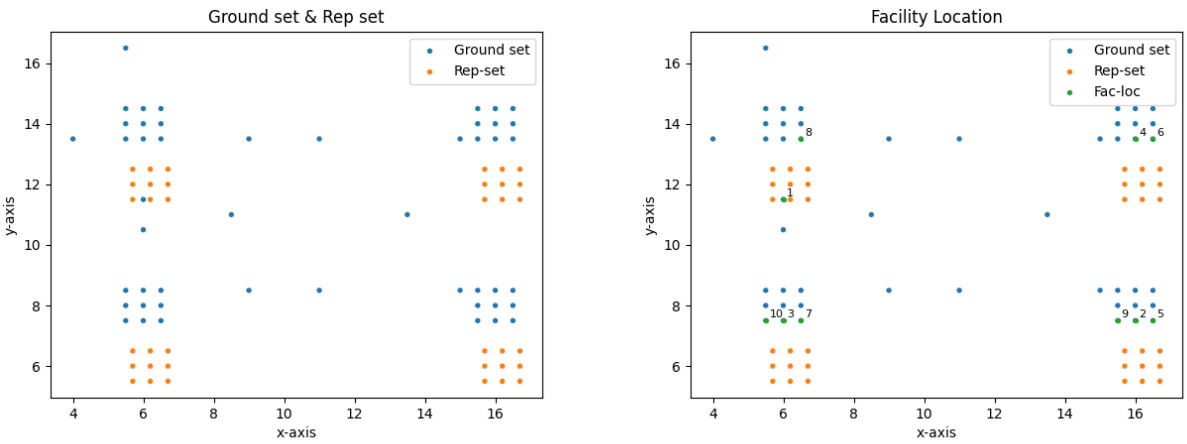

Figure 1: Ground-Set, Rep-Set and Representative Subset chosen using Facility Location, Graph Cut with $\lambda = -3, 0, 3$ in the order and Disparity Sum Function

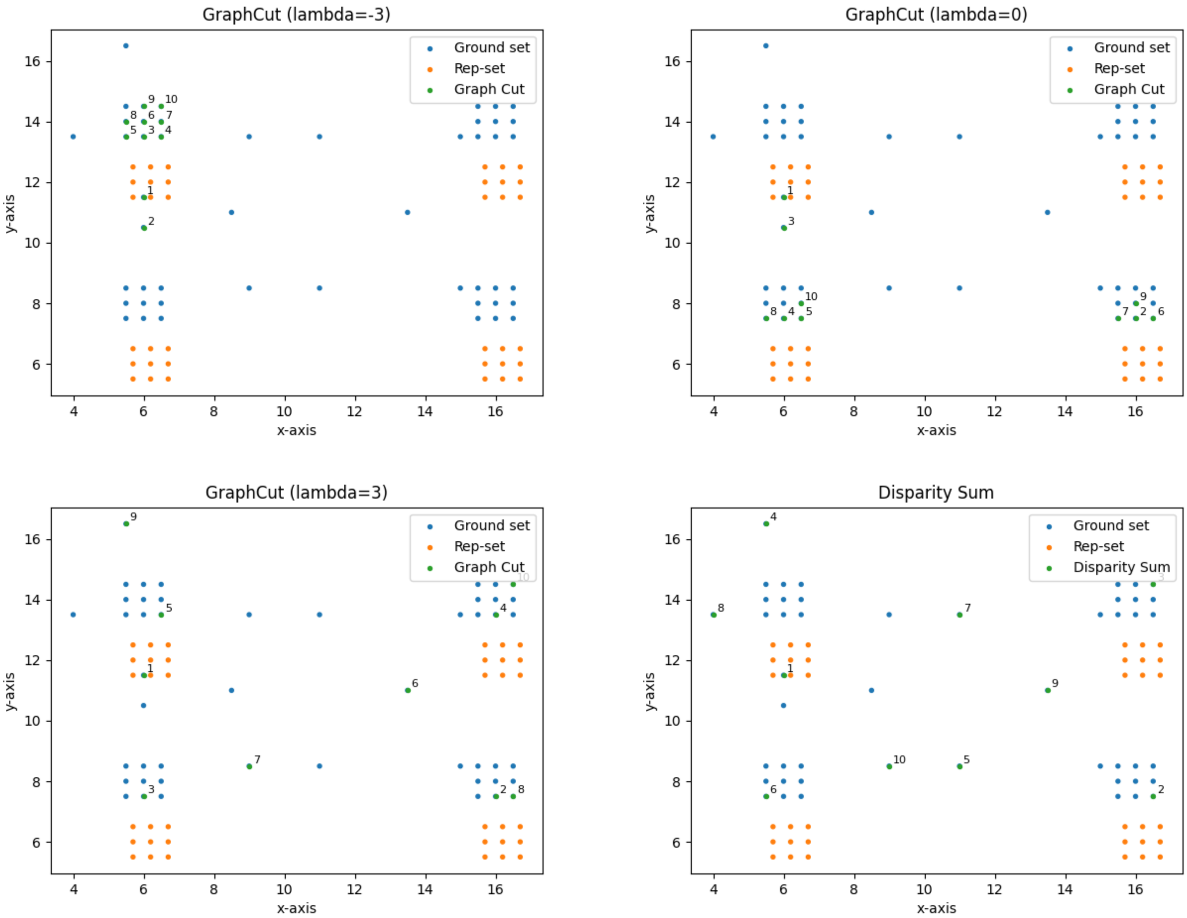

Figure 2: Ground-Set, Rep-Set and Representative Subset chosen using Facility Location, Graph Cut with $\lambda = -3, 0, 3$ in the order and Disparity Sum Function

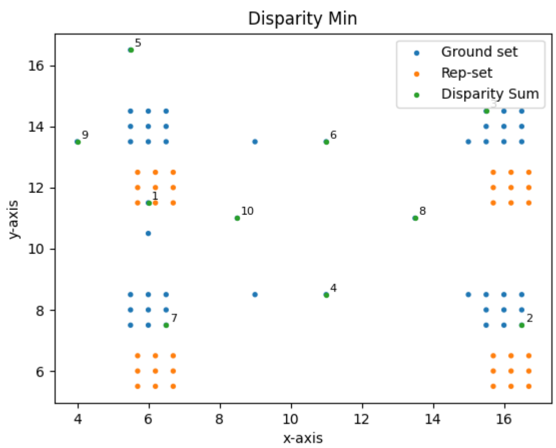

Figure 3: Representative Subset chosen using Disparity Min function

2. Part 2

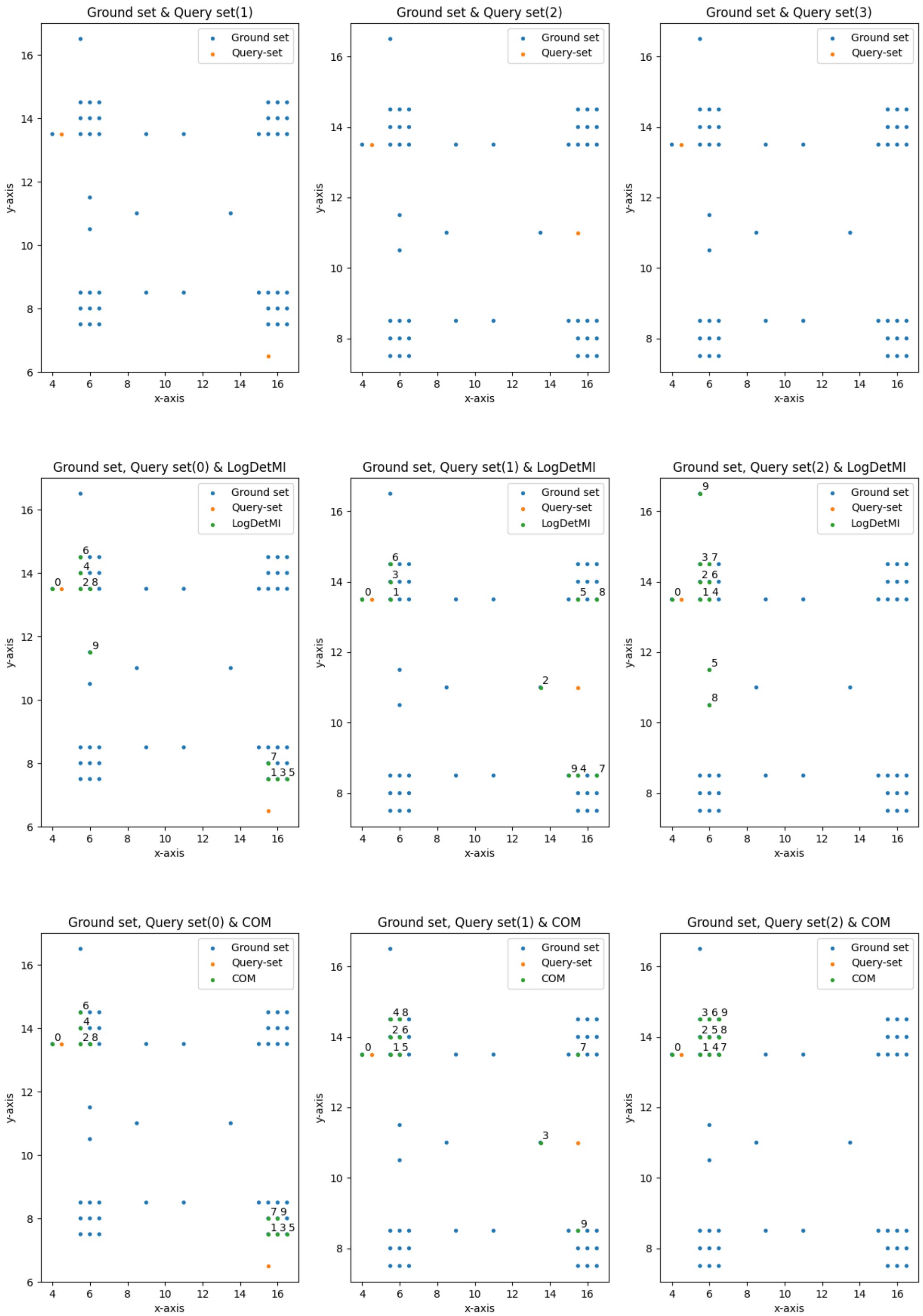

Figure 4: Ground-Set, Query-Set for three query sets is represented above. The first two query sets had multiple queries (2) while last one had only single query

5 Question 5

For the given `prox` function, for different range of values of z, $f_x(z) = min(\frac{1}{2\gamma}||x - z||^2 + h(x))$ which we are trying to minimise will be given as,

$$f(z) = \begin{cases} min(\frac{1}{2\gamma}||z||^2, 100) & z \leq 0 \\ 0 & 0 \leq z \leq \theta \\ min(\frac{1}{2\gamma}||\theta - z||^2, 100) & \theta \leq z \end{cases} \tag{15}$$

The above function is calculated on certain values of x.

1. **Case 1:** When $0 \leq z \leq \theta$

   In this case, $f_x(z) = min(\frac{1}{2\gamma}||x - z||^2 + h(x))$. One can see that if $0 \leq x \leq \theta$, then clearly $f_x(z)$ minimizes to $f_x(z) = \frac{1}{2\gamma}||x - z||^2$, since $h(x) = 0$ in this range of x. Also for $x = z$ in this range, $f_x(z) = 0$. Therefore,

   $$\boxed{x = z, 0 \leq z \leq \theta}$$

2. **Case 2:** When $z \geq \theta$

   In this case if $0 \leq x \leq \theta$, then $f_x(z) = min_x(\frac{1}{2\gamma}||x - z||^2)$. Also observe that the norm in the function will only be minimised for $x = \theta$. But for $x \geq \theta$, $f_x(z) = min_x(\frac{1}{2\gamma}||x - z||^2 + 100)$ which can only be minimized for $x = z$ in this range. But if we see both the values collectively, then the function will become $f_x(z) = min(\frac{1}{2\gamma}||\theta - z||^2, 100)$. Therefore, we again have two cases:-

   (a) $f_x(z) = \frac{1}{2\gamma}||\theta - z||^2$, then

   $$\frac{1}{2\gamma}||\theta - z||^2 \leq 100$$
   $$\implies ||\theta - z||^2 \leq 200\gamma$$
   $$\implies -\sqrt{200\gamma} + \theta \leq z \leq \sqrt{200\gamma} + \theta$$

   Intersecting the above interval with range of z in this case, we have

   $$\boxed{x = \theta,\ if\ \theta \leq z \leq \sqrt{200\gamma} + \theta}$$

   (b) $f_x(z) = 100$, then

   $$100 \leq \frac{1}{2\gamma}||\theta - z||^2$$
   $$\implies 200\gamma \leq ||\theta - z||^2$$
   $$\implies \sqrt{200\gamma} + \theta \leq z$$

   Therefore, for this case we have,

   $$\boxed{x = z,\ if\ \sqrt{200\gamma} + \theta \leq z}$$

3. **Case 3:** When $z \leq 0$

   This case is similar to *case 2*, only difference being that $f_X(z) = min(\frac{1}{2\gamma}||z||^2, 100)$ with same logic as *case 2*. Therefore again we will have two cases

(a) $f_x(z) = min(\frac{1}{2\gamma}||z||^2$, then

$$\frac{1}{2\gamma}||z||^2 \leq 100$$

$$\implies ||z||^2 \leq 200\gamma$$

$$\implies -\sqrt{200\gamma} \leq z \leq \sqrt{200\gamma}$$

Intersecting the above interval with range of z in this case, we have

$$\boxed{x = 0,\ if\ \ -\sqrt{200\gamma} \leq z \leq 0}$$

(b) $f_x(z) = 100$, then

$$100 \leq \frac{1}{2\gamma}||\theta - z||^2$$

$$\implies 200\gamma \leq ||z||^2$$

Also intersecting with the interval of this case we will have,

$$\implies z \leq -\sqrt{200\gamma}$$

Therefore, for this case we have,

$$\boxed{x = z,\ if\ z \leq -\sqrt{200\gamma}}$$

Therefore, combining results of all the three cases, we will have,

$$prox_h(z) = \begin{cases} x = z & z \leq -\sqrt{200\gamma} \\ x = 0 & -\sqrt{200\gamma} \leq z \leq 0 \\ x = z & 0 \leq z \leq \theta \\ x = \theta & \theta \leq z \leq \sqrt{200\gamma} + \theta \\ x = z & \sqrt{200\gamma} + \theta \leq z \end{cases}$$

6 Question 6

Projected Gradient Descent algorithm is given as follows,

$$x_{t+1} = x_t - \alpha_t \nabla f(x_t)$$

$$x_{t+1} = P_C(x_{t+1})$$

We are given the following optimization problem,

$$\min_{x \in R} f(\mathbf{x})$$

$$subject\ to\ \mathbf{x} \leq r\ and\ \mathbf{x} \geq l$$

where $P_C(z) = argmin_x \frac{1}{2}||x - z||^2$

Since the function to be minimized is convex function in $P_c(x)$, therefore, we can use the KKT conditions to solve the problem, which are given as,

1. $\frac{1}{2}\nabla_x ||x - z||^2 + \nabla_x \lambda_1(x - r) + \nabla_x \lambda_2(l - x) = 0$
2. $l - x \leq 0$
3. $x - r \leq 0$
4. $\lambda_1(l - x) = 0$

5. $\lambda_2(x-r)=0$
6. $\lambda_1 \geq 0$
7. $\lambda_2 \geq 0$

Above all the conditons imply optimality at the $l \leq x \leq r$ Solving the first condition we will get,

$$(x-z)-\lambda_1+\lambda_2=0 \tag{16}$$

$$x=z+\lambda_1-\lambda_2 \tag{17}$$

Now we will consider two cases,

1. **Case 1:** When $\mathbf{l} < \mathbf{x} < \mathbf{r}$,

   In this case, the condition 2 and 4 above will imply, $\lambda_1 = 0$ and $\lambda_2 = 0$. So by using equation 21, we will have $x = z$. For this to hold z must lie in interval (l, r).

2. **Case 2:** If the first case doesn't hold then only two options we have are $x = l$ or $x = r$, If $x = l$, then $\lambda_2 = 0$ according to condition 5 and $x = z + \lambda_1 - \lambda_2 \implies z = l - \lambda_1 + \lambda_2 \implies z \leq l$. Similarly, for the case $x = r$

Therefore, the above solution can be summarized as,

$$P_C(z) = \begin{cases} l & z \leq l \\ z & l < z < r \\ r & z \geq r \end{cases}$$

In short this can be written as, $P_C(z) = \min(\max(z,l),r)$. Hence the complete optimization step can be given as, $x_{t+1} = \min(\max(z, x_{t+1}), r)$

7 Question 7

We can write $I_f(A;B;C)$ as,

$$\begin{aligned} I_f(A;B;C) &= f(A)+f(B)+f(C)-f(A\cup B)-f(A\cup C)-f(B\cup C)+f(A\cup B\cup C) \\ &= I_f(A;B)+I_f(C;B)-f(B)-f(A\cup C)+f(A\cup B\cup C) \\ &= I_f(A;B)+I_f(C;B)-I_f(A\cup C;B) \end{aligned}$$

We can see that if function, $I_f(X;B)$ is submodular for a fixed set B then, we can write,

$$I_f(A;B)+I_f(C;B) \geq I_f(A\cup C;B)+I_f(A\cap C;B)$$

$$\implies I_f(A;B)+I_f(C;B)-I_f(A\cup C;B) \geq I_f(A\cap C;B)$$

$$\implies I_f(A;B;C) \geq I_f(A\cap C;B)$$

Since, $I_f(X;B) \geq 0$ because of submodularity, thus

$$\implies \boxed{I_f(A;B;C) \geq 0}$$

Similarly, we can see that if $I_f(X;B)$ is supermodular for fixed set B. Then, we can write,

$$I_f(A;B)+I_f(C;B) \leq I_f(A\cup C;B)+I_f(A\cap C;B)$$

$$\implies I_f(A;B)+I_f(C;B)-I_f(A\cup C;B) \leq I_f(A\cap C;B)$$

Since, $I_f(X;B) \leq 0$ because of submodularity, thus

$$\implies \boxed{I_f(A;B;C) \leq 0}$$

8 Question 8

False, it is not true for all k. We saw in the above example in question 2, that it is true for k=2. However, consider the following monotone submodular function, we do not have the inequality to be satisfied:

$$f(A) = \min(|A|, 6)$$

Clearly, $f(A)$ is monotonic, as well as submodular. We now consider the 5-way mutual information given by,

$$I_f(A, B, C, D, E)$$

let A, B, C, D, E be the disjoint sets such that $|A| = |B| = |C| = |D| = 2, |E| = 1$. RHS is given by $\min f(A), f(B), f(C), f(D), f(E) = 1$.
I_f term has 5 ($\binom{4}{1} + 1$) terms involving singletons: $f(A)$, 10 ($\binom{4}{2} + \binom{4}{1}$) terms like $f(A \cup B)$, 10 ($\binom{4}{3} + \binom{4}{2}$) terms like $f(A \cup B \cup C)$, 5 ($\binom{4}{4} + \binom{4}{3}$) terms like $f(A \cup B \cup C \cup D)$ and 1 ($\binom{1}{1}$) term $f(A \cup B \cup C \cup D \cup E)$. Thus

$$\begin{aligned} I_f(A; B; C; D; E) &= (4(2) + 1) - (6(4) + 4(3)) + (4(6) + 6(5)) - (1(6) + 4(6)) + 6 \\ &= 3 \geq 1 = \min f(A), f(B), f(C), f(D), f(E) = 1. \end{aligned}$$

Hence, this is against the given statement. Therefore, the given relation does not hold for all the values of k.