CS769 (March 2023) - Optimization in Machine Learning
Assignment 2

1 Assignment Policies

The following are the policies regarding this assignment.

1. This assignment needs **to be done individually** by everyone. You are expected to work on the assignments on your own. Plagiarism leads to disciplinary action as per DPAC policies.
2. Use Latex (online platforms such as overleaf) for writing theory questions.
3. Submit the assignment as a zip file containing latex source, and PDF file.
4. Maximum Marks is for 50 marks.
5. Assignment is due on 17th April 2023

Questions

Question 1 . (8 marks)

Let there be a graph G with V vertices and E edges. Comment on the modularity of the following functions and give reasons.

1. Let $I(V_1)$ = set of edges with at least one end point in V_1, $V_1 \in V(G)$. Then $|I|$ is?
2. Let $cut(V_1)$ = set of all branches with only one endpoint in V_1, $V_1 \in V(G)$. Then $|cut|$ is?
3. $|I(V_1)| + |cut(V_1)|$ is?
4. Let $B = (V_L, V_R, E)$. Let $E_L(X)$ = set of all vertices in V_R adjacent only to vertices in X, $X \in V_L$. E_R is defined similarly on subsets of V_R. Then $|E_L|$, $|E_R|$ are?

Question 2. (1+2x4=9 marks)
Let Ω be a universal set and $A, B \subseteq \Omega$. For a monotone submodular f, submodular mutual information between the sets A, B denoted by $I_f(A; B)$ is defined as $I_f(A; B) \triangleq f(A) + f(B) - f(A \cup B)$.

1. What would the expression for of $I_f(A_1; A_2; \ldots; A_k)$?

2. Show that $I_f(A; B) \geq 0$ and $I_f(A; B|C) \geq 0$
3. Show that $\min(f(A), f(B)) \geq I_f(A; B) \geq f(A \cap B)$
4. Show that $\min(f(A|C), f(B|C)) \geq I_f(A; B|C) \geq f(A \cap B|C)$
5. Show that $I_f(A; B)$ can be lower bounded by $f(A) - \sum_{j \in A \setminus B} f(j|B) \leq I_f(A; B)$ and upper bounded by $I_f(A; B) \leq f(A) - \sum_{j \in A \setminus B} f(j|\Omega \setminus j) \leq f(A)$. Are upper/lower bounds submodular?

Question 3. (4 marks)

We have a regression function for the variable x, given by $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$. There are n instances. The coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained by solving the given optimization problem -:

$$\hat{\beta}_0, \hat{\beta}_1 = \text{argmin}_{\beta_0, \beta_1} \sum_{i=1}^{n} w_i(x)(y_i - \beta_0 - \beta_1 x_i)^2 \tag{1}$$

where the weight factors $w_i(x)$ do not depend on either of the coefficients.

1. Show that the above function can be written in the form -:

$$(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})^\top g(x)(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}) \tag{2}$$

   where $\boldsymbol{y} = [y_1 \ y_2 \ \cdots \ y_n]^\top$, $\boldsymbol{a} = [\beta_0 \ \beta_1]^\top$, $\boldsymbol{B} = [1 \ x_1; \ 1 \ x_2; \ \cdots \ 1 \ x_n]$, and $g(x)$ is a diagonal matrix whose diagonal entries are the weights $w_i(x)$.

2. Using the formulation above, show that $\hat{f}(x)$ is a linear combination of $\boldsymbol{y}$. In other words, $\hat{f}(x) = \sum_{i=1}^{n} h_i(x) y_i$, where $h_i(x)$ is some function over x. Clearly mention what $h_i(x)$ is.

Question 4. (2x6=12 marks)

In this problem, we will explore the various types of submodular functions by plotting the data against the optimal set obtained by greedy inference. You are encouraged to use libraries such as submodlib for submodular functions and matplotlib for plotting. You may access the text files here.

1. In this part of the problem, you are provided with two sets of files - 'gset_1.txt' and 'rep.txt'. The former contains the ground set of points (whose subset we want) and the latter contains the target set of points (whose representation we want). Each line of a file represents an (x, y)-coordinate.

   (a) First, plot the data from both sets of files using differing colors. Legends and axis should be present and labelled properly.

   Then, plot the observations using the functions given below for optimal set budget of size 10. Use the naive greedy algorithm and make sure to plot points from the optimal set obtained in a different color. Report the order in which the points were picked up by your algorithm -:

   (b) Facility Location

(c) Graph Cut with varying parameter λ

(d) Set Cover

(e) Disparity Sum

2. In the next part of the problem, you will explore mutual-information based submodular functions for query-focused summarization. In this case, you are given ground set 'gset_2.txt' and 'qset.txt' (which consists of queries). Note that here is no overlap between the query points and the ground set data points. Multiple queries may be given on a single line, in which case they need to be considered by the function jointly.

   (a) Plot the ground set and query set data using different colors.

   Further, plot the summarization results obtained using the following submodular mutual information functions in a similar manner to the previous part. The optimal set budget continues to be 10 -:

   (b) Log-determinant Mutual Information

   (c) Concave over Modular

Question 5. (3 marks)

Recall that the Prox operator of a function $\underline{h}$ for some argument z is

$$\text{prox}_h(z) = \text{argmin}_x \frac{1}{2\gamma}\|x - z\|^2 + h(x).$$

Compute the Prox operator for $h(x)$ defined as $h(x) = 0$ if $0 \leq x \leq \theta$ (for some fixed $\theta > 0$) and $h(x) = 100$ for any other value of x.

Question 6. (5 marks)

Often, in optimization problems in machine learning, we have a simple constraint requiring that parameters lie in a particular interval. Such optimization problems can be effectively solved using the projected gradient descent algorithm discussed in the class.

Derive the exact projection operation of the projected gradient descent algorithm, for the following optimization problem which has the simplest form of such an interval constraint:

$$\begin{array}{ll} \min_{\mathbf{x}\in\Re} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \leq r \text{ and } \quad \mathbf{x} \geq l \end{array}$$

for some fixed given scalars $l < r \in \Re$ and for some machine learning convex loss function $f(\mathbf{x})$. You can use the Karush-Kuhn-Tucker (KKT) conditions for deriving the projection step.

Question 7. (3 marks)

Show that the 3-way submodular mutual information $I_f(A; B; C) \geq 0$ if $I_f(A; B)$ is submodular in A for a fixed set B. Similarly show that, $I_f(A; B; C) \leq 0$ if $I_f(A; B)$ is supermodular in A for a fixed set B.

Question 8. (6 marks)

Given a monotone submodular function f, does the inequality: $I_f(A_1; A_2; \cdots; A_k) \leq \min(f(A_1), \cdots, f(A_k))$ always hold for any $k \in \mathbb{N}$?