Problem 1

a

We know that according to the discrete derivative definition of submodularity, for some $S \subseteq R \subseteq X$ and $e \in X \setminus R$,

$$f(S \cup \{e\}) - f(S) \leq f(R \cup \{e\}) - f(R) \tag{1}$$

Now, Let $B \subseteq A \subseteq X$. Given f_{max}, we can conclude the following,

$$\begin{aligned} f(B) &\leq f(A) \\ \implies -f(B) &\geq -f(A) \end{aligned} \tag{2}$$

Now, let $e \in X \setminus A$. We find the discrete derivatives for the following cases,
If $w_e < f(B) \leq f(A)$ (weight of e is lesser than max weight in set B and subsequently than in set A),

$$\begin{aligned} f(B \cup \{e\}) - f(B) = f(B) - f(B) = 0, \\ f(A \cup \{e\}) - f(A) = 0, \\ \implies f(A \cup \{e\}) - f(A) = f(B \cup \{e\}) - f(B) \end{aligned} \tag{3}$$

If $f(B) \leq w_e \leq f(A)$ (weight of e is greater(or equal) than max value in B but not than in A),

$$\begin{aligned} f(B \cup \{e\}) - f(B) = w_e - f(B) \geq 0, \\ f(A \cup \{e\}) - f(A) = f(A) - f(A) = 0, \\ \implies f(A \cup \{e\}) - f(A) \leq f(B \cup \{e\}) - f(B) \end{aligned} \tag{4}$$

If $w_e > f(A) \geq f(B)$ (weight of e is greater than max weight in set A and subsequently than in set B), using 2,

$$\begin{aligned} f(B \cup \{e\}) - f(B) = w_e - f(B) > 0, \\ f(A \cup \{e\}) - f(A) = w_e - f(A) > 0, \\ \implies f(A \cup \{e\}) - f(A) \leq f(B \cup \{e\}) - f(B) \end{aligned} \tag{5}$$

Using 3, 4 and 5, we observe that $\forall e \in X \setminus A$,

$$f(A \cup \{e\}) - f(A) \leq f(B \cup \{e\}) - f(B) \tag{6}$$

Hence, according to 1, the function f_{max} is sub-modular.

b

Let $B \subseteq A \subseteq X$. Given f_{min}, we can conclude the following,

$$\begin{aligned} f(B) &\geq f(A) \\ \implies -f(B) &\leq -f(A) \end{aligned} \tag{7}$$

Now, let $e \in X \setminus A$. We find the discrete derivatives for the following cases,
If $w_e < f(A) \leq f(B)$ (weight of e is lesser than min weight in set A and subsequently than in set B), using 7,

$$\begin{aligned} f(A \cup \{e\}) - f(A) = w_e - f(A) < 0, \\ f(B \cup \{e\}) - f(B) = w_e - f(B) < 0, \\ \implies f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B) \end{aligned} \tag{8}$$

From 8 we observe that f_min does not satisfy 1, and hence its **not** submodular.

Problem 2

Let's define a function for $i \in V$ and $A \subseteq V$, $n_i(A)$ = no. of edges connecting vertex i to vertices in set A.
Now let $A \subseteq B \subseteq V$ and $e \in V \setminus B$. We can write the following,

$$c(A \cup \{e\}) - c(A) = c(A) - n_e(A) + n_e(\bar{A}) - c(A) = n_e(\bar{A}) - n_e(A) \tag{9}$$

$$c(B \cup \{e\}) - c(B) = c(B) - n_e(B) + n_e(\bar{B}) - c(B) = n_e(\bar{B}) - n_e(B) \tag{10}$$

Since $A \subseteq B$, we can conclude the following,

$$\begin{gathered} n_e(A) \leq n_e(B), \\ n_e(\bar{A}) \geq n_e(\bar{B}) \end{gathered} \tag{11}$$

Using 9, 10 and 11, we see that,

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B) \tag{12}$$

Since c satisfies 1, it is submodular.

Problem 3

(a)

Let P denote the set of papers ($|N| = n$) and R the set of reviewers ($|R| = k$). Given that R is divided in m subsets. The let the quality of the review function for any paper $p_i \in P$ be $f_i(A)$ where $A \subseteq R$. Let $e_1, e_2 \in R$ be two reviewers. Taking the quality of independent review of each reviewer to be equal, we could say that,

$$f(\{e_1\}) = f(\{e_2\}) \tag{13}$$

Given that f_i gives diminishing returns, using 14 we can say that for some $e_1, e_2 \in R$ and $A \subseteq R$,

$$\begin{aligned} & f(A \cup \{e_1\} \cup \{e_2\}) - f(A \cup \{e_1\}) \leq f(A \cup \{e_1\}) - f(A) \\ \implies & f((A \cup \{e_2\}) \cup \{e_1\}) - f(A \cup \{e_2\}) \leq f(A \cup \{e_1\}) - f(A) \end{aligned} \tag{14}$$

Now since $A \subseteq (A \cup \{e_2\}) \subseteq R$ and $e_1 \in R \setminus (A \cup \{e_2\})$, using 1 we can say that f_i is submodular.
And since positive linear combination of submodular functions are also submodular, the function f s.t.,

$$f = \sum_i^n f_i \tag{15}$$

is also submodular.
Now adding the additional constraint of choosing at-most one reviewer from each lab (group out of m), we can see that this forms a matroidal constraint on $A \subseteq R$, more specifically a partition

matroid constraint. Hence the function remains sub-modular.
The maximization problem now becomes,

$$\max_{A \subseteq R} f(A) \tag{16}$$

Still, the problem is trivial. Now, let's assume that only $0 \geq l \leq n$ papers can be reviewed by a reviewer. So, to formulate the problem now, we will choose a subset A of reviewers who have to review all of the papers in a subset K of P. Also, they can only review those K set of papers. So, $|K| \leq l$ is the new constraint. Now we have the submodular function f_{K_i} for a subset K_i of P which is the sum of quality of reviews of all the papers in K_i (previous arguments of submodularity still hold).
Let S be a subcollection of the set of all subsets of P such that all the subsets in S are mutually exclusive and exhaustive sets of P. From [1], the new problem formed,

$$\max_{K_i \in S} \cup_{K_i}^{P} f(K_i) \tag{17}$$

with the uniform matroid constraint $|K_i| < l$ is also submodular.

Problem 4

(a)

In the implementation, a gaussian was fit onto the training data using the Gaussian Process Regression algorithm from the *scikit-learn* library in python. The optimizer for the hyperparameters (maximizing log maximal likelihood) is also run. The kernel used is a combination of the libraries Constant, Exponential Sine Squared, and White Noise kernel. The effective equation was,

$$k_y(x_p, x_q) = \sigma_f^2 \exp(-2\frac{sin(\frac{\pi}{\tau|x_p - x_q|})^2}{l}) + \sigma_n^2 \tag{18}$$

where σ_f^2 is signal variance(constant kernel), l is scale factor(RBF kernel), σ_n^2 is noise variance(White Noise kernel) and τ is the periodicity.
The values of the above mentioned hyperparameters that were learned are :

$$\sigma_f = 1.06,$$
$$l = 2.01$$
$$\tau = 2.08$$
$$\sigma_n = 0.083$$

The mean square error of the estimated function with the ground truth ($y = sin(3x)$) is = **0.038**. The final plot is given in figure 1.

(b)

In the implementation, a similar approach as above for fitting the data is used. But, as the dataset was quite large and a powerful computer wasn't available, to cut down the time for testing different kernels, n random samples was chosen from the training data and the fitting was done m times.

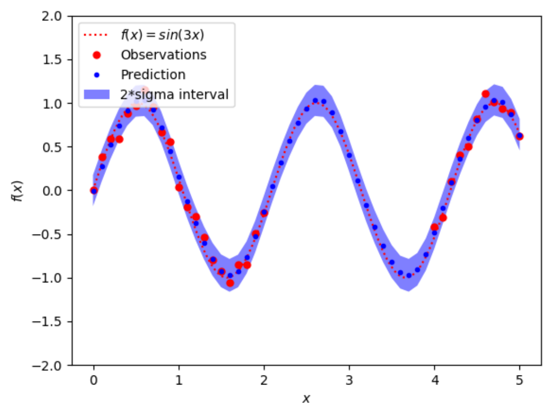

Figure 1: Output plot for Problem 4(a)

Since the parameters were being learned, the approximate parameters learned on each iteration improved the optimization time for the next iteration bringing the time to run the whole code down. Due to the size of the dataset, this approach could be assumed to give results very close to that by running the algorithm once on the whole dataset.
The kernel used is a combination of the libraries Constant, RBF(Radial Basis function, same as square-exponential) and White Noise kernel. The effective equation was,

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(\frac{-1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \tag{19}$$

where σ_f^2 is signal variance(constant kernel), l is scale factor(RBF kernel), σ_n^2 is noise variance(White Noise kernel).
The values of the above mentioned hyperparameters that were learned are :

$$\sigma_f = 43.9,$$

$$l = 53.2$$

$$\sigma_n = 3.95$$

$$n = 500$$

$$m = 50$$

The mean square error of the estimated function with the ground truth is = **4.04** (Since random samples are chosen each time, this error varied from 4.04-4.2).

NOTE

Instruction to run code given in README.md file.

References

[1] Andreas Krause and Daniel Golovin. *Submodular function maximization.* 2014.