1.

Consider the following data set for regression:

Index	X	Y
1	0	1
2	-2	2
3	-1	- 2

(a) Show all possible bootstrap samples of the dataset that have the same size as this dataset. Note that permutations of the same data set are considered the same dataset, for example $\{1, 2, 3\}$ and $\{2, 3, 1\}$ are the same dataset., in which 1,2,3 are indices of the data points.

(b) Construct a KNN regression model for all bootstrap samples in part 5a that contain the data point $(x, y) = (0, 1)$ with $K = 2$ and predict the label for the test point $x^* = 0$ by averaging the predictions of those bootstrap models.

2.

Consider the following dataset:

Index	X_1	X_2	Y
1	0	-1	1
2	1	0	-1
3	-1	1	2
4	1	-1	-2
5	-1	-1	0

Determine the leave-one-out cross validation estimate of the MSE $\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ of the null model $\hat{y}_i = \hat{\beta}_0$ on this dataset.

3.

Consider the following data set for regression:

Index	X_1	X_2	Y
1	0	-1	1
2	-2	0	2
3	-2	3	- 2
4	2	1	3
5	-5	6	2
6	1	-9	-1
7	19	-10	0
8	0	15	3

Assume that we split this dataset into 4 folds: the first two data points are the first fold, the second two data points are the second fold, so on so forth (folds are separated using double lines in the table).

Calculate the 4-fold cross-validation error of KNN regression for the above data set, using $k = 2$ nearest neighbors and the Manhattan distance between the training point $\mathbf{x} = (x_1, x_2)$ and the test point $\mathbf{x}^* = (x_1^*, x_2^*)$, i.e. $d_M(\mathbf{x}, \mathbf{x}^*) = |x_1 - x_1^*| + |x_2 - x_2^*|$. If there is a tie in nearest neighbors, break it in any way you want. Suggestion: to organize your calculations, you may want to use tables.

4.

Consider the following data set for regression:

Index	X	Y
1	0	1
2	-2	2
3	-1	- 2

(a) Show all possible bootstrap samples of the dataset that have the same size as this dataset. Note that permutations of the same data set are considered the same dataset, for example $\{1, 2, 3\}$ and $\{2, 3, 1\}$ are the same dataset., in which 1,2,3 are indices of the data points.

(b) Construct a KNN regression model for all bootstrap samples in part 5a that contain the data point $(x, y) = (0, 1)$ with $K = 2$ and predict the label for the test point $x^* = 0$ by averaging the predictions of those bootstrap models.

5.

Consider the training set $\{1, 2, 3\}$. There are 10 *distinct* bootstrap samples with the same size that can be drawn from this training set. (For example, $\{1, 2, 2\}$ is not distinct from $\{2, 2, 1\}$). Using those bootstrap samples, build an 80% bootstrap confidence interval for the mean.