Note:
- Write all required theory or calculations steps.
- Symbols have their usual meanings.

Q1. Following is a regression data:

ID	Height (x_1)	Weight (x_2)	label (y)
1	1	2	1
2	0	1	1
3	3	0	1
4	1	1	-1
5	2	1	-1

A gradient boosting regression tree is determined using a model $F(x) + h(x)$. Suppose you build $F(x)$ and $h(x)$ as trees with only two terminal nodes (decision stumps). The loss functions is mean squared error. Using a single cut across height at 1.1 and a single cut across weight at 1.1.

    a. Find the regression model $F(x)$.   [2]
    b. Find $h(x)$.   [2]

Q2. Consider the Adaboost algorithm. Consider the following classification dataset

ID	Height (x_1)	Weight (x_2)	Class (y)
1	1	2	1
2	0	1	1
3	3	0	1
4	1	1	-1
5	2	1	-1
6	2	2	-1

    a. Find a decision tree classifier with single cut (decision stump). Make one cut across x_1 at 1.1 and one cut across x_2 at 1.1. Find the classifier using weighted zero-one loss. Let this be $h_1(x)$.   [2]
    b. Compute α_1 and updated weights.   [1]
    c. Compute $h_2(x)$ using a cut across x_1 at 1.1 and a cut across x_2 at 1.1.   [2]
    d. Find the output for $(x_1, x_2) = (3,0)$ using the final boosted classifier $f(x)$.   [1]
Note: In case the loss is equal in two cases, chose the tree with cut across **height** or x_1.

Q3. It is IPL season. We have a match between Gujarat and Mumbai today. Lets suppose that you decide to check with your friends that which team will win a match with two inputs – form (x_1) and fitness (x_2). If players are in form, then $(x_1= 1)$ else $(x_1=0)$, if players are fit $(x_2= 1)$ else $(x_2=0)$. Suppose you collected samples by asking 5 of your friends whose replies are
$X = \{x_1, x_2\} = \{(1,0), (0,1), (1,1), (0,0), (0,1)\}$ for Gujarat
$X = \{x_1, x_2\} = \{(0,0), (1,1), (1,1), (1,0), (0,0)\}$ for Mumbai

Assume that other required things are equal, who will win the match if $(x_1, x_2)=(1,0)$? Show the required theory. [3]

Q4. Consider the following data

Sample	Class 1			Class 2		
no.	(x_1)	(x_2)	(x_3)	(x_1)	(x_2)	(x_3)
1	1	0	-1	1	1	0
2	0	0	0	0	1	0
3	-1	1	1	0	0	0
4	No sample			-1	0	0

Suppose we use a trained encoder from an auto-encoder with the following structure. Let the output node has a linear activation.

x_1 w_1

x_2 y

w_2

x_3 w_3
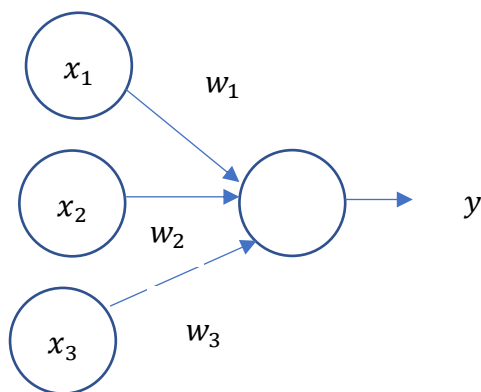
a. Find y for all inputs. Weights are $w_1=1, w_2=0, w_3=-1$. [1]
b. Once you find y, compute mean and unbiased variance MLE estimate for each class using y. Assume a Gaussian distribution for y. [1]
c. Find the decision boundary between two classes using the components of part (b). [2]