This homework is done by Tianwei Mo (Bill).

1.

a) Possible predictors are: The frequency of the sound, the volume of the sound, etc.
The response variables are male and female.

b) Possible predictors are: The writing time, the numbers of horizontal and vertical lines used, the length of the recorded motion track, etc.
The response variables are all letters, such as a, b, A, B, …, and all numbers, such as 1, 2, ….

2.

a) Since $y = 0, 1$,

$$P(y = 0|\boldsymbol{x}) = 1 - P(y = 1|\boldsymbol{x})$$
$$\frac{1}{1 + e^{-z}} > 0.5$$
$$e^{-z} < 1$$
$$z > 0$$
$$2x_1 + 3x_2 > -1$$

Thus, the set of x is set S_1 such that $2x_1 + 3x_2 > -1$ for $x_1, x_2 \in S_1$.

b)

$$\frac{1}{1 + e^{-z}} > 0.8$$
$$z > \ln 4$$
$$2x_1 + 3x_2 > \ln 4 - 1$$

Thus, the set of x is set S_2 such that $2x_1 + 3x_2 > \ln 4 - 1$ for $x_1, x_2 \in S_2$.

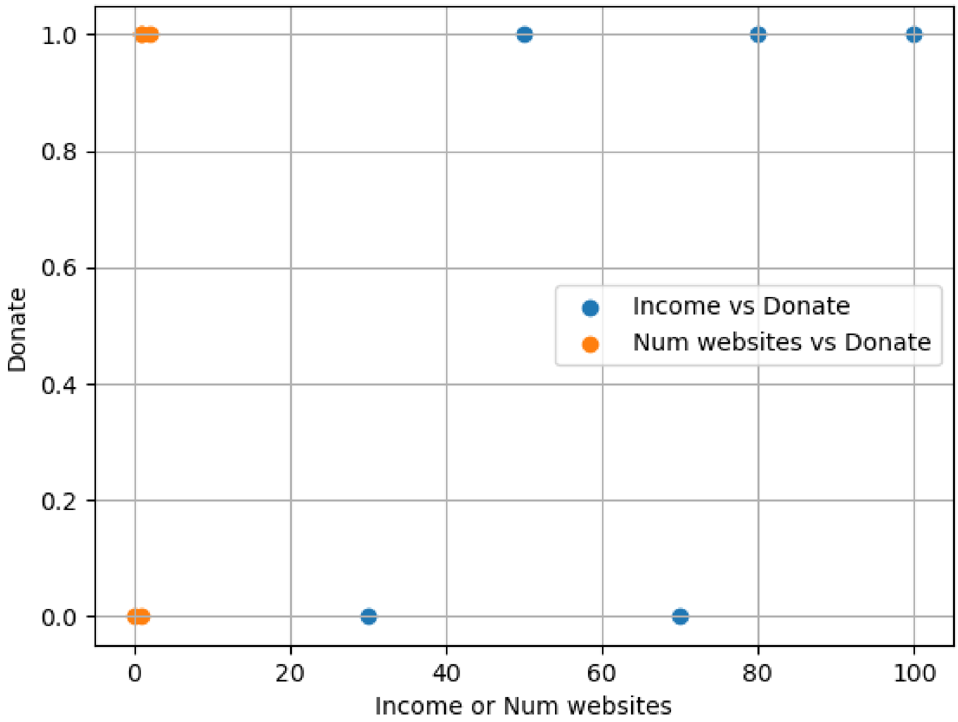
c)

$$2x_1 + 1.5 > \ln 4 - 1$$
$$x_1 > \ln 2 - 1.25$$

Thus, the set of x_1 is $(\ln 2 - 1.25, +\infty)$

3.

a) The scatter plot:



b)

$$\mathbf{w} = [0,2],$$
$$b = -1.$$

c) a

$$z_1 = -1$$
$$z_2 = 1$$
$$z_3 = 1$$
$$z_4 = 3$$
$$z_5 = 1$$

Since $\frac{1}{(1+e^{-z})}$ is a increasing function, the smallest z has the least likelihood.

Thus, sample 1 is the least likely.

d) It wouldn't change the y hat since z_i is a linear expression, multiplying a positive scalar won't change its positive or negative. It would change the likelihood.

$$z_i' = \mathbf{w}'^{\mathbf{T}}\mathbf{x}_i + b' = \alpha\mathbf{w}^{\mathbf{T}}\mathbf{x}_i + \alpha b = \alpha z_i$$
$$P'(y_i - 1|x_i) = \frac{1}{1+e^{-z_i'}} = \frac{1}{1+e^{-\alpha z_i}}$$

4.

a)

$$P(y_i = 1|x_i) = \frac{1}{1 + e^{-z_i}}, z_i = w^T x_i + b$$

Let $b = -6, w = [0.05, 1], x_i = [40, 3.5],$

$$z_i = -0.5$$
$$P(y_i = 1|x_i) = 0.3775$$

b) Let $b = -6, w = [0.05, 1], x_i = [x_1, 3.5], P = 0.5$

$$\frac{1}{1 + e^{-z_i}} = 1$$
$$z = 0$$
$$x_1 = 50$$

The student need to study for 50 hours.

5.

a)

$$\frac{\partial z_i}{\partial \beta_0} = 1$$
$$\frac{\partial z_i}{\partial \beta_1} = x_1$$
$$\frac{\partial z_i}{\partial \beta_2} = x_2$$

b)

$$\frac{\partial J(\beta)}{\partial \beta_0} = \frac{\partial J(\beta)}{\partial z_i}\frac{\partial z_i}{\partial \beta_0} = \sum_{i=1}^{N}\left(\frac{1}{1 + e^{z_i}} - y_i\right) * 1 = \sum_{i=1}^{N}\frac{1}{1 + e^{z_i}} - y_i$$

$$\frac{\partial J(\beta)}{\partial \beta_1} = \frac{\partial J(\beta)}{\partial z_i}\frac{\partial z_i}{\partial \beta_1} = x_1\sum_{i=1}^{N}\frac{1}{1 + e^{z_i}} - y_i$$

$$\frac{\partial J(\beta)}{\partial \beta_2} = \frac{\partial J(\beta)}{\partial z_i}\frac{\partial z_i}{\partial \beta_2} = x_2\sum_{i=1}^{N}\frac{1}{1 + e^{z_i}} - y_i$$

c) Let $\frac{\partial J(\beta)}{\partial \beta_0} = 0,$

$$\sum_{i=1}^{N}\frac{1}{1 + e^{z_i}} - y_i = 0$$

Let $\frac{\partial J(\beta)}{\partial \beta_1} = 0,$

$$x_1 \sum_{i=1}^{N} \frac{1}{1 + e^{z_i}} - y_i = 0$$

Let $\frac{\partial J(\beta)}{\partial \beta_2} = 0$,

$$x_2 \sum_{i=1}^{N} \frac{1}{1 + e^{z_i}} - y_i = 0$$

There is no closed form solution for the derivatives. We can use gradient decent to optimize the loss function.