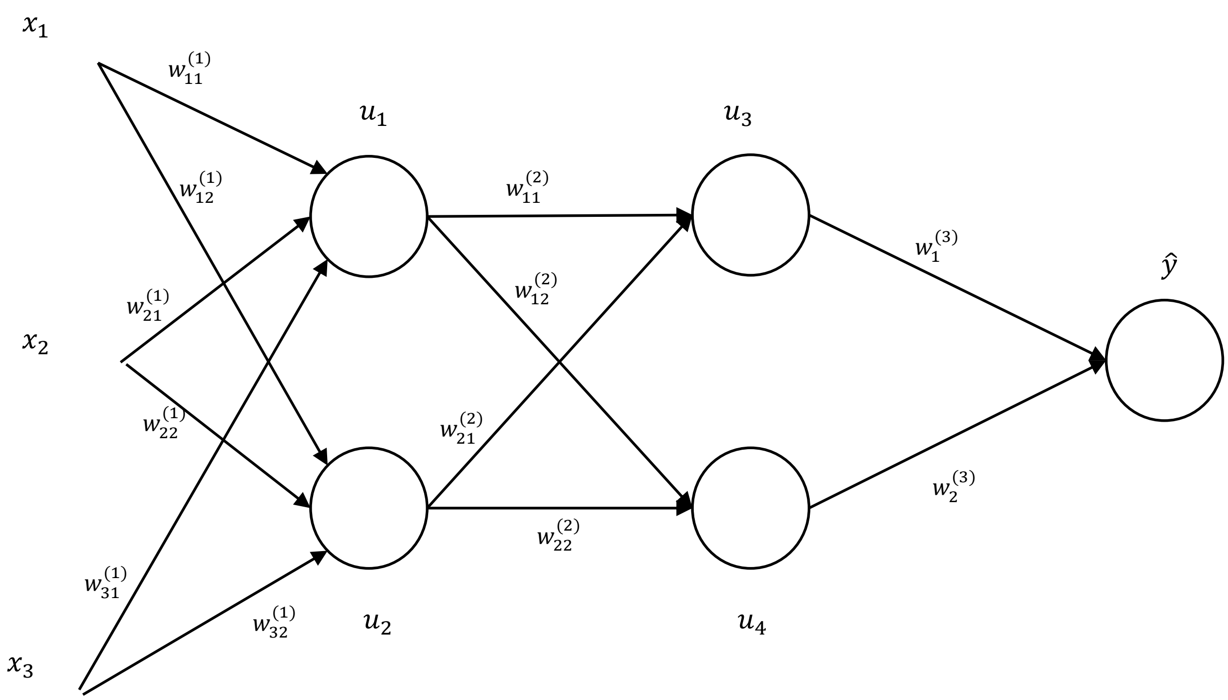
Question 1:

The neural network will look as follows:



First, we will define the functions u_1, u_2, u_3, u_4 with ReLU activation functions, and $\hat{y}$ with identity function. All the weights are 1.

The loss function is RSS.

$$u_1 = \max\left(0, w_{11}^{(1)}x_1 + w_{21}^{(1)}x_2 + w_{31}^{(1)}x_3\right)$$

$$u_2 = \max\left(0, w_{12}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{32}^{(1)}x_3\right)$$

$$u_3 = \max\left(0, w_{11}^{(2)}u_1 + w_{21}^{(2)}u_2\right)$$

$$u_4 = \max\left(0, w_{12}^{(2)}u_1 + w_{22}^{(2)}u_2\right)$$

$$\hat{y} = w_1^{(3)}u_3 + w_2^{(3)}u_4$$

In backpropagation, we want to derive the loss function w.r.t the weights.

$$Loss = \sum_{I=1}^{N}(y_I - \hat{y_i})^2$$

$$\frac{dLoss}{dw_1^{(3)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_1^{(3)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*u_3}$$

$$\frac{dLoss}{dw_2^{(3)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_2^{(3)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*u_4}$$

$$\frac{dLoss}{dw_{11}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{11}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_3}*\frac{du_3}{dw_{11}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*u_1}$$

$$\frac{dLoss}{dw_{21}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{21}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_3}*\frac{du_3}{dw_{21}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*u_2}$$

$$\frac{dLoss}{dw_{12}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{12}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_4}*\frac{du_4}{dw_{12}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_2^{(3)}*u_1}$$

$$\frac{dLoss}{dw_{22}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{22}^{(2)}} = -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_4}*\frac{du_3}{dw_{22}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_2^{(3)}*u_2}$$

$$\begin{aligned}\frac{dLoss}{dw_{11}^{(1)}} &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{11}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_3}*\frac{du_3}{du_1}*\frac{du_1}{dw_{11}^{(1)}} - 2(y-\hat{y})*\frac{d\hat{y}}{du_4}*\frac{du_4}{du_1}*\frac{du_1}{dw_{11}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*\frac{du_1}{dw_{11}^{(1)}} - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*\frac{du_1}{dw_{11}^{(1)}}\\ &= \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_1 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_1}\end{aligned}$$

$$\begin{aligned}\frac{dLoss}{dw_{21}^{(1)}} &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{21}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_3}*\frac{du_3}{du_1}*\frac{du_1}{dw_{21}^{(1)}} - 2(y-\hat{y})*\frac{d\hat{y}}{du_4}*\frac{du_4}{du_1}*\frac{du_1}{dw_{21}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*\frac{du_1}{dw_{21}^{(1)}} - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*\frac{du_1}{dw_{21}^{(1)}}\\ &= \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_2 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_2}\end{aligned}$$

$$\begin{aligned}\frac{dLoss}{dw_{31}^{(1)}} &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{dw_{31}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*\frac{d\hat{y}}{du_3}*\frac{du_3}{du_1}*\frac{du_1}{dw_{31}^{(1)}} - 2(y-\hat{y})*\frac{d\hat{y}}{du_4}*\frac{du_4}{du_1}*\frac{du_1}{dw_{31}^{(1)}}\\ &= -2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*\frac{du_1}{dw_{21}^{(1)}} - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*\frac{du_1}{dw_{31}^{(1)}}\\ &= \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_3 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_3}\end{aligned}$$

$$\frac{dLoss}{dw_{12}^{(1)}} = -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{dw_{12}^{(1)}}$$

$$= -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{du_3} * \frac{du_3}{du_2} * \frac{du_2}{dw_{12}^{(1)}} - 2(y - \hat{y}) * \frac{d\hat{y}}{du_4} * \frac{du_4}{du_2} * \frac{du_2}{dw_{12}^{(1)}}$$

$$= -2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * \frac{du_2}{dw_{12}^{(1)}} - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * \frac{du_2}{dw_{12}^{(1)}}$$

$$= \boldsymbol{-2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * x_1 - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * x_1}$$

$$\frac{dLoss}{dw_{22}^{(1)}} = -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{dw_{22}^{(1)}}$$

$$= -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{du_3} * \frac{du_3}{du_2} * \frac{du_2}{dw_{22}^{(1)}} - 2(y - \hat{y}) * \frac{d\hat{y}}{du_4} * \frac{du_4}{du_2} * \frac{du_2}{dw_{22}^{(1)}}$$

$$= -2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * \frac{du_2}{dw_{22}^{(1)}} - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * \frac{du_2}{dw_{22}^{(1)}}$$

$$= \boldsymbol{-2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * x_2 - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * x_2}$$

$$\frac{dLoss}{dw_{32}^{(1)}} = -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{dw_{32}^{(1)}} =$$

$$= -2 * (y_i - \hat{y}_i) * \frac{d\hat{y}}{du_3} * \frac{du_3}{du_2} * \frac{du_2}{dw_{32}^{(1)}} - 2(y - \hat{y}) * \frac{d\hat{y}}{du_4} * \frac{du_4}{du_2} * \frac{du_2}{dw_{32}^{(1)}}$$

$$= -2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * \frac{du_2}{dw_{32}^{(1)}} - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * \frac{du_2}{dw_{32}^{(1)}}$$

$$= \boldsymbol{-2 * (y_i - \hat{y}_i) * w_1^{(3)} * w_{21}^{(2)} * x_3 - 2 * (y_i - \hat{y}_i) * w_2^{(3)} * w_{22}^{(2)} * x_3}$$

$$y = 0$$

$$u_1 = \max(0, 1 * 1 + 1 * 2 + (-1) * 1) = 2$$

$$u_2 = \max(0, 1 * 1 + 1 * 2 + (-1) * 1) = 2$$

$$u_3 = \max(0, 2 * 1 + 2 * 1) = 4$$

$$u_4 = \max(0, 2 * 1 + 2 * 1) = 4$$

$$\hat{y} = 4 * 1 + 4 * 1 = 8$$

$$\frac{dLoss}{dw_1^{(3)}} = \boldsymbol{-2 * (y_i - \hat{y}_i) * u_3} = -2 * (0 - 8) * 4 = 64$$

$$\frac{dLoss}{dw_2^{(3)}} = \boldsymbol{-2 * (y_i - \hat{y}_i) * u_4} = -2 * (0 - 8) * 4 = 64$$

$$\frac{dLoss}{dw_{11}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*u_1} = -2*(0-8)*2 = 32$$

$$\frac{dLoss}{dw_{21}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*u_2} = -2*(0-8)*2 = 32$$

$$\frac{dLoss}{dw_{12}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_2^{(3)}*u_1} = -2*(0-8)*2 = 32$$

$$\frac{dLoss}{dw_{22}^{(2)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_2^{(3)}*u_2} = -2*(0-8)*2 = 32$$

$$\frac{dLoss}{dw_{11}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_1 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_1}$$
$$= -2*(0-8) - 2*(0-8) = 32$$

$$\frac{dLoss}{dw_{21}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_2 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_2}$$
$$= -2*(0-8)*2 - 2*(0-8)*2 = 64$$

$$\frac{dLoss}{dw_{31}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{11}^{(2)}*x_3 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{12}^{(2)}*x_3}$$
$$= -2*(0-8)*(-1) - 2*(0-8)*(-1) = -32$$

$$\frac{dLoss}{dw_{12}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{21}^{(2)}*x_1 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{22}^{(2)}*x_1}$$
$$= -2*(0-8) - 2*(0-8) = 32$$

$$\frac{dLoss}{dw_{22}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{21}^{(2)}*x_2 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{22}^{(2)}*x_2}$$
$$= -2*(0-8)*2 - 2*(0-8)*2 = 64$$

$$\frac{dLoss}{dw_{32}^{(1)}} = \boldsymbol{-2*(y_i - \hat{y}_i)*w_1^{(3)}*w_{21}^{(2)}*x_3 - 2*(y_i - \hat{y}_i)*w_2^{(3)}*w_{22}^{(2)}*x_3}$$
$$= -2*(0-8)*(-1) - 2*(0-8)*(-1) = -32$$

Question 2

Code attached in zip file.

Question 3

Given a set of points in general positions in the plane $\{(x_1^1, x_2^1), \cdots, (x_1^N, x_2^N)\}$, and a real vector $Y = \{y_1, \dots, y_N\}$.

Fitting X to Y is as regressions problem. Basically, we want a neural network that for each $\{x_1^i, y_1^i\}$ will fit a $\hat{y}$ so that $y - \hat{y} < \epsilon$.

We will build a network with input layer that consists of 3 neurons (x_1, x_2, b), we need $2(N)$ neurons in the hidden layer. Each tower can be modeled by subtracting two step functions. Each step function can be represented by a single neuron in the hidden layer.

We build the neurons in the following way: for two points $\{x_1^i, x_2^i\}$, the line connecting them will be:

$l_i = w_{1,i}^1 * x_1 + x_{2,i}^1 * x_2 + x_{3,i}^1 * b = 0$. Because of the general position, we know this line will capture only two points.

For each line, we will create two parallels, with space of $\pm\epsilon$:

$l_+^i = w_{1,i}^1 * x_1 + x_{2,i}^1 * x_2 + x_{3,i}^1 * b + \epsilon = 0$; $l_-^i = w_{1,i}^1 * x_1 + x_{2,i}^1 * x_2 + x_{3,i}^1 * b - \epsilon = 0$, so that the only two points in the strip between l_+^i, l_-^i are x_1, x_2. Each of these lines is a linear combination, and therefore can be represented by a single neuron: one neuron for l_+^i and another for l_-^i.

We have a finite number of points, so we can do this process for each set of points in the input matrix X.

Using the activation step function, we define:

$$\begin{cases} a_{i_1} = -1\,;\ w_{1,i}^1 * x_1 + x_{2,i}^1 * x_2 + x_{3,i}^1 * b - \epsilon > 0 \\ a_{i_1} = 0\,;\ \ otherwise \end{cases}$$

$$\begin{cases} a_{i_2} = 1\,;\ w_{1,i}^1 * x_1 + x_{2,i}^1 * x_2 + x_{3,i}^1 * b + \epsilon > 0 \\ a_{i_2} = 0\,;\ \ otherwise \end{cases}$$

So that if the sum of $a_{i1} + a_{i2}$ is 0 for all the points out of the strip, and 1 for all the points inside the strip (like we did in class).

From the hidden layer to the output layer, which consists of a single neuron, we will define:

$$w_{2i-1}^2 = w_{2i}^2 = w_i$$

Because there are only two strips containing each x_i we will get the equation: $y_i = w_i + w_{i-1}$.

We do this process for each input. For y_1, y_N we will use the line connecting them.

Because the all the y_i are known, we get a matrix of N*N which looks:

$$\begin{pmatrix} w_1 + 0 * w_2 + \ldots + 0 * w_{n-1} + w_n = y_1 \\ 0 * w_1 + w_2 + w_3 + \cdots + 0 * w_n = y_2 \\ \dots \\ \dots \\ \dots \end{pmatrix}$$

This matrix needs to be of full dimension to support N equations.

Question 4

Code attached in zip file.