Question 1

1.1

$$\nabla_b \sum_i (x_i - b)^2 = 0 \Rightarrow 2\sum_i (x_i - b) = 0$$
$$\Rightarrow \sum_i x_i - nb = 0$$
$$\Rightarrow \sum_i x_i = nb$$
$$\Rightarrow \boxed{b = \frac{1}{n}\sum_i x_i}$$

1.2

The optimal value for b is the average of a normal distribution

1.3

The minimum value for this loss function is going to be 0, because of the absolute value.

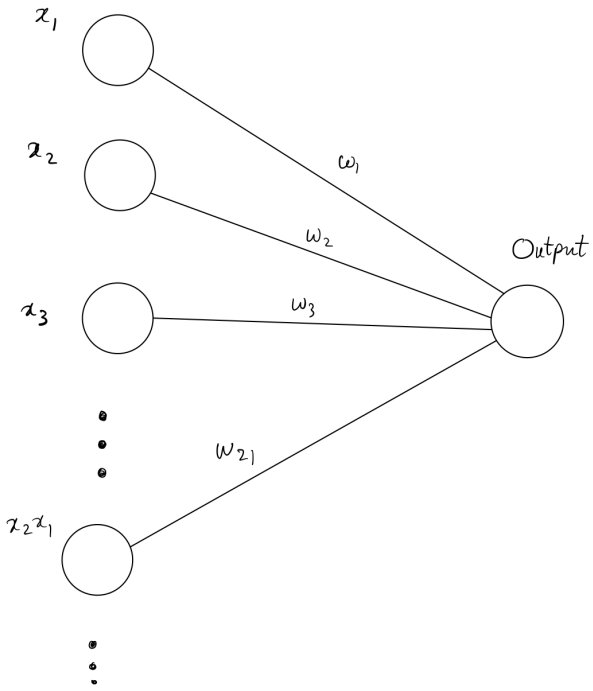
$$\sum_i |x_i - b| = 0 \Rightarrow \boxed{b = \frac{1}{n}\sum_i x_i}$$

Question 2

Linear function of a column vector $[x_1, x_2, ..., 1]^T$, can be expressed as a linear transformation:

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{w} + b = [\mathbf{x},\, 1][\mathbf{w},\, b]^T$$

Question 3



Question 4

4.1

The matrix $\mathbf{X}^T\mathbf{X}$ won't be invertible. Thus, no unique solution will exist for the optimal weight $\mathbf{w}^*$.

4.2

One way to potentially fix it is to use a technique to approximate the inverse of $\mathbf{X}^T\mathbf{X}$. Introducing noise will do the following:

$$(\mathbf{X} + \epsilon)^T(\mathbf{X} + \epsilon) = \mathbf{X}^T\mathbf{X} + \mathbf{X}^T\epsilon + \epsilon^T\mathbf{X} + \epsilon^T\epsilon$$

where ϵ is a matrix of all noise entries.

4.3

$$\begin{aligned} &\mathbb{E}[\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\epsilon + \epsilon^T\mathbf{X} + \epsilon^T\epsilon] \\ &= \mathbb{E}[\mathbf{X}^T\mathbf{X}] + \mathbb{E}[\mathbf{X}^T\epsilon] + \mathbb{E}[\epsilon^T\mathbf{X}] + \epsilon^T\epsilon \end{aligned}$$

4.4

The SGD algorithm is unable to find a global minimum.

Question 5

5.1

$$-\log P(\mathbf{y}|\mathbf{X}) = \sum_i |y - \mathbf{w}^T x^{(i)} - b|$$

5.2

Unable to provide closed form solution.

5.3

I think using the standard minibatch SGD update is sufficient:

$$\mathbf{w} \leftarrow \mathbf{w} - \frac{\eta}{|\beta|} \sum_i \mathbf{x}^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b - y^{(i)})$$

Question 6

It doesn't work because it needs at least one more layer as the final output layer.

Question 7

To be added...

Question 8

To be added...