CSCI 378
HW11

In maximum likelihood estimation (MLE) of ordinary linear regression, we want to compute the point estimate $\hat{w}$ of the weight vector that maximizes the likelihood of the responses $y^{(n)}$, given the evidence vectors $x^{(n)}$. In other words:

$$\hat{w} = \text{argmax}_w \prod_{n=1}^{N} P(y^{(n)} \mid w, x^{(n)})$$

As we saw, this can be re-expressed as:

$$\hat{w} = \text{argmax}_w \sum_{n=1}^{N} \log P_\epsilon(y^{(n)} - w^T x^{(n)})$$

where P_ϵ is the probability distribution for our stochastic term ϵ.

In ordinary linear regression, $P_\epsilon \sim \text{Normal}(0, \sigma^2)$.

We also mentioned robust regression, which is the same as ordinary linear regression, except that $P_\epsilon \sim \text{Laplace}(0, b)$ for some positive scaling factor b.

Show that in this case:

$$\hat{w} = \text{argmin}_w \sum_{n=1}^{N} |y^{(n)} - w^T x^{(n)}|$$

In other words, the loss function for robust regression is $L_{\text{robust}}(w) = \sum_{n=1}^{N} |y^{(n)} - w^T x^{(n)}|$.

Hint: the probability density function for $\text{Laplace}(\mu, b)$ is:

$$P(z) = \frac{1}{2b} \exp\left(-\frac{|z - \mu|}{b}\right)$$