Exercise 1

Question 1 (5pt)

Imagine we have two possibilities: We can scan and email the image, or we can use an optical character reader (OCR) and send the text file. Discuss the advantage and disadvantages of the two approaches in a comparative manner. When would one be preferable over the other? (5pt)

Question 2 (5pt)

One source of noise is error in the labels. Can you propose a method to find data points that are highly likely to be mislabeled?

Question 3 (10pt)

In lecture 2, you will learn that the loss function of linear regression model is:

$$L(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)} \right)^2, \tag{1}$$

where $h(x) = wx + b$.

The loss function of logistic regression model is:

$$L(w, b) = -\frac{1}{m} \sum_{i=1}^{m} \left(y^{(i)} \log(h(x^{(i)})) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) \right), \tag{2}$$

where $h(x) = \frac{1}{1+e^{-(wx+b)}}$.

Please prove that although both models have different loss functions, their optimization are same since they have the same derivatives $\frac{L(w,b)}{\partial w}$, $\frac{L(w,b)}{\partial b}$.