A Single Neuron

- Data
- Regression vs Classification
  - Neuron

\[ \hat{y} = \theta, X + b \]

Data

\[ D = \{X, Y\} \]

Input \rightarrow Label (Known) \rightarrow \text{Supervised Learning}

Example Problem: Image of Digit Classification

\( X \): all of our training images
\( Y \): the label for each image

- Capital \( \Rightarrow \) matrix
- Lower case \( \Rightarrow \) column vector

Matrix

\[ X = \begin{bmatrix} x_1 \cdots x_n \end{bmatrix} \]

Shape

Column vector is \((784, 1)\)
\[ y(x) = \frac{2}{x} \]

One-hot encoding

\[
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

Linear Regression

What is this line?

Model: \[ \hat{y} = mx + b \]

What is the relationship between the input \( x \) and output \( y \)?

New input value, we don't know the true value for \( y \)

Binary Classifier

\[ \hat{y} = mx + b \geq 0 \]

We want the output to be \( T/F, 1/0, 1/-1 \)
Goal: We want to find values for $w_k$ and $b$ such that $\hat{y}^{(i)} = y^{(i)}$ for all values of $i$.

What is a good objective function? 

loss

$L ( \hat{y}^{(i)}, y^{(i)} ) = \hat{y}^{(i)} - y^{(i)}$

Does not work very well.
<table>
<thead>
<tr>
<th>( \hat{y} )</th>
<th>( y )</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

This is a problem.

\[
L(\hat{y}^{(n)}, y^{(n)}) = |\hat{y}^{(n)} - y^{(n)}| \quad \text{MAD}
\]