

1. Section 1

Derive the analytical solution of w_0 and w_1

dataset: $n=1, \dots, N$ Observations

x_n -> one dimensional variable

t_n -> label

linear model: $f(x_n; w_0, w_1) = w_0 + w_1 * x_n$

squared loss function: $L = \frac{1}{N} \sum_N (f(x_n) - t_n)^2$

mean value of $x_n = \bar{x}$

mean value of $t_n = \bar{t}_n$

Helpful Equation: $\sum N x_n = N \bar{x}_n$

Helpful Equation: $\sum N t_n = N \bar{t}_n$

$$L = \frac{1}{N} \sum_N (w_0 + w_1 x_n - t_n)^2$$

$$\begin{aligned} \frac{\partial L}{\partial w_0} &= \frac{1}{N} \sum_N 2(w_0 + w_1 x_n - t_n) \\ &= \frac{2}{N} \sum_N (w_0 + w_1 x_n - t_n) = 0 \\ &= \frac{2}{N} \left(\sum_N w_0 + \sum_N w_1 x_n - \sum_N t_n \right) = 0 \\ &= \frac{2 \sum_N w_0}{N} + \frac{2 \sum_N w_1 x_n}{N} - \frac{2 \sum_N t_n}{N} = 0 \\ &= \frac{2Nw_0}{N} + \frac{2w_1 N \bar{x}_n}{N} - \frac{2N \bar{t}_n}{N} = 0 \\ &= 2w_0 + 2w_1 \bar{x}_n - 2\bar{t}_n = 0 \end{aligned} \tag{1.1}$$

$$\begin{aligned} \frac{\partial L}{\partial w_1} &= \frac{1}{N} \sum_N 2(w_0 + w_1 x_n - t_n) x_n \\ &= \frac{2}{N} \sum_N (w_0 + w_1 x_n - t_n) x_n = 0 \\ &= \frac{2}{N} \left(\sum_N w_0 x_n + \sum_N w_1 x_n^2 - \sum_N t_n x_n \right) = 0 \\ &= \frac{2}{N} \left(w_0 \sum_N x_n + w_1 \sum_N x_n^2 - \sum_N t_n x_n \right) = 0 \\ &= \frac{2w_0 \sum_N x_n}{N} + \frac{2w_1 \sum_N x_n^2}{N} - \frac{2 \sum_N t_n x_n}{N} = 0 \\ &= \frac{2w_0 N \bar{x}_n}{N} + \frac{2w_1 N \overline{x_n^2}}{N} - \frac{2N \overline{t_n x_n}}{N} = 0 \\ &= 2w_0 \bar{x}_n + 2w_1 \overline{x_n^2} - 2\overline{t_n x_n} = 0 \end{aligned} \tag{1.2}$$

$$\begin{aligned} \frac{\partial L}{\partial w_0} &= 2w_0 + 2w_1 \bar{x}_n - 2\bar{t}_n = 0 \\ \frac{\partial L}{\partial w_1} &= 2w_0 \bar{x}_n + 2w_1 \overline{x_n^2} - 2\overline{t_n x_n} = 0 \end{aligned} \tag{1.3}$$

$$\begin{aligned}
w_0 &= \frac{\overline{x_n t_n} - \overline{x_n t_n}}{\overline{x_n^2} - \overline{x_n^2}} \\
w_1 &= \overline{t_n} - \overline{x_n} \frac{\overline{x_n t_n} - \overline{x_n t_n}}{\overline{x_n^2} - \overline{x_n^2}}
\end{aligned}
\tag{1.4}$$

2. Section 2

Derive the analytical solution of w in vector and matrix format

dataset: $n=1, \dots, N$ Observations

$$W = [w_0, w_1]^T$$

$$X_n = [1, x_n]^T$$

linear model: $f(x_n; w_0, w_1) = w_0 + w_1 * x_n = W^T X_n$

squared loss function: $L = \frac{1}{N} \sum_N (f(x_n) - t_n)^2$

$$L = \frac{1}{2N} \sum_N (W^T X_n - t_n)^2$$

$$\begin{aligned} L &= \frac{1}{N} \sum_N (W^T X_n - t_n)^2 \\ &= \frac{1}{N} (X_n W - t_n)^T (X_n W - t_n) \\ &= ((X_n W)^T - t_n^T) (X_n W - t_n) \\ &= (X_n W)^T X_n W - (X_n W)^T t_n - t_n^T (X_n W) + t_n^T t_n \\ &= W^T X_n^T X_n W - 2(X_n W)^T t_n + t_n^T t_n \end{aligned} \tag{2.1}$$

$$\begin{aligned} \frac{\partial L}{\partial W} &= 2X_n^T X_n W - 2X_n^T t_n = 0 \\ &= X_n^T X_n W = X_n^T t_n \\ &= (X_n^T X_n)^{-1} X_n^T t_n \end{aligned} \tag{2.2}$$

3. Section 3

Discuss the selection of L2-norm vs L1-norm as the loss function for linear modeling (regression). You are asked to use examples of justify your answer.

When comparing L2 and L1 norms one of the biggest differences comes when your data contains some form of outliers. If your data does contain outliers then the L1 norm is going to move a much greater distance towards that outlier then the L2 norm would. A good example of this is shown in the figure below.

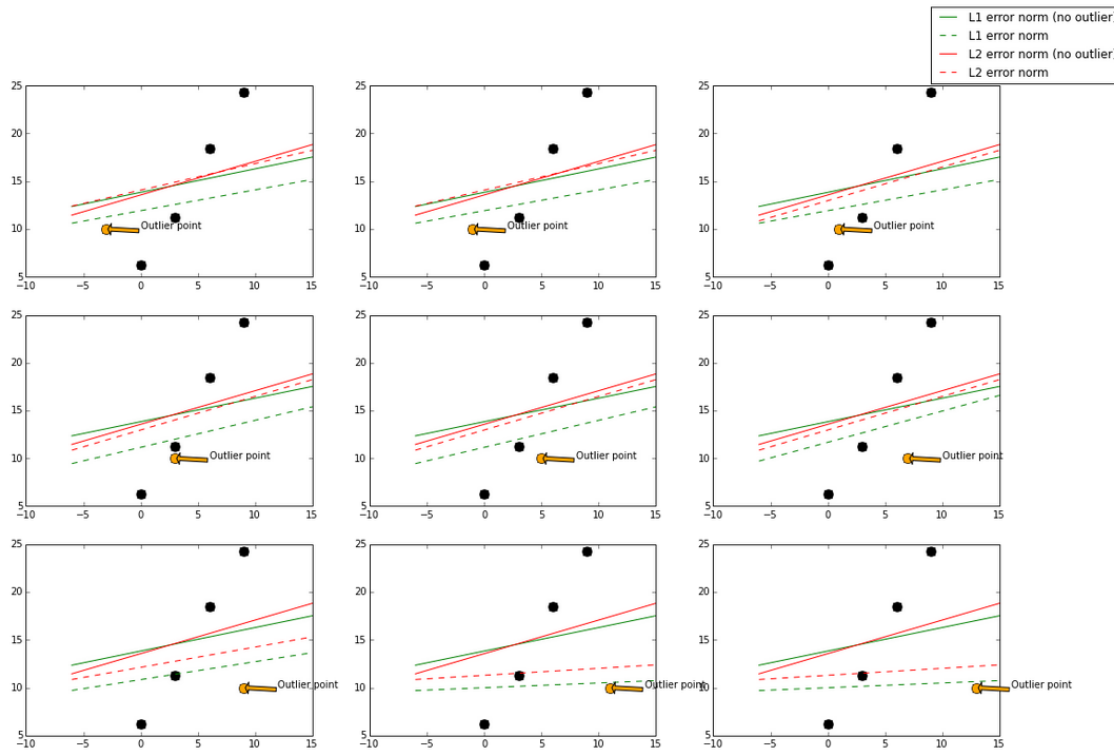


Figure 3.1: programmatic-L1-vs-L2-visualization

4. Section 4

S = (1, 1), (2, 2), (3, 3)

$$w_0 = 0$$

$$w_1 = 0$$

Step size = 0.1

$$f(x) = w_0 + w_1 * x_n$$

$$n = 1$$

$$w_0 = 0 + 0.1 * (1 - (0 + 1 * 0)) * 1 = 0.1$$

$$w_1 = 0 + 0.1 * (1 - (0 + 1 * 0)) * 1 = 0.1$$

$$n = 2$$

$$w_0 = 0.1 + 0.1 * (2 - (0.1 + 2 * 0.1)) * 1 = 0.27$$

$$w_1 = 0.1 + 0.1 * (2 - (0.1 + 2 * 0.1)) * 2 = 0.44$$

$$n = 3$$

$$w_0 = 0.27 + 0.1 * (3 - (0.27 + 3 * 0.44)) * 1 = 0.411$$

$$w_1 = 0.44 + 0.1 * (3 - (0.27 + 3 * 0.44)) * 3 = 0.863$$

(4.1)

$$n = 1$$

$$w_0 = 0.411 + 0.1 * (1 - (0.411 + 1 * 0.863)) * 1 = 0.3836$$

$$w_1 = 0.863 + 0.1 * (1 - (0.411 + 1 * 0.863)) * 1 = 0.8356$$

$$n = 2$$

$$w_0 = 0.3836 + 0.1 * (2 - (0.3836 + 2 * 0.8356)) * 1 = 0.37812$$

$$w_1 = 0.8356 + 0.1 * (2 - (0.3836 + 2 * 0.8356)) * 2 = 0.82464$$

$$n = 3$$

$$w_0 = 0.37812 + 0.1 * (3 - (0.37812 + 3 * 0.82464)) * 1 = 0.392916$$

$$w_1 = 0.82464 + 0.1 * (3 - (0.37812 + 3 * 0.82464)) * 3 = 0.869028$$