CMSC510 - Assignment 1

Steven M. Hernandez, V00528395

September 2020

1 Derivations

$$\begin{aligned} L(h(x,w),y) &= (h(x,w)-y)^2 \\ &= \left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)^2 \\ &= \left(w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y\right)^2 \end{aligned} \tag{1}$$

To calculate $\nabla_w L(x,y)$, we need to find the partial derivatives of $L(x,y)$. Using the chain rule, $L(x,y) = f(g(x))$ where $f(x) = x^2$ and $g(x) = \left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)$, thus the derivative can be calculated as $L'(x,y) = f'(g(x))g'(x)$. We know for all w_j, $f'(g(x)) = 2\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)$. Thus, our final question is determining $g'(x)$ for all w_j.

$$\begin{aligned} \frac{\delta g}{\delta w_0} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_0} w_0 = 1 \end{aligned} \tag{2}$$

$$\begin{aligned} \frac{\delta g}{\delta w_1} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_1} w_1 x = x \end{aligned} \tag{3}$$

$$\begin{aligned} \frac{\delta g}{\delta w_2} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_2} w_2 x^2 = x^2 \end{aligned} \tag{4}$$

$$\begin{aligned} \frac{\delta g}{\delta w_3} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_3} w_3 x^3 = x^3 \end{aligned} \tag{5}$$

$$\begin{aligned} \frac{\delta g}{\delta w_4} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_4} w_4 x^4 = x^4 \end{aligned} \tag{6}$$

$$\begin{aligned} \frac{\delta g}{\delta w_5} w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5 - y \\ = \frac{\delta g}{\delta w_5} w_5 x^5 = x^5 \end{aligned} \tag{7}$$

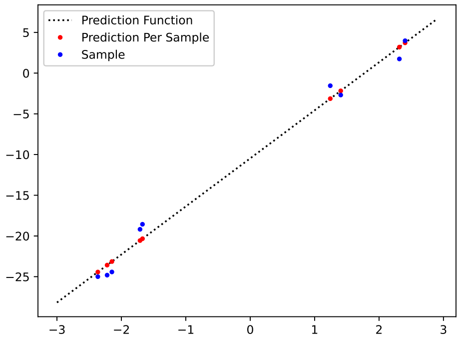
Which we can describe simply as $\frac{\delta g}{\delta w_j} = x^j$. Thus, we know that $\frac{\delta L}{\delta w_j}(x,y) = 2x^j\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)$. Written out in full:

$$
\begin{aligned}
\frac{\delta L}{\delta w_0}(x,y) &= 2\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)\\
\frac{\delta L}{\delta w_1}(x,y) &= 2x\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)\\
\frac{\delta L}{\delta w_2}(x,y) &= 2x^2\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)\\
\frac{\delta L}{\delta w_3}(x,y) &= 2x^3\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)\\
\frac{\delta L}{\delta w_4}(x,y) &= 2x^4\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)\\
\frac{\delta L}{\delta w_5}(x,y) &= 2x^5\left(\left(\sum_{j=0}^{n} w_j x^j\right) - y\right)
\end{aligned}
\tag{8}
$$

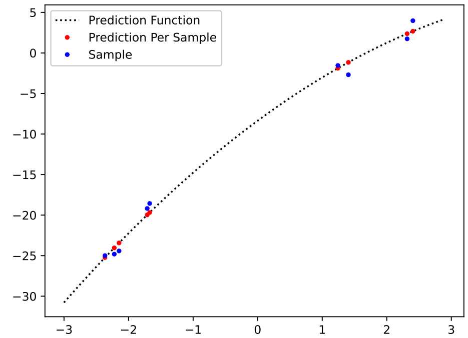
so for $n = 5$:

$$
\nabla_w L = \left\{\frac{\delta L}{\delta w_0}, \frac{\delta L}{\delta w_1}, \frac{\delta L}{\delta w_2}, \frac{\delta L}{\delta w_3}, \frac{\delta L}{\delta w_4}, \frac{\delta L}{\delta w_5}\right\}
\tag{9}
$$

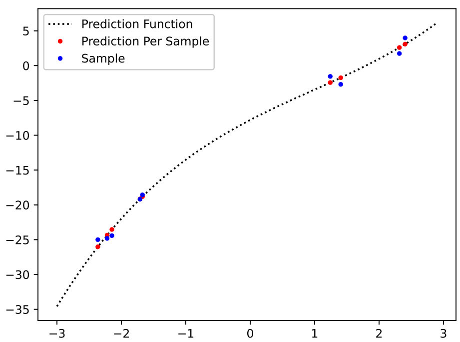
2 Plots



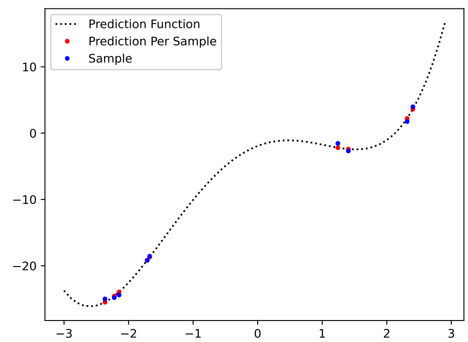
(a) $n = 1$



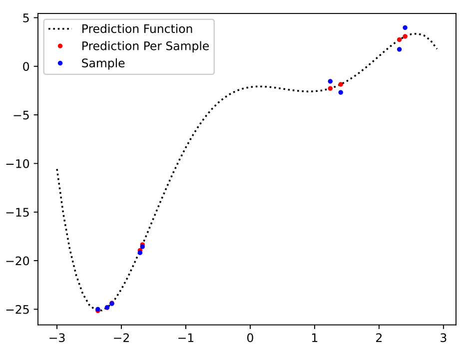
(b) $n = 2$



(c) $n = 3$



(d) $n = 4$

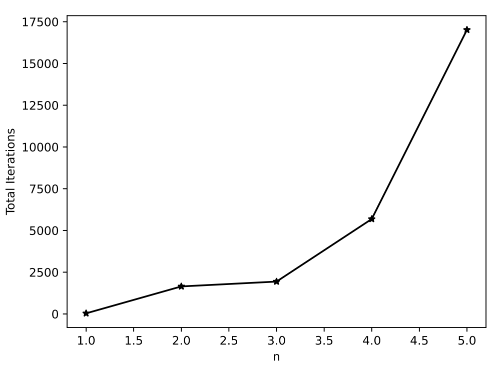


(e) $n = 5$

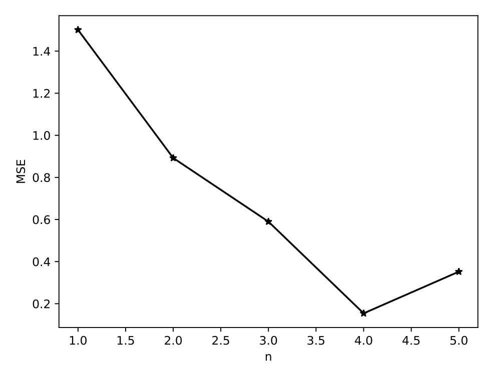
Figure 1: Predictions per Sample versus actual Sample values

n	γ	threshold
1	10^{-1}	10^{-5}
2	10^{-2}	10^{-5}
3	10^{-2}	10^{-5}
4	10^{-3}	10^{-5}
5	10^{-4}	10^{-5}

Table 1: Parameters used for training (selected experimentally)



(a) As n increases, the number of iterations also increases



(b) MSE decrease up until $n = 4$.

Figure 2: To decide the best value of n, we must consider the trade-offs with time to compute (number of iterations) and the accuracy achieved or Mean Squared Error (MSE)