



Department of Computer Engineering

Artificial Intelligence

Assignment 6 Part 1

Dr. Rohban

Parsa Mohammadian — 98102284

December 17, 2021

Contents

| | | |
|----------|-------|----------|
| 1 | | 1 |
| 1.1 | | 1 |
| 1.2 | | 1 |

1

1.1

$$\begin{aligned}
f(w_2, w_1, w_0) &= \sum_i (\hat{y}^i - y^i)^2 = \sum_i (w_2 x_i^2 + w_1 x_i + w_0 - y^i)^2 \\
&= \sum_i w_2^2 x_i^4 + w_2 x_i^2 w_1 x_i + w_2 x_i^2 w_0 - w_2 x_i^2 y^i + w_1 x_i w_2 x_i^2 + w_1^2 x_i^2 \\
&\quad + w_1 x_i w_0 - w_1 x_i y^i + w_0 w_2 x_i^2 + w_0 w_1 x_i + w_0^2 - w_0 y^i \\
&\quad - y^i w_2 x_i^2 - y^i w_1 x_i - y^i w_0 - (y^i)^2 \\
&= \sum_i w_2^2 x_i^4 + 2w_2 x_i^3 w_1 + 2w_2 x_i^2 w_0 - 2w_2 x_i^2 y^i + w_1^2 x_i^2 \\
&\quad + 2w_1 x_i w_0 - 2w_1 x_i y^i + w_0^2 - 2w_0 y^i - (y^i)^2 \\
\nabla f &= \begin{bmatrix} \frac{\partial f(w_2, w_1, w_0)}{\partial w_0} \\ \frac{\partial f(w_2, w_1, w_0)}{\partial w_1} \\ \frac{\partial f(w_2, w_1, w_0)}{\partial w_2} \end{bmatrix} = \begin{bmatrix} \sum_i 2w_2 x_i^2 + 2w_1 x_i + 2w_0 - 2y^i \\ \sum_i 2w_2 x_i^3 + 2w_1 x_i^2 + 2x_i w_0 - 2x_i y^i \\ \sum_i 2w_2 x_i^4 + 2x_i^3 w_1 + 2x_i^2 w_0 - 2x_i^2 y^i \end{bmatrix}
\end{aligned}$$

1.2

$$\begin{aligned}
w^+ &= w - \alpha \nabla f(w) \\
\begin{bmatrix} w_0^+ \\ w_1^+ \\ w_2^+ \end{bmatrix} &= \begin{bmatrix} w_0 - \alpha \frac{\partial f(w_2, w_1, w_0)}{\partial w_0} \\ w_1 - \alpha \frac{\partial f(w_2, w_1, w_0)}{\partial w_1} \\ w_2 - \alpha \frac{\partial f(w_2, w_1, w_0)}{\partial w_2} \end{bmatrix} \\
&= \begin{bmatrix} w_0 - \alpha \sum_i 2w_2 x_i^2 + 2w_1 x_i + 2w_0 - 2y^i \\ w_1 - \alpha \sum_i 2w_2 x_i^3 + 2w_1 x_i^2 + 2x_i w_0 - 2x_i y^i \\ w_2 - \alpha \sum_i 2w_2 x_i^4 + 2x_i^3 w_1 + 2x_i^2 w_0 - 2x_i^2 y^i \end{bmatrix}
\end{aligned}$$

By increasing the learning rate α to a large number, it is possible that algorithm does not converge. At the other hand, when we decrease the learning rate enough, we can be sure that algorithm converges to optimal solution, But the smaller the learning rate, the more time it takes to converge. We should choose the learning rate that is appropriate for our problem.