

# Machine Learning

## MW 3

Armen Mkrtumyan

18/02/25

1.

$x_1$	$x_2$	$x_3$	$y$
5	8	6	1
6	4	3	1
8	5	9	0
7	9	1	0

$$\eta = 0.1$$

$$w_1 = 0.08$$

$$w_2 = 0.06$$

$$w_3 = 0.03$$

$$b = 0$$

Epoch 1)  $z_1 = 0.08 \cdot 5 + 0.06 \cdot 8 + 0.03 \cdot 6 + 0 = 1.06$

$$\sigma(1.06) = \frac{1}{1 + e^{-1.06}} \approx 0.742$$

$$\text{error}_1 = 0.742 - 1 = -0.258$$

$$\frac{\partial L}{\partial w_1} = (\text{error}_1) x_1 = -0.258 \cdot 5 = -1.29$$

$$\frac{\partial L}{\partial w_2} = -0.258 \cdot 8 = -2.064$$

$$\frac{\partial L}{\partial w_3} = -0.258 \cdot 6 = -1.548$$

$$\frac{\partial L}{\partial b} = -0.258$$

$$w_1 = 0.08 - 0.1(-1.29) = 0.209$$

$$w_2 = 0.06 - 0.1(-2.064) = 0.2664$$

$$w_3 = 0.03 - 0.1(-1.548) = 0.1848$$

$$b = 0 - 0.258 \cdot 0.1 = -0.0258$$



$$z_2 = 0.209 \cdot 6 + 0.2664 \cdot 4 + 0.1848 \cdot 3 + 0.0258 \approx 2.8998$$

$$\sigma(2.8998) = \frac{1}{1 + e^{-2.8998}} \approx 0.948$$

$$\text{error}_2 = 0.948 - 1 = -0.052$$

$$\frac{\partial L}{\partial w_1} = -0.052 \cdot 6 = -0.312$$

$$w_1 = 0.209 - 0.1(-0.312) = 0.2402$$

$$\frac{\partial L}{\partial w_2} = -0.052 \cdot 4 = -0.208$$

$$w_2 = 0.2664 - 0.1(-0.208) = 0.2872$$

$$\frac{\partial L}{\partial w_3} = -0.052 \cdot 3 = -0.156$$

$$w_3 = 0.1848 - 0.1(-0.156) = 0.2004$$

$$\frac{\partial L}{\partial b} = -0.052 = -0.052$$

$$b = 0.0258 - 0.1(-0.052) = 0.031$$

$$z_3 = 0.2402 \cdot 8 + 0.2872 \cdot 5 + 0.2004 \cdot 9 + 0.031 \approx 5.1922$$

$$\sigma(5.1922) = \frac{1}{1 + e^{-5.1922}} \approx 0.9945$$

$$\text{error}_3 = 0.9945 - 0 = 0.9945$$

$$\frac{\partial L}{\partial w_1} = 0.9945 \cdot 8 = 7.956$$

$$w_1 = 0.2402 - 0.1 \cdot 7.956 = -0.5554$$

$$\frac{\partial L}{\partial w_2} = 0.9945 \cdot 5 = 4.9725$$

$$w_2 = 0.2872 - 0.1 \cdot 4.9725 = -0.2101$$



$$\frac{\partial L}{\partial w_3} = 0.9945 \cdot 9 = 8.9505 \quad w_3 = 0.2004 - 0.1 \cdot 8.9505 = -0.6947$$

$$\frac{\partial L}{\partial b} = 0.9945 \quad b = 0.031 - 0.1 \cdot 0.9945 = -0.0685$$

$$z_4 = -0.5554 \cdot 7 - 0.2101 \cdot 9 - 0.6947 \cdot 1 - 0.0685 \approx -6.5419$$

$$\sigma(-6.5419) = \frac{1}{1 + e^{+6.5419}} \approx 0.0014$$

$$\text{error}_4 = (0.0014) - 0 = 0.0014$$

$$\frac{\partial L}{\partial w_1} = 0.0014 \cdot 7 = 0.0098 \quad w_1 = -0.5554 - 0.1 \cdot 0.0098 \approx -0.5564$$

$$\frac{\partial L}{\partial w_2} = 0.0014 \cdot 9 = 0.0126 \quad w_2 = -0.2101 - 0.1 \cdot 0.0126 \approx -0.2114$$

$$\frac{\partial L}{\partial w_3} = 0.0014 \cdot 1 = 0.0014 \quad w_3 = -0.6947 - 0.1 \cdot 0.0014 \approx -0.6948$$

$$\frac{\partial L}{\partial b} = 0.0014 \quad b = -0.0685 - 0.1(0.0014) = -0.0686$$

Epoch 2)

$$z_1 = -0.5564 \cdot 5 - 0.2114 \cdot 8 - 0.6948 \cdot 6 - 0.0686 = -8.7106$$

$$\sigma(-8.7106) = \frac{1}{1 + e^{+8.7106}} = 0.00016$$

$$\text{error}_1 = 0.00016 - 1 = -0.99984$$



$$\frac{\partial L}{\partial w_1} = 0.99984 \cdot 5 = -4.9992 \quad w_1 = -0.5564 - 0.1(-4.9992) \approx -0.0565$$

$$\frac{\partial L}{\partial w_2} = 0.99984 \cdot 8 = -7.99872 \quad w_2 = -0.2114 - 0.1(-7.99872) \approx 0.5885$$

$$\frac{\partial L}{\partial w_3} = 0.99984 \cdot 6 = -5.99904 \quad w_3 = -0.6948 - 0.1(-5.99904) \approx -0.0949$$

$$\frac{\partial L}{\partial b} = -0.99984 \quad b = -0.0686 - 0.1(-0.99984) \approx 0.0314$$

$$Z_2 = -0.0565 \cdot 6 + 0.5885 \cdot 4 - 0.0949 \cdot 3 + 0.0314 = 1.7617$$

$$\sigma(1.7617) = \frac{1}{1 + e^{-1.7617}} \approx 0.854$$

$$\text{error}_2 = 0.854 - 1 = -0.146$$

$$\frac{\partial L}{\partial w_1} = -0.146 \cdot 6 = -0.876 \quad w_1 = -0.0565 - 0.1(-0.876) = 0.0311$$

$$\frac{\partial L}{\partial w_2} = -0.146 \cdot 4 = -0.584 \quad w_2 = 0.5885 - 0.1(-0.584) = 0.6469$$

$$\frac{\partial L}{\partial w_3} = -0.146 \cdot 3 = -0.438 \quad w_3 = -0.0949 - 0.1(-0.438) = -0.0511$$

$$\frac{\partial L}{\partial b} = -0.146 = 0.146 \quad b = 0.0314 - 0.1(-0.146) = 0.046$$



$$z_3 = 0.0311 \cdot 8 + 0.6469 \cdot 5 - 0.0511 \cdot 9 + 0.046 = 3.0694$$

$$\sigma(3.0694) = \frac{1}{1 + e^{-3.0694}} \approx 0.957$$

$$\text{error}_3 = 0.957 - 0 = 0.957$$

$$\frac{\partial L}{\partial w_1} = 0.957 \cdot 8 = 7.656 \quad w_1 = 0.0311 - 0.1(7.656) = -0.7345$$

$$\frac{\partial L}{\partial w_2} = 0.957 \cdot 5 = 4.785 \quad w_2 = 0.6469 - 0.1(4.785) = 0.1684$$

$$\frac{\partial L}{\partial w_3} = 0.957 \cdot 9 = 8.613 \quad w_3 = -0.0511 - 0.1(8.613) = -0.9124$$

$$\frac{\partial L}{\partial b} = 0.957 \quad b = 0.046 - 0.1(0.957) = -0.0497$$

$$z_4 = -0.7345 \cdot 7 + 0.1684 \cdot 9 - 0.9124 \cdot 1 - 0.0497 \approx -4.588$$

$$\sigma(-4.588) = \frac{1}{1 + e^{+4.588}} \approx 0.0101$$

$$\text{error}_4 = 0.0101 - 0 = 0.0101$$

$$\frac{\partial L}{\partial w_1} = 0.0101 \cdot 7 = 0.0707 \quad w_1 = -0.7345 - 0.1(0.0707) \approx \underline{-0.7416}$$

$$\frac{\partial L}{\partial w_2} = 0.0101 \cdot 9 = 0.0909 \quad w_2 = 0.1684 - 0.1(0.0909) \approx \underline{0.1593}$$

$$\frac{\partial L}{\partial w_3} = 0.0101 \cdot 1 = 0.0101 \quad w_3 = -0.9124 - 0.1(0.0101) \approx \underline{-0.9134}$$

$$\frac{\partial L}{\partial b} = 0.0101 = 0.0101 \quad b = -0.0497 - 0.1(0.0101) \approx \underline{-0.0507}$$



Now let's do with regularization ( $\lambda = 0.5$ )

Each gradient will have the term  $\frac{\lambda}{n} w_j = \frac{0.5}{4} w_j = 0.125 w_j$  added to the end of its calculation

Epoch 1)

$$z_1 = 1.06 \text{ (recall from previous one)}$$

$$\sigma(1.06) \approx 0.742$$

$$\text{error}_1 = -0.258$$

$$\frac{\partial L}{\partial w_1} = -0.258 \cdot 5 + 0.125 \cdot 0.08 = -1.28 \quad w_1 = 0.08 - 0.1(-1.28) = 0.208$$

$$\frac{\partial L}{\partial w_2} = -0.258 \cdot 8 + 0.125 \cdot 0.06 = -2.0565 \quad w_2 = 0.06 - 0.1(-2.0565) = 0.2657$$

$$\frac{\partial L}{\partial w_3} = -0.258 \cdot 6 + 0.125 \cdot 0.03 = -1.54425 \quad w_3 = 0.03 - 0.1(-1.54425) = 0.1844$$

$$\frac{\partial L}{\partial b} = -0.258 \quad b = 0 - 0.1(-0.258) = 0.0258$$

$$z_2 = 6 \cdot 0.208 + 0.2657 \cdot 4 + 0.1844 \cdot 3 + 0.0258 = 2.8898$$

$$\sigma(2.8898) = \frac{1}{1 + e^{-2.8898}} = 0.947$$

$$\text{error}_2 = 0.947 - 1 = -0.053$$

$$\frac{\partial L}{\partial w_1} = -0.053 \cdot 6 + 0.125 \cdot 0.208 = -0.292$$

$$w_1 = 0.208 - 0.1(-0.292) = 0.2372$$



$$\frac{\partial L}{\partial w_2} = -0.053 \cdot 4 + 0.125(0.2652) = -0.17879$$

$$w_2 = 0.2652 - 0.1(-0.17879) = 0.2836$$

$$\frac{\partial L}{\partial w_3} = -0.053 \cdot 3 + 0.125(0.1844) = -0.13595$$

$$w_3 = 0.1844 - 0.1(-0.13595) = 0.197$$

$$\frac{\partial L}{\partial b} = -0.053$$

$$b = 0.0258 - 0.1(-0.053) = 0.0311$$

$$z_3 = 0.2372 \cdot 8 + 0.2836 \cdot 5 + 0.197 \cdot 9 + 0.0311 = 5.1197$$

$$\sigma(5.1197) = \frac{1}{1 + e^{-5.1197}} = 0.984$$

$$\text{error}_3 = 0.984 - 0 = 0.984$$

$$\frac{\partial L}{\partial w_1} = 0.984 \cdot 8 + 0.125 \cdot 0.2372 = 7.98165 \quad w_1 = 0.2372 - 0.1(7.98165) \approx -0.5610$$

$$\frac{\partial L}{\partial w_2} = 0.984 \cdot 5 + 0.125 \cdot 0.2836 = 5.00545 \quad w_2 = -0.2836 - 0.1(5.00545) \approx -0.2169$$

$$\frac{\partial L}{\partial w_3} = 0.984 \cdot 9 + 0.125 \cdot 0.197 = 8.970625 \quad w_3 = 0.197 - 0.1(8.970625) \approx -0.7001$$

$$\frac{\partial L}{\partial b} = 0.984 \quad b = 0.0311 - 0.1(0.984) = -0.0683$$

$$z_4 = -0.5610 \cdot 7 - 0.2169 \cdot 9 - 0.7001 \cdot 1 - 0.0683 = -6.6475$$

$$\sigma(-6.6475) = \frac{1}{1 + e^{6.6475}} = 0.0013$$

$$\text{error}_4 = 0.0013 - 0 = 0.0013$$



$$\frac{\partial L}{\partial w_1} = 0.0013 \cdot 7 + 0.125 \cdot (-0.5610) = -0.061025$$

$$\frac{\partial L}{\partial w_2} = 0.0013 \cdot 9 + 0.125 \cdot (-0.2169) = -0.0151125$$

$$\frac{\partial L}{\partial w_3} = 0.0013 \cdot 1 + 0.125 \cdot (-0.7001) = -0.0862125$$

$$\frac{\partial L}{\partial b} = 0.0013$$

$$w_1 = -0.5610 - 0.1(-0.061025) \approx -0.5549$$

$$w_2 = -0.2169 - 0.1(-0.0151125) \approx -0.2154$$

$$w_3 = -0.7001 - 0.1(-0.0862125) \approx -0.6915$$

$$b = -0.0683 - 0.1(0.0013) \approx -0.0684$$

Epoch 2)

$$z_1 = -0.5549 \cdot 5 - 0.2154 \cdot 8 - 0.6915 \cdot 6 - 0.0684 \approx -8.7151$$

$$\sigma(-8.7151) = \frac{1}{1 + e^{8.7151}} \approx 0.00016$$

$$\text{error}_1 = 0.00016 - 1 = -0.99984$$

$$\frac{\partial L}{\partial w_1} = -0.99984 \cdot 5 + 0.125(-0.5549) = -5.0685625$$

$$\frac{\partial L}{\partial w_2} = -0.99984 \cdot 8 + 0.125(-0.2154) = -8.025645$$

$$\frac{\partial L}{\partial w_3} = -0.99984 \cdot 6 + 0.125(-0.6915) = -6.0854775$$

$$\frac{\partial L}{\partial b} = -0.99984$$



$$w_1 = -0.5549 - 0.1(-5.0685625) \approx -0.0480$$

$$w_2 = -0.2154 - 0.1(-8.025645) \approx 0.5871645 \approx 0.5872$$

$$w_3 = -0.6915 - 0.1(-6.0854775) \approx -0.083$$

$$b = -0.0684 - 0.1(-0.99984) \approx 0.0316$$

$$z_2 = -0.0480 \cdot 6 + 0.5872 \cdot 4 - 0.083 \cdot 3 + 0.0316 \approx 1.8434$$

$$\sigma(1.8434) = \frac{1}{1 + e^{-1.8434}} = 0.8634$$

$$\text{error}_2 = 0.8634 - 1 \approx -0.137$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = -0.137 \cdot 6 + 0.125(-0.0480) = -0.828$$

$$\frac{\partial \mathcal{L}}{\partial w_2} = -0.137(4) + 0.125(0.5872) = -0.4746$$

$$\frac{\partial \mathcal{L}}{\partial w_3} = -0.137 \cdot 3 + 0.125(-0.083) = -0.42375$$

$$\frac{\partial \mathcal{L}}{\partial b} = -0.137$$

$$w_1 = -0.0480 - 0.1(-0.828) = 0.0348$$

$$w_2 = 0.5872 - 0.1(-0.4746) = 0.6347$$

$$w_3 = -0.083 - 0.1(-0.42375) = -0.0409$$

$$b = 0.0316 - 0.1(-0.137) = 0.0453$$

$$z_3 = 0.0348 \cdot 8 + 0.6347 \cdot 5 - 0.0409 \cdot 9 + 0.0453 = 3.1291$$

$$\sigma(3.1291) = \frac{1}{1 + e^{-3.1291}} \approx 0.958$$

$$\text{error}_3 = 0.958 - 0 = 0.958$$



$$\frac{\partial L}{\partial w_1} = 0.958(8) + 0.125(0.0348) = 7.66835$$

$$\frac{\partial L}{\partial w_2} = 0.958(5) + 0.125(0.6347) = 4.8693375$$

$$\frac{\partial L}{\partial w_3} = 0.958(9) + 0.125(-0.0409) = 8.6168875$$

$$\frac{\partial L}{\partial b} = 0.958 + 0 = 0.958$$

$$w_1 = 0.0348 - 0.1(7.66835) = -0.732$$

$$w_2 = 0.6347 - 0.1(4.8693375) = 0.1478$$

$$w_3 = -0.0409 - 0.1(8.6168875) = -0.9026$$

$$b = 0.0453 - 0.1(0.958) = -0.0505$$

$$z_4 = -0.732 \cdot 7 + 0.1478 \cdot 9 - 0.9026 \cdot 1 - 0.0505 = -4.7469$$

$$\sigma(-4.7469) = \frac{1}{1 + e^{4.7469}} \approx 0.0086$$

$$\text{error}_4 = 0.0086 - 0 = 0.0086$$

$$\frac{\partial L_4}{\partial w_1} = 0.0086(7) + 0.125(-0.732) = -0.0313$$

$$\frac{\partial L_4}{\partial w_2} = 0.0086(9) + 0.125(0.1478) = 0.095875$$

$$\frac{\partial L_4}{\partial w_3} = 0.0086(1) + 0.125(-0.9026) = -0.104225$$

$$\frac{\partial L_4}{\partial b} = 0.0086$$



$$w_1 = -0.732 - 0.1(-0.0313) = \underline{-0.7289}$$

$$w_2 = 0.1478 - 0.1(0.095875) = \underline{0.1382}$$

$$w_3 = -0.9026 - 0.1(-0.104225) = \underline{-0.8922}$$

$$b = -0.0505 - 0.1(0.0086) = \underline{-0.0514}$$

Let's compare the results

No regularization

Regularization

$$w_1 = -0.7416$$

$$-0.7289$$

$$w_2 = 0.1593$$

$$0.1382$$

$$w_3 = -0.9134$$

$$-0.8922$$

$$b = -0.0507$$

$$-0.0514$$

Although it's not toooo big, we still can see that L2 regularization has made the weights smaller than for the "no regularization" case. It brought them more toward zero.