Machine Learning

Armen Mkrtumyan
08/02/25

HW 1

Ex 1.

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

$\alpha = 0.1$
$w_1 = 0.08$
$w_2 = 0.06$
$w_3 = 0.03$
$b = 0$

Epoch 1.

$z_1 = 5 \cdot 0.08 + 0.06 \cdot 8 + 0.03 \cdot 6 + 0 = 1.05 \quad G(1.05) = 1.06$
$error_1 = 1 - 1.05 = 0.06$
$z_2 = 6 \cdot 0.08 + 4 \cdot 0.06 + 3 \cdot 0.03 + 0 = 0.81 \quad G(0.81) = 0.81$
$error_2 = 1 - 0.81 = 0.19$
$z_3 = 8 \cdot 0.08 + 5 \cdot 0.06 + 9 \cdot 0.03 + 0 = 1.21 \quad G(1.21) = 1.21$
$error_3 = 0 - 1.21 = -1.21$
$z_4 = 7 \cdot 0.08 + 9 \cdot 0.06 + 1 \cdot 0.03 + 0 = 1.13 \quad G(1.13) = 1.13$
$error_4 = 0 - 1.13 = -1.13$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} error_i^2 = \frac{1}{4} \left((0.06)^2 + 0.19^2 + (-1.21)^2 + (-1.13)^2 \right) \approx 0.7$$

$$\frac{\partial L}{\partial w_j} = -\frac{2}{n} \sum_i \left(y^i - G(z^i) \right) x_i$$

$$\Delta w_j = \alpha \cdot \frac{\partial L}{\partial w_j}$$

$$w_j = w_j + \Delta w_j$$

$$\frac{\partial L}{\partial w_1} = -\frac{2}{4}(-0.06 \cdot 5 + 0.19 \cdot 6 - 1.21 \cdot 8 - 1.13 \cdot 7) = +8.38$$

$$\frac{\partial L}{\partial w_2} = -\frac{2}{4}(-0.06 \cdot 8 + 0.19 \cdot 4 - 1.21 \cdot 5 - 1.13 \cdot 9) = +7.97$$

$$\frac{\partial L}{\partial w_3} = -\frac{2}{4}(-0.06 \cdot 6 + 0.19 \cdot 3 - 1.21 \cdot 9 - 1.13 \cdot 4) = +5.9$$

$$\frac{\partial L}{\partial b} = -\frac{2}{4}(-0.06 + 0.19 - 1.21 - 1.13) = +1.11$$

$\Delta w_1 = -0.1 \cdot (+8.38) = -0.838$
$\Delta w_2 = -0.1 \cdot (+7.97) = -0.797$
$\Delta w_3 = -0.1 \cdot (+5.9) = -0.59$
$\Delta b = -0.1 \cdot (+1.11) = -0.111$

$w_1 = 0.08 - 0.838 = -0.758 \approx -0.76$
$w_2 = 0.06 - 0.797 = -0.737 \approx -0.74$
$w_3 = 0.03 - 0.59 = -0.56 \approx -0.56$
$b = 0 - 0.111 = -0.111 \approx -0.11$

Epoch 2

$z_1 = 5 \cdot (-0.76) + 8 \cdot (-0.74) + 6 \cdot (-0.56) - 0.11 = -13.19$
$G(-13.19) = -13.19$ $\quad error_1 = 1 - (-13.19) = 14.19$
$z_2 = 6 \cdot (-0.76) + 4(-0.74) + 3(-0.56) - 0.11 = -9.31$
$G(-9.31) = -9.31$ $\quad error_2 = 1 - (-9.31) = 10.31$

$z_3 = 8 \cdot (-0.76) + 5(-0.74) + 9(-0.56) - 0.11 = -14.93$

$G(-14.93) = -14.93 \quad error = 0 - (-14.93) = 14.93$

$z_4 = 7(-0.76) + 9(-0.74) + 1(-0.56) - 0.11 = -12.65$

$G(-12.65) = -12.65 \quad error = 0 - (-12.65) = 12.65$

$MSE = \frac{1}{4}(14.19^2 + 10.31^2 + 14.93^2 + 12.65^2) = 172.65$

$\frac{\partial L}{\partial w_1} = -0.5(14.19 \cdot 5 + 10.31 \cdot 6 + 14.93 \cdot 8 + 12.65 \cdot 7) = -170.4$

$\frac{\partial L}{\partial w_2} = -0.5(14.19 \cdot 8 + 10.31 \cdot 4 + 14.93 \cdot 5 - 12.65 \cdot 9) = -171.63$

$\frac{\partial L}{\partial w_3} = -0.5(14.19 \cdot 6 + 10.31 \cdot 3 + 14.93 \cdot 9 + 12.65 \cdot 1) = -131.55$

$\frac{\partial L}{\partial b} = -0.5(14.19 + 10.31 + 14.93 + 12.65) = -26.04$

$\Delta w_1 = -0.1(-170.4) = 17.04 \quad w_1 = -0.76 + 17.04 = 16.28$

$\Delta w_2 = -0.1(-171.63) = 17.16 \quad w_2 = -0.74 + 17.16 = 16.42$

$\Delta w_3 = -0.1(-131.55) = 13.16 \quad w_3 = -0.56 + 13.16 = 12.59$

$\Delta b = -0.1(-26.04) = 2.6 \quad b = -0.11 + 26 = 2.49$

Epoch 3

$z_1 = 5 \cdot 16.28 + 8 \cdot 16.42 + 6 \cdot 12.59 + 2.49 = 290.79$

$G(290.79) = 290.79 \quad error_1 = 1 - 290.79 = -289.79$

$z_2 = 6 \cdot 16.28 + 4 \cdot 16.42 + 3 \cdot 12.59 + 2.49 = 203.62$

$G(203.62) = 203.62 \quad error_2 = 1 - 203.62 = -202.62$

$z_3 = 8 \cdot 16.28 + 5 \cdot 16.42 + 9 \cdot 12.59 + 2.49 = 328.14$

$G(328.14) = 328.14 \quad error_3 = 0 - 328.14 = -328.14$

$z_4 = 7 \cdot 16.28 + 9 \cdot 16.42 + 1 \cdot 12.59 + 2.49 = 276.82$

$G(276.82) = 276.82 \quad error_4 = 0 - 276.82 = -276.82$

$MSE = \frac{1}{4}\left((-289.79)^2 + (-202.62)^2 + (-328.14)^2 + (-276.82)^2\right) =$

$\approx 77378.$

$\frac{\partial L}{\partial w_1} = -0.5(-289.79 \cdot 5 - 202.62 \cdot 6 - 328.14 \cdot 8 - 276.82 \cdot 7 \approx 3613.77$

$\frac{\partial L}{\partial w_2} = -0.5(-289.79 \cdot 8 - 202.62 \cdot 4 - 328.14 \cdot 5 - 276.82 \cdot 9 \approx 3630.44$

$\frac{\partial L}{\partial w_3} = -0.5(-289.79 \cdot 6 - 202.62 \cdot 3 - 328.14 \cdot 9 - 276.82 \cdot 1) \approx 2788.34$

$\frac{\partial L}{\partial b} = -0.5(-289.79 - 202.62 - 328.14 - 276.82) \approx 548.69$

$\Delta w_1 = -0.1(3613.77) = -361.38 \quad w_1 = 76.28 - 361.38 = -345.1$

$\Delta w_2 = -0.1(3630.44) = -363.04 \quad w_2 = 16.42 - 363.04 = -346.62$

$\Delta w_3 = -0.1(2788.34) = -278.83 \quad w_3 = 12,59 - 278.83 = -266.24$

$\Delta b = -0.1(548.69) = -54.87 \quad b = 2.49 - 54,87 = -52.38$

Epoch 4.

$z_1 = 5(-345.1) + 8(-346.62) + 6(-266.24) - 52,38 = -6148.28$

$G(-6148.28) = -6148.28 \quad error_1 = 1 - (-6148.28) = 6149.28$

$z_2 = 6(-345.1) + 4(-346.62) + 3(-266.24) - 52.38 = -4308.18$

$G(-4308.18) = -4308.18 \quad error_2 = 1 - (-4308.18) = 4309.18$

$z_3 = 8 \cdot (-345.1) + 5(-346.62) + 9(-266.24) - 52.38 = -6941.44$

$G(-6941.44) = -6941.44 \quad error_3 = 0 - (-6941.44) = 6941.44$

$z_4 = 7(-345.1) + 9(-346.62) + 1(-266.24) - 52.38 = -5853.9$

$G(-5853.9) = -5853.9 \quad error_4 = 0 - (-5853.9) = 5853.9$

$MSE = \frac{1}{4}(6149.28^2 + 4309.18^2 + 6941.44^2 + 5853.9^2) \approx 34.31 \cdot 10^6$

$\frac{\partial L}{\partial w_1} = -0.5(6149.28 \cdot 5 + 4309.18 \cdot 6 + 6941.44 \cdot 8 + 5853.9 \cdot 8) \approx -76555$

$\frac{\partial L}{\partial w_2} = -0.5(6149.28 \cdot 8 + 4309.18 \cdot 4 + 6941.44 \cdot 5 + 5853.9 \cdot 5) \approx -76912$

$\frac{\partial L}{\partial w_3} = -0.5(6149.28 \cdot 6 + 4309.18 \cdot 3 + 6941.44 \cdot 9 + 5853.9 \cdot 2) \approx -59075$

$\frac{\partial L}{\partial b} = -0.5(6149.28 + 4309.18 + 6941.44 + 5853.9) \approx -11627$

$\Delta w_1 = -0.1(-76555) = 7655.5 \quad w_1 = -345.1 + 7655.5 = 7310.44$

$\Delta w_2 = -0.1(-76912) = 7691.2 \quad w_2 = -346.62 + 7691.2 = 7344.54$

$\Delta w_3 = -0.1(-59075) = 5907.5 \quad w_3 = -266.24 + 5907.5 = 5641.26$

$\Delta b = -0.1(-11627) = 1162.7 \quad b = -52.38 + 1162.7 = 1110.32$

Our predictions are $\begin{cases} 1; z >= 0.5 \\ 0; \text{otherwise} \end{cases}$

1. $5 \cdot 7310.44 + 8 \cdot 7344.54 + 6 \cdot 5641.26 + 1110.32 >> 0.5 \Rightarrow$ predict 1
2. $6 \cdot 7310.44 + 4 \cdot 7344.54 + 3 \cdot 5641.26 + 1110.32 >> 0.5 \Rightarrow$ predict 1
3. $8 \cdot 7310.44 + 5 \cdot 7344.54 + 9 \cdot 5641.26 + 1110.32 >> 0.5 \Rightarrow$ predict 1
4. $7 \cdot 7310.44 + 9 \cdot 7344.54 + 1 \cdot 5641.26 + 1110.32 >> 0.5 \Rightarrow$ predict 1

Conclusion →

Conclusion

Our MSE, over each epoch was increasing very quickly. In the end, becoming 34.7 million. So our training process was diverging from true values.

Since weights have become extremely large, & we had huge MSE => learning rate was too high; the term $\frac{2}{n}$ could have been removed, because it was just a constant & we could've configured everything by just using learning rate.

My model predicted all classes as 1, resulting in terrible performance. An absolute disaster!!!

Problem 3

1) No, here is why
   - → 0 misclassified example ≠ minimal MSE (doesn't always mean), we might still achieve better classifier if we continue
   - → classifier might not generalize well for future data (overfit)
2) Yes, here is why
   - → small batch sizes can help to escape local minima & are slow to converge
   - → big batches can lead to faster convergence & provide more accurate gradient estimates