XGBoost Exercises

Course: Data Mining

Professor: Dr. Tahaei

Authors:
Mohammad Saeed Arvenaghi
Shadmehr Salehi

Subject: XGBoost

December 2024

Questions

Question 1: Logistic Gradients + Best Split

We are performing one iteration of XGBoost for binary classification with **logistic loss**. We have the following 5 data points:

Sample i	x_i	y_i	$\hat{y}_i$ (current log-odds)
1	1	0	-0.2
2	2	0	0.0
3	3	1	0.3
4	4	1	1.0
5	5	1	2.0

Assume:

$$\lambda = 1, \quad \gamma = 0.1.$$

(A) For each sample i:

1. Convert $\hat{y}_i$ (log-odds) to probability $p_i = \sigma(\hat{y}_i)$.
2. Compute the gradient $g_i = p_i - y_i$.
3. Compute the hessian $h_i = p_i(1 - p_i)$.

(B) Consider a possible split at $x < 3$ vs. $x \geq 3$:

1. Compute $\sum_{i \in L} g_i$, $\sum_{i \in L} h_i$ for $L = \{1, 2\}$.
2. Compute $\sum_{i \in R} g_i$, $\sum_{i \in R} h_i$ for $R = \{3, 4, 5\}$.
3. Compute $\sum g_i, \sum h_i$ for all 5 points.
4. Calculate the split gain:

$$\text{Gain} = \frac{1}{2}\left(\frac{(\sum_L g_i)^2}{\sum_L h_i + \lambda} + \frac{(\sum_R g_i)^2}{\sum_R h_i + \lambda} - \frac{\left(\sum g_i\right)^2}{\sum h_i + \lambda}\right) - \gamma.$$

(C) Check if Gain > 0. If so, the split is beneficial.

Question 2: Comparing Two Splits + Optimal Leaf Weights

We have 6 samples (x_i, y_i) for a **binary classification** task:

i	x_i	y_i	$\hat{y}_i$ (log-odds)
1	1.5	0	0.2
2	2.1	1	-0.5
3	2.9	0	0.0
4	3.2	0	0.3
5	4.1	1	1.2
6	4.8	1	1.8

Again, let $\lambda = 1$. Now $\gamma = 0.2$. We consider two splits of this single node containing all 6 samples:

$$\text{Split A: } x < 3 \quad \text{vs.} \quad x \geq 3, \qquad \text{Split B: } x < 4 \quad \text{vs.} \quad x \geq 4.$$

(1) Compute p_i, g_i, h_i for each i:

$$p_i = \sigma(\hat{y}_i), \quad g_i = p_i - y_i, \quad h_i = p_i(1 - p_i).$$

(2) For each split:

1. Determine the left set L and the right set R.
2. Compute $\sum_L g_i$, $\sum_L h_i$, $\sum_R g_i$, $\sum_R h_i$, and $\sum g_i, \sum h_i$.
3. Use the same Gain formula (subtracting $\gamma = 0.2$) to see if the split is beneficial.

(3) Compare Gain_A and Gain_B. Whichever is larger is chosen by XGBoost.

(4) Suppose the chosen split is, say, the better one. Compute the **optimal leaf weights**:

$$w^* = -\frac{\sum g_i}{\sum h_i + \lambda}$$

for each leaf.

Solutions

Solution to Question 1

Detailed Derivation of Gradients and Hessians

The logistic loss function L for binary classification is defined as:

$$L = -y \log(p) - (1 - y) \log(1 - p),$$

where $p = \sigma(\hat{y}) = \frac{1}{1+e^{-\hat{y}}}$, and $\hat{y}$ is the log-odds predicted by the model. This loss function measures the discrepancy between the true label $y \in \{0, 1\}$ and the predicted probability $p \in [0, 1]$.

1. Compute the Gradient To derive the gradient, we differentiate the loss L with respect to $\hat{y}$:

$$\frac{\partial L}{\partial \hat{y}} = \frac{\partial L}{\partial p} \cdot \frac{\partial p}{\partial \hat{y}}.$$

First, compute $\frac{\partial L}{\partial p}$:

$$\frac{\partial L}{\partial p} = -\frac{y}{p} + \frac{1 - y}{1 - p}.$$

Next, compute $\frac{\partial p}{\partial \hat{y}}$. Since $p = \sigma(\hat{y}) = \frac{1}{1+e^{-\hat{y}}}$, we have:

$$\frac{\partial p}{\partial \hat{y}} = p(1 - p).$$

Combining the two:

$$\frac{\partial L}{\partial \hat{y}} = \left(-\frac{y}{p} + \frac{1 - y}{1 - p} \right) \cdot p(1 - p).$$

Simplify the expression:

$$\frac{\partial L}{\partial \hat{y}} = p - y.$$

Thus, the gradient for each sample i is:

$$g_i = p_i - y_i.$$

2. Compute the Hessian The Hessian is the second derivative of the loss with respect to $\hat{y}$:

$$\frac{\partial^2 L}{\partial \hat{y}^2} = \frac{\partial}{\partial \hat{y}} (p - y).$$

Since $p - y$ depends on p, we compute the derivative of p with respect to $\hat{y}$:

$$\frac{\partial^2 L}{\partial \hat{y}^2} = \frac{\partial p}{\partial \hat{y}} = p(1 - p).$$

Thus, the Hessian for each sample i is:

$$h_i = p_i(1 - p_i).$$

Summary of Formulas For each sample i:

- Compute the probability p_i from the log-odds $\hat{y}_i$ using the sigmoid function:

$$p_i = \sigma(\hat{y}_i) = \frac{1}{1 + e^{-\hat{y}_i}}.$$

- Compute the gradient:

$$g_i = p_i - y_i.$$

- Compute the Hessian:

$$h_i = p_i(1 - p_i).$$

These calculations are performed for each data point during an iteration of XGBoost.

(A) Compute p_i, g_i, h_i

For the 5 data points:

$$(\hat{y}_1, y_1) = (-0.2,\ 0), \quad (\hat{y}_2, y_2) = (0.0,\ 0), \quad (\hat{y}_3, y_3) = (0.3,\ 1), \quad (\hat{y}_4, y_4) = (1.0,\ 1), \quad (\hat{y}_5, y_5) = (2.0,\ 1).$$

$$p_i = \sigma(\hat{y}_i) = \frac{1}{1 + e^{-\hat{y}_i}}.$$

Let us write approximate numeric values (rounded to 6 decimals):

$$\begin{aligned} p_1 &= \sigma(-0.2) \approx 0.450166, \\ p_2 &= \sigma(0.0) = 0.500000, \\ p_3 &= \sigma(0.3) \approx 0.574443, \\ p_4 &= \sigma(1.0) \approx 0.731059, \\ p_5 &= \sigma(2.0) \approx 0.880797. \end{aligned}$$

$$g_i = p_i - y_i, \quad h_i = p_i(1 - p_i).$$

So:

$$\begin{aligned} g_1 &= 0.450166 - 0 = 0.450166, & h_1 &= 0.450166 \times 0.549834 \approx 0.247516, \\ g_2 &= 0.500000 - 0 = 0.500000, & h_2 &= 0.500000 \times 0.500000 = 0.250000, \\ g_3 &= 0.574443 - 1 = -0.425557, & h_3 &= 0.574443 \times 0.425557 \approx 0.244458, \\ g_4 &= 0.731059 - 1 = -0.268941, & h_4 &= 0.731059 \times 0.268941 \approx 0.196612, \\ g_5 &= 0.880797 - 1 = -0.119203, & h_5 &= 0.880797 \times 0.119203 \approx 0.105050. \end{aligned}$$

(B) Splitting at $x < 3$ vs. $x \geq 3$

$$L = \{1, 2\}, \quad R = \{3, 4, 5\}.$$

$$\sum_L g_i = g_1 + g_2 = 0.450166 + 0.500000 = 0.950166, \quad \sum_L h_i = h_1 + h_2 = 0.247516 + 0.250000 = 0.497516.$$

$$\sum_R g_i = g_3 + g_4 + g_5 = -0.425557 + (-0.268941) + (-0.119203) = -0.813701,$$

$$\sum_R h_i = h_3 + h_4 + h_5 = 0.244458 + 0.196612 + 0.105050 = 0.546120.$$

$$\sum_{i=1}^{5} g_i = 0.950166 + (-0.813701) = 0.136465, \quad \sum_{i=1}^{5} h_i = 0.497516 + 0.546120 = 1.043636.$$

$$\text{Gain} = \frac{1}{2}\Big(\frac{(0.950166)^2}{0.497516+1} + \frac{(-0.813701)^2}{0.546120+1} - \frac{(0.136465)^2}{1.043636+1}\Big) - \gamma.$$

We take $\gamma = 0.1$. Numerically:

$$(0.950166)^2 \approx 0.902815, \quad (-0.813701)^2 \approx 0.662157, \quad (0.136465)^2 \approx 0.018625.$$

$$0.497516 + 1 = 1.497516, \quad 0.546120 + 1 = 1.546120, \quad 1.043636 + 1 = 2.043636.$$

$$\frac{0.902815}{1.497516} \approx 0.603, \quad \frac{0.662157}{1.546120} \approx 0.4283, \quad \frac{0.018625}{2.043636} \approx 0.009119.$$

Inside the parentheses:

$$0.603 + 0.4283 - 0.009119 \approx 1.022181.$$

Half of that is ≈ 0.51109. Finally subtract $\gamma = 0.1$:

$$\text{Gain} \approx 0.51109 - 0.1 = 0.41109.$$

Since $0.41109 > 0$, the split is **beneficial**.

(C) Conclusion

We have Gain ≈ 0.41109, which is positive. XGBoost would therefore perform this split.

Solution to Question 2

We have 6 samples:

$$\begin{aligned}
(x_1, y_1, \hat{y}_1) &= (1.5,\ 0,\ 0.2), \\
(x_2, y_2, \hat{y}_2) &= (2.1,\ 1,\ -0.5), \\
(x_3, y_3, \hat{y}_3) &= (2.9,\ 0,\ 0.0), \\
(x_4, y_4, \hat{y}_4) &= (3.2,\ 0,\ 0.3), \\
(x_5, y_5, \hat{y}_5) &= (4.1,\ 1,\ 1.2), \\
(x_6, y_6, \hat{y}_6) &= (4.8,\ 1,\ 1.8).
\end{aligned}$$

Let $\lambda = 1,\ \gamma = 0.2$.

(1) Gradients and Hessians

Compute:

$$p_i = \sigma(\hat{y}_i), \quad g_i = p_i - y_i, \quad h_i = p_i(1 - p_i).$$

Approximations:

$$\begin{array}{lll}
p_1 = \sigma(0.2) \approx 0.549834, & g_1 = 0.549834 - 0 = 0.549834, & h_1 = 0.549834 \times 0.450166 \approx 0.247516, \\
p_2 = \sigma(-0.5) \approx 0.377541, & g_2 = 0.377541 - 1 = -0.622459, & h_2 = 0.377541 \times 0.622459 \approx 0.234216, \\
p_3 = \sigma(0.0) = 0.500000, & g_3 = 0.500000 - 0 = 0.500000, & h_3 = 0.500000 \times 0.500000 = 0.250000, \\
p_4 = \sigma(0.3) \approx 0.574443, & g_4 = 0.574443 - 0 = 0.574443, & h_4 = 0.574443 \times 0.425557 \approx 0.244458, \\
p_5 = \sigma(1.2) \approx 0.768525, & g_5 = 0.768525 - 1 = -0.231475, & h_5 = 0.768525 \times 0.231475 \approx 0.177893, \\
p_6 = \sigma(1.8) \approx 0.858149, & g_6 = 0.858149 - 1 = -0.141851, & h_6 = 0.858149 \times 0.141851 \approx 0.121732.
\end{array}$$

$$\sum g_i \approx 0.549834 + (-0.622459) + 0.500000 + 0.574443 + (-0.231475) + (-0.141851) = 0.628492,$$

$$\sum h_i \approx 0.247516 + 0.234216 + 0.250000 + 0.244458 + 0.177893 + 0.121732 = 1.275815.$$

(2) Split A vs. Split B

Split A: $x < 3$ vs. $x \geq 3$

$$L_A = \{1, 2, 3\}, \quad R_A = \{4, 5, 6\}.$$

$$\sum_{L_A} g_i = g_1 + g_2 + g_3 \approx 0.549834 + (-0.622459) + 0.500000 = 0.427375,$$

$$\sum_{L_A} h_i = h_1 + h_2 + h_3 \approx 0.247516 + 0.234216 + 0.250000 = 0.731732.$$

$$\sum_{R_A} g_i = g_4 + g_5 + g_6 \approx 0.574443 + (-0.231475) + (-0.141851) = 0.201117,$$

$$\sum_{R_A} h_i = h_4 + h_5 + h_6 \approx 0.244458 + 0.177893 + 0.121732 = 0.544083.$$

$$\sum g_{\text{all}} = 0.628492, \quad \sum h_{\text{all}} = 1.275815.$$

$$\text{Gain}_A = \frac{1}{2}\Big[\frac{(0.427375)^2}{0.731732+1} + \frac{(0.201117)^2}{0.544083+1} - \frac{(0.628492)^2}{1.275815+1}\Big] - 0.2.$$

$$(0.427375)^2 \approx 0.182648, \quad (0.201117)^2 \approx 0.040448, \quad (0.628492)^2 \approx 0.395029.$$

$$0.731732+1 = 1.731732, \quad 0.544083+1 = 1.544083, \quad 1.275815+1 = 2.275815.$$

$$\frac{0.182648}{1.731732} \approx 0.1055, \quad \frac{0.040448}{1.544083} \approx 0.0262, \quad \frac{0.395029}{2.275815} \approx 0.1736.$$

Inside brackets:

$$0.1055 + 0.0262 - 0.1736 = -0.0419.$$

Half of that is -0.02095. Then subtract $\gamma = 0.2$:

$$\text{Gain}_A = -0.02095 - 0.2 \approx -0.22095.$$

So $\text{Gain}_A \approx -0.221 < 0$. **Not beneficial.**

Split B: $x < 4$ vs. $x \geq 4$

$$L_B = \{1,2,3,4\}, \quad R_B = \{5,6\}.$$

$$\sum_{L_B} g_i = g_1 + g_2 + g_3 + g_4 \approx 1.002, \quad \sum_{L_B} h_i = h_1 + h_2 + h_3 + h_4 \approx 0.976190.$$

$$\sum_{R_B} g_i = g_5 + g_6 \approx -0.373326, \quad \sum_{R_B} h_i = h_5 + h_6 \approx 0.299625.$$

Totals: $\sum g_{\text{all}} \approx 0.628674$, $\sum h_{\text{all}} \approx 1.275815$.

$$\text{Gain}_B = \frac{1}{2}\Big[\frac{(1.002)^2}{0.976190+1} + \frac{(-0.373326)^2}{0.299625+1} - \frac{(0.628674)^2}{1.275815+1}\Big] - 0.2.$$

$$(1.002)^2 \approx 1.004, \quad (-0.373326)^2 \approx 0.13937, \quad (0.628674)^2 \approx 0.39524.$$

$$0.976190+1 = 1.97619, \quad 0.299625+1 = 1.299625, \quad 1.275815+1 = 2.275815.$$

$$\frac{1.004}{1.97619} \approx 0.508, \quad \frac{0.13937}{1.299625} \approx 0.1072, \quad \frac{0.39524}{2.275815} \approx 0.1737.$$

Inside the brackets:

$$0.508 + 0.1072 - 0.1737 = 0.4415.$$

Half of that is 0.22075. Subtract $\gamma = 0.2$:

$$\text{Gain}_B = 0.22075 - 0.2 = 0.02075.$$

$\text{Gain}_B \approx 0.021 > 0$. **Beneficial.**

(3) Which Split is Better?

$$\text{Gain}_A \approx -0.221, \quad \text{Gain}_B \approx 0.021.$$

Thus, Split B is chosen because it has positive (and higher) gain.

(4) Optimal Leaf Weights (for the Chosen Split B)

For a leaf with sum of gradients G and sum of hessians H:

$$w^* = -\frac{G}{H + \lambda}.$$

$$w_{L_B} = -\frac{\sum_{i \in L_B} g_i}{\sum_{i \in L_B} h_i + 1} = -\frac{1.002}{0.976190 + 1} \approx -\frac{1.002}{1.97619} \approx -0.507.$$

$$w_{R_B} = -\frac{\sum_{i \in R_B} g_i}{\sum_{i \in R_B} h_i + 1} = -\frac{-0.373326}{0.299625 + 1} = \frac{0.373326}{1.299625} \approx 0.287.$$

Conclusion for Question 2:

- Split A yields $\text{Gain}_A < 0$, so it is not used.
- Split B yields $\text{Gain}_B > 0$, so XGBoost picks Split B.
- Leaf weights after Split B are approximately:

$$w_{L_B} \approx -0.507, \quad w_{R_B} \approx 0.287.$$