Equivalence of Standardized Score Sum and F1 Score Optimization

Greg Jennings

December 2024

Equivalence of Standardized Score Sum and F1 Score Optimization

1 Setup and Definitions

Let's consider a binary classification problem with:

- Total dataset size: N
- Number of positive examples: P
- Number of negative examples: $N - P$

For standardized scores:

- Let $\mu = \frac{P}{N}$ be the mean (proportion of positive examples)
- Let $\sigma = \sqrt{\frac{P(N-P)}{N^2}}$ be the standard deviation

Therefore:

- Positive class gets score: $a = \frac{1-\mu}{\sigma}$
- Negative class gets score: $-b = \frac{0-\mu}{\sigma}$

2 For Any Fixed Window Size

For a window containing n points:

- Let TP be true positives
- $FP = n - TP$ will be false positives

2.1 Standardized Score Sum

The sum S of standardized scores in the window:

$$
\begin{aligned}
S &= TP \cdot a + FP \cdot (-b) \\
&= TP \cdot a + (n - TP) \cdot (-b) \\
&= TP(a + b) - nb
\end{aligned}
$$

2.2 F1 Score

F1 score is defined as:

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

where:

$$
\begin{aligned}
precision &= \frac{TP}{TP + FP} = \frac{TP}{n} \\
recall &= \frac{TP}{P}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
F1 &= 2 \cdot \frac{\frac{TP}{n} \cdot \frac{TP}{P}}{\frac{TP}{n} + \frac{TP}{P}} \\
&= 2 \cdot \frac{TP^2}{nP} \cdot \frac{nP}{TP(n + P)} \\
&= 2 \cdot \frac{TP}{n + P}
\end{aligned}
$$

3 Proof of Optimization Equivalence

1. For fixed window size n, the standardized score sum S is:
   - Linear in TP
   - Has slope $(a + b)$ which is positive (since $a > 0$ and $b < 0$)
   - Has constant term $-nb$
2. Similarly, F1 score for fixed n is:
   - Linear in TP
   - Has slope $\frac{2}{n+P}$ (positive)
   - Has zero constant term
3. Therefore, for any fixed window size n:

- Both metrics are linear functions of TP
- Both have positive slopes
- The constant terms don't affect optimization

4. As we slide our window:

   - For each window size n
   - Both metrics will achieve their maximum at the same value of TP
   - Therefore they will select the same optimal interval

4 Why This Matters

The equivalence shows that:

1. Maximizing the sum of standardized scores
2. Maximizing F1 score

Are functionally equivalent optimization problems when finding optimal intervals. This means we can use the computationally efficient dynamic programming approach to find optimal F1 score intervals.

5 Additional Note

The key insight is that standardization creates a scoring system where:

- Positive examples contribute positive scores
- Negative examples contribute negative scores
- The ratio of these scores is optimally balanced to align with F1 score optimization