Correctness of ArgMax Transformation

Alban Grastien

February 2026

This document is meant to provide a proof that the transformation of ArgMax as proposed in Caisar (or, at least, Caisar's `transform` branch as it stands at the time of writing) preserves its semantics.

We assume that the input is a tensor I of order 1 (i.e., an array) of k floats denoted $I = [I_1, \dots, I_k]$. Generalisation to tensors of higher order is assumed trivial and left for future work.

1 Semantics of ArgMax

$\texttt{ArgMax} : \mathbf{F}^k \to \mathbf{F}$ is an ONNX operator that takes I as input and returns the tensor $O = \texttt{ArgMax}(I)$ of order 0 (a scalar) such that $O \in \{1, \dots, k\}$ is the only index that satisfies the following constraint:

$$\forall j \in \{1, \dots, k\}. \left\{ \begin{array}{l} j < O \Rightarrow \quad I_O > I_j \\ j > O \Rightarrow \neg(I_j > I_O). \end{array} \right.$$

2 Transformation

We propose the following transformation: $\texttt{ArgMax}(I)$ is being replaced with $I \triangleright \texttt{diff} \triangleright \texttt{sign} \triangleright \texttt{ReLU} \triangleright \texttt{score} \triangleright \texttt{ReLU} \triangleright \texttt{raise}$ where $\triangleright$ is the reverse function applicator (i.e., $x \triangleright f$ is equivalent to $f(x)$) and the different operators are:

- $\texttt{diff}$ is the operator $\mathbf{F}^k \to \mathbf{F}^{C_k^2}$ such that $\texttt{diff}_{(i,j)}(I) = I_j - I_i$ for all $(i, j) \in C_k^2$ $(i < j)$;[1]
- $\texttt{sign}$ is the operator $\mathbf{F}^s \to \{-1, 0, 1\}^s$ where s is context-dependent such that

$$\texttt{sign}_i(V) = \left\{ \begin{array}{ll} 1 & \text{if } V_i > 0 \\ -1 & \text{if } V_i < 0 \\ 0 & \text{if } V_i = 0; \end{array} \right.$$

- $\texttt{ReLU}$ is the operator $\mathbf{F}^s \to \mathbf{F}^s$ (or $ints^s \to ints^s$) where s is context-dependent such that

$$\texttt{ReLU}_i(V) = \left\{ \begin{array}{ll} V_i & \text{if } V_i > 0 \\ 0 & \text{if } V_i \leq 0; \end{array} \right.$$

- $\texttt{score}$ is a function $\mathbf{F}^{C_k^2} \to \mathbf{F}^k$ such that $\texttt{score}_i(V) = (1 - i) + \Sigma_{j<i}\ V_{(j,i)} - \Sigma_{j>i}\ V_{(i,j)}$;
- $\texttt{raise}$ is the operator $\mathbf{F}^s \to \mathbf{F}$ where s is context-dependent such that

$$\texttt{raise}(V) = \Sigma_{i \in \{1,\dots,\text{degree}(V)\}}\ V_j \times j.$$

[1] The exact implementation of this operator is explained later.

3 Proof

We use the following notation to refer to the intermediate tensors during the computation.

- $I^d = I \triangleright \texttt{diff}$,
- $I^{ds} = I \triangleright \texttt{diff} \triangleright \texttt{sign}$,
- $I^{dsr} = I \triangleright \texttt{diff} \triangleright \texttt{sign} \triangleright \texttt{ReLU}$,
- $I^{dsrs} = I \triangleright \texttt{diff} \triangleright \texttt{sign} \triangleright \texttt{ReLU} \triangleright \texttt{score}$,
- $I^{dsrsr} = I \triangleright \texttt{diff} \triangleright \texttt{sign} \triangleright \texttt{ReLU} \triangleright \texttt{score} \triangleright \texttt{ReLU}$, and
- $I^{dsrsrr} = I \triangleright \texttt{diff} \triangleright \texttt{sign} \triangleright \texttt{ReLU} \triangleright \texttt{score} \triangleright \texttt{ReLU} \triangleright \texttt{raise}$.

Theorem 1 *Given the input I and $O = \texttt{ArgMax}(I)$, the following properties holds:*

1. $\{O\} = \Big\{j \in \{1, \ldots, k\} \mid \forall \ell \in \{1, \ldots, k\}. \Big(\ell < j \Rightarrow I^d_{(\ell,j)} > 0\Big) \wedge \Big(\ell > j \Rightarrow I^d_{(j,\ell)} \leq 0\Big)\Big\}$;
2. $I^{ds} \in \{-1, 0, 1\}^{C_k^2}$;
3. $\{O\} = \Big\{j \in \{1, \ldots, k\} \mid \forall \ell \in \{1, \ldots, k\}. \Big(\ell < j \Rightarrow I^{ds}_{(\ell,j)} = 1\Big) \wedge \Big(\ell > j \Rightarrow I^{ds}_{(j,\ell)} \leq 0\Big)\Big\}$;
4. $I^{dsr} \in \{0, 1\}^{C_k^2}$;
5. $\{O\} = \Big\{j \in \{1, \ldots, k\} \mid \forall \ell \in \{1, \ldots, k\}. \Big(\ell < j \Rightarrow I^{dsr}_{(\ell,j)} = 1\Big) \wedge \Big(\ell > j \Rightarrow I^{dsr}_{(j,\ell)} = 0\Big)\Big\}$;
6. $\{1, 0, -1, -2, \ldots\}^{C_k^2} \in I^{dsrs}$;
7. $\{O\} = \big\{j \in \{1, \ldots, k\} \mid I^{dsrs}_j = 1\big\}$;
8. $\{0, 1\}^{C_k^2} \in I^{dsrsr}$;
9. $\{O\} = \big\{j \in \{1, \ldots, k\} \mid I^{dsrsr}_j = 1\big\}$;
10. $O = I^{dsrsrr}$.

Proof of Property 1.1

$I^d_{(i,j)}$ is computed as the difference between I_j and I_i. Furthermore, O is the ArgMax of I, i.e., it is the (only) index that satisfies $I_O \geq I_i$ for all $i > O$ and $I_O > I_i$ for all $i < O$. This is precisely what Property 1.1 describes.

Proof of Property 1.2

$I^{ds} = I^d \triangleright \texttt{sign}$. Thus, the values in I^{ds} belong to $\{-1, 0, 1\}$.

Proof of Property 1.3

$I^{ds} = I^d \triangleright \texttt{sign}$. Property 1.3 is the one-to-one translation of Property 1.1.

Proof of Property 1.4

$I^{dsr} = I^{ds} \triangleright \texttt{ReLU}$. From Property 1.2, we know $I^{ds} \in \{-1, 0, 1\}^{C_k^2}$. Thus, it holds: $I^{dsr} \in \{0, 1\}^{C_k^2}$.

Proof of Property 1.5

$I^{dsr} = I^{ds} \triangleright$ `ReLU`. Property 1.5 is the one-to-one translation of Property 1.3.

Proof of Property 1.6 and Property 1.7

$I^{dsrs} = I^{dsrs} \triangleright$ `score`. First proving that $I_O^{dsrs} = 1$, and then that $I_\ell^{dsrs} \neq 1$ for $\ell \neq O$.

First, $I_O^{dsrs} = (1 - O) + \Sigma_{j<O}\ I_{(j,O)}^{dsrs} - \Sigma_{j>O}\ I_{(O,j)}^{dsrs} = (1 - O) + \Sigma_{j<O}\ 1 - \Sigma_{j>O}\ 0 = 1 - O + O + 0 = 1$.

Second, for another index ℓ, there exists an index j such that $V_{(j,\ell)} = 0$ or $V_{(\ell,j)} = 1$ (since O is the only index for which the values are as in the equation above). Thus, for every j that violates the property, the sum is reduced by 1 (and therefore, equals 0, or -1, etc.).

Proof of Property 1.8

$I^{dsrsr} = I^{dsrs} \triangleright$ `ReLU`. From Property 1.6, we know $I^{dsrs} \in \{1, 0, -1, -2, \ldots\}^{C_k^2}$. Thus, it holds: $I^{dsrsr} \in \{0, 1\}^{C_k^2}$.

Proof of Property 1.9

$I^{dsrsr} = I^{dsrs} \triangleright$ `ReLU`. From Properties 1.6 and 1.7, we know that the value associated with O is 1, so $I_O^{dsrsr}1$ Furthermore, we know that the value associated with any $\ell \neq O$ is in $\{1, 0, -1, \ldots\} \setminus \{1\}$, thus $I_\ell^{dsrsr} = 0$. Therefore, O is the only index such that $I_O^{dsrsr} = 1$.

Proof of Property 1.10

From Properties 1.8 and 1.9, we know that all values are 0 except for O which is associated with 1, i.e., $I^{dsrsrr} = \Sigma_{j\in\{1,\ldots,k\}}\ j \times I_j^{dsrsr} = O \times 1 = O$.

4 Notes

4.1 Implementation of the `diff` operator

The `diff` operator is implemented via a General Matrix Multiplication (`gemm`). As a reminder, the Gemm operation (with no scalar multiplier, no transpose, and no bias) is done between two matrices A and B of shape (M, L) and (L, N) respectively and returns a matrix $C = A \otimes B$ of shape (M, N) such that:

$$C_{i,j} = \sum_{\ell\in\{1,\ldots,L\}} A_{i,\ell} \times B_{\ell,j}.$$

Here, A is the input vector I of shape $(1, k)$ and B is a constant matrix of shape (k, C_k^2). For any pair (i, j) such that $1 \leq i < j \leq k$, let $\ell_{i,j}$ be an index from 1 to C_k^2 that is unique to this pair. Then, $B_{x,z}$ is:

- -1 if there exists y such that $\ell_{x,y} = z$;
- 1 if there exists y such that $\ell_{y,x} = z$;
- 0 otherwise.

It should be clear that the property $C_{\ell_{i,j}} = A_j - A_i$ holds for any pair (i, j), i.e., this is a correct implementation of `diff`.

Operators `score` and `raise` are similarly implemented through a `gemm` operator that includes a constant matrix (and a bias for `score`).