# Notation

The following notation is used throughout this work.

| | |
|---|---|
| $a$ | a scalar (real, integer, or word/token) |
| $\tilde{a}$ | a vector, nominally a column vector |
| $A$ | a matrix |
| $\mathcal{A}$ | a sequence, including a dataset or a sequence of words |
| $\tilde{x}_{[i]}$ | the $i$th element of the vector $\tilde{x}$ |
| $X_{[i,j]}$ | the row $i$ and column $j$ element $X$ |
| $X_{[:,i]}$ | the $i$th column *vector* of the matrix $X$ |
| $X_{[i,:]}$ | the $i$th row *vector* of the matrix $X$ |
| $w^{(t)}$ | a scalar $t$th element of some sequence |
| $W^{\mathrm{f}}$ | a matrix disambiguated by the name $f$ |
| $[A\ B]$ | the horizontal concatenation of $A$ and $B$ |
| $[A; B]$ | the vertical concatenation of $A$ and $B$ |
| $P(\ldots)$ | A probability (estimated or ground truth) |

**Words are treated as integers**

We consistently notate words, as if they were scalar integer values. writing for example $w^{(1)}$ as to be the first word in a sequence. Which is then used an an index: $C_{[:,w^{(i)}]}$ is it's corresponding word vector, from the embedding matrix $C$.

**Superscripts are never exponents**

Just to reiterate, we use $x^{(t)}$ not to represent $\prod_{i=1}^{i=t} x$, but as a variable name for a sequence element. The only exponential that occurs in this work is the natural exponential, which we write $\exp x$.
Readers may be more familiar with subscripts being used to allow more variable names e.g. $x_t$. However, we use this for indexing: that would be the $t$th element of $x$.
In the case of recurrent neural networks, we often will want to reference both the $i$ element of the $t$th timestep variable: $x^{(t)}$. This is thus unambiguously written written $x^{(t)}_{[i]}$.