| $$\begin{bmatrix}The\\brown\\fox\\jumps\end{bmatrix}\xrightarrow[tokenizer]{}\begin{bmatrix}101\\2078\\513\\778\end{bmatrix}\xrightarrow[layer]{embedding}\begin{bmatrix}\begin{bmatrix}\phantom{-}0.21&-1.03&\phantom{-}0.87&\phantom{-}0.58&-0.76\end{bmatrix}\\\begin{bmatrix}-0.45&\phantom{-}0.92&-0.23&\phantom{-}1.12&\phantom{-}0.09\end{bmatrix}\\\begin{bmatrix}\phantom{-}0.98&-0.14&\phantom{-}0.45&-0.88&\phantom{-}0.63\end{bmatrix}\\\begin{bmatrix}-0.67&\phantom{-}0.73&\phantom{-}0.36&\phantom{-}0.21&-0.47\end{bmatrix}\end{bmatrix}$$ |
| --- |
|  |
| $$\begin{bmatrix}pos=0\\pos=1\\pos=2\\pos=3\end{bmatrix}\xrightarrow[encoder]{positional}\begin{bmatrix} \begin{bmatrix} PE{(0,0)} & PE{(0,1)} & PE{(0,2)} & PE{(0,3)} & PE{(0,4)} \end{bmatrix} \\ \begin{bmatrix} PE{(1,0)} & PE{(1,1)} & PE{(1,2)} & PE{(1,3)} & PE{(1,4)} \end{bmatrix} \\ \begin{bmatrix} PE{(2,0)} & PE{(2,1)} & PE{(2,2)} & PE{(2,3)} & PE{(2,4)} \end{bmatrix} \\ \begin{bmatrix} PE{(3,0)} & PE{(3,1)} & PE{(3,2)} & PE{(3,3)} & PE{(3,4)} \end{bmatrix} \end{bmatrix}$$ |
| $$=\begin{bmatrix} \begin{bmatrix} \sin\left(\frac{0}{10000^{0/5}}\right) & \cos\left(\frac{0}{10000^{0/5}}\right) & \sin\left(\frac{0}{10000^{2/5}}\right) & \cos\left(\frac{0}{10000^{2/5}}\right) & \sin\left(\frac{0}{10000^{4/5}}\right) \end{bmatrix} \\ \begin{bmatrix} \sin\left(\frac{1}{10000^{0/5}}\right) & \cos\left(\frac{1}{10000^{0/5}}\right) & \sin\left(\frac{1}{10000^{2/5}}\right) & \cos\left(\frac{1}{10000^{2/5}}\right) & \sin\left(\frac{1}{10000^{4/5}}\right) \end{bmatrix} \\ \begin{bmatrix} \sin\left(\frac{2}{10000^{0/5}}\right) & \cos\left(\frac{2}{10000^{0/5}}\right) & \sin\left(\frac{2}{10000^{2/5}}\right) & \cos\left(\frac{2}{10000^{2/5}}\right) & \sin\left(\frac{2}{10000^{4/5}}\right) \end{bmatrix} \\ \begin{bmatrix} \sin\left(\frac{3}{10000^{0/5}}\right) & \cos\left(\frac{3}{10000^{0/5}}\right) & \sin\left(\frac{3}{10000^{2/5}}\right) & \cos\left(\frac{3}{10000^{2/5}}\right) & \sin\left(\frac{3}{10000^{4/5}}\right) \end{bmatrix} \end{bmatrix}$$ |
|  |
| $$=\begin{bmatrix}\begin{bmatrix}0.000&1.000&0.000&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.841&1.000&0.001&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.909&0.999&0.001&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.141&0.997&0.002&1.000&0.000\end{bmatrix}\end{bmatrix}(3\:d.p.)$$ |
| $$\begin{bmatrix}\begin{bmatrix}0.000&1.000&0.000&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.841&1.000&0.001&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.909&0.999&0.001&1.000&0.000\end{bmatrix}\\\begin{bmatrix}0.141&0.997&0.002&1.000&0.000\end{bmatrix}\end{bmatrix}+\begin{bmatrix}\begin{bmatrix}\phantom{-}0.21&-1.03&\phantom{-}0.87&\phantom{-}0.58&-0.76\end{bmatrix}\\\begin{bmatrix}-0.45&\phantom{-}0.92&-0.23&\phantom{-}1.12&\phantom{-}0.09\end{bmatrix}\\\begin{bmatrix}\phantom{-}0.98&-0.14&\phantom{-}0.45&-0.88&\phantom{-}0.63\end{bmatrix}\\\begin{bmatrix}-0.67&\phantom{-}0.73&\phantom{-}0.36&\phantom{-}0.21&-0.47\end{bmatrix}\end{bmatrix}$$  $$=\begin{bmatrix}\begin{bmatrix}\phantom{-}0.21&-0.03&\phantom{-}0.87&\phantom{-}1.58&-0.76\end{bmatrix}\\\begin{bmatrix}\phantom{-}0.39&\phantom{-}1.92&-0.23&\phantom{-}2.12&\phantom{-}0.09\end{bmatrix}\\\begin{bmatrix}\phantom{-}1.89&\phantom{-}0.86&\phantom{-}0.45&\phantom{-}0.12&\phantom{-}0.63\end{bmatrix}\\\begin{bmatrix}-0.53&\phantom{-}1.73&\phantom{-}0.36&\phantom{-}1.21&-0.47\end{bmatrix}\end{bmatrix}$$ |
| $$ \begin{bmatrix} \begin{bmatrix} \phantom{-}0.21 & -0.03 & \phantom{-}0.87 & \phantom{-}1.58 & -0.76 \end{bmatrix} \\ \begin{bmatrix} \phantom{-}0.39 & \phantom{-}1.92 & -0.23 & \phantom{-}2.12 & \phantom{-}0.09 \end{bmatrix} \\ \begin{bmatrix} \phantom{-}1.89 & \phantom{-}0.86 & \phantom{-}0.45 & \phantom{-}0.12 & \phantom{-}0.63 \end{bmatrix} \\ \begin{bmatrix} -0.53 & \phantom{-}1.73 & \phantom{-}0.36 & \phantom{-}1.21 & -0.47 \end{bmatrix} \end{bmatrix} \xrightarrow[\text{multi-head}]{\text{attention}} \begin{bmatrix} \begin{bmatrix} \phantom{-}0.18 & -0.02 & \phantom{-}0.68 & \phantom{-}1.37 & -0.66 \end{bmatrix} \\ \begin{bmatrix} \phantom{-}0.07 & \phantom{-}0.18 & -0.02 & \phantom{-}0.32 & \phantom{-}0.01 \end{bmatrix} \\ \begin{bmatrix} \phantom{-}0.19 & \phantom{-}0.05 & \phantom{-}0.03 & \phantom{-}0.12 & \phantom{-}0.04 \end{bmatrix} \\ \begin{bmatrix} -0.09 & \phantom{-}0.31 & \phantom{-}0.02 & \phantom{-}0.23 & -0.06 \end{bmatrix} \end{bmatrix} $$ |
| $$x\_{0}-\mu\_{0}=[x\_{0,0}-\mu\_0\quad x\_{0,1}-\mu\_0\quad x\_{0,2}-\mu\_0\quad x\_{0,3}-\mu\_0\quad x\_{0,4}-\mu\_0]$$  $$=[0.18-0.31\quad-0.02-0.31\quad 0.68-0.31\quad 1.37-0.31\quad-0.66-0.31]$$  $$=[-0.13\quad-0.33\quad \phantom{-}0.37\quad \phantom{-}1.06\quad-0.97]$$ |
| $$\frac{x\_{i}-\mu\_i}{\sqrt{\sigma\_i^2}}=\left[\frac{x\_{0,0}-\mu\_0}{\sqrt{\sigma\_0^2}},\quad\frac{x\_{0,1}-\mu\_0}{\sqrt{\sigma\_0^2}},\quad\frac{x\_{0,2}-\mu\_0}{\sqrt{\sigma\_0^2}},\quad\frac{x\_{0,3}-\mu\_0}{\sqrt{\sigma\_0^2}},\quad\frac{x\_{0,4}-\mu\_0}{\sqrt{\sigma\_0^2}}\right]$$  $$=\left[\frac{-0.13}{0.7631}\quad\frac{-0.33}{0.7631}\quad\frac{0.37}{0.7631}\quad\frac{1.06}{0.7631}\quad\frac{-0.97}{0.7631}\right]$$  $$=\begin{bmatrix}-0.17&-0.43&\phantom{-}0.48&\phantom{-}1.39&-1.27\end{bmatrix}(3\:d.p.)$$ |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |