

Diagram illustrating the multiplication of a 2x4 matrix X (green) by a 4x3 matrix W^Q (red) to produce a 2x3 matrix Q (red).

The diagram shows a green 2x4 matrix labeled X multiplied by a blue 4x3 matrix labeled W^K , resulting in a blue 2x3 matrix labeled K . The matrices are represented as grids of colored squares: green for X , blue for W^K and K . The multiplication is indicated by a large \times symbol, and the result is indicated by an equals sign $=$.

The diagram shows a green 2x4 matrix labeled X multiplied by a blue 4x3 matrix labeled W^v , resulting in a blue 2x3 matrix labeled V . The matrices are represented by grids of colored squares (green for X , blue for W^v and V), and the operation is indicated by a multiplication symbol (\times) and an equals sign ($=$).

$$\text{Softmax} \left(\frac{\begin{matrix} \textcolor{red}{Q} & & \textcolor{blue}{K}^T \\ \begin{matrix} \square & \square & \square \\ \square & \square & \square \end{matrix} & \times & \begin{matrix} \square & \square \\ \square & \square \\ \square & \square \end{matrix} \end{matrix}}{\sqrt{d_k}} \right) \begin{matrix} \textcolor{blue}{V} \\ \begin{matrix} \square & \square & \square \\ \square & \square & \square \end{matrix} \end{matrix} = \begin{matrix} \textcolor{red}{Z} \\ \begin{matrix} \square & \square & \square \\ \square & \square & \square \end{matrix} \end{matrix}$$