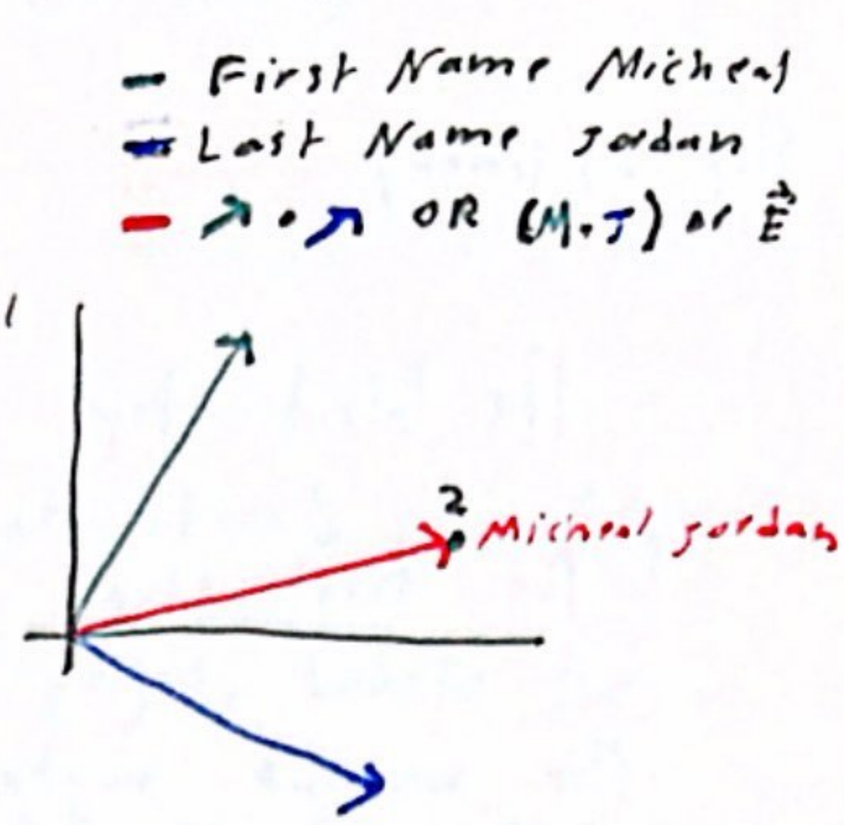
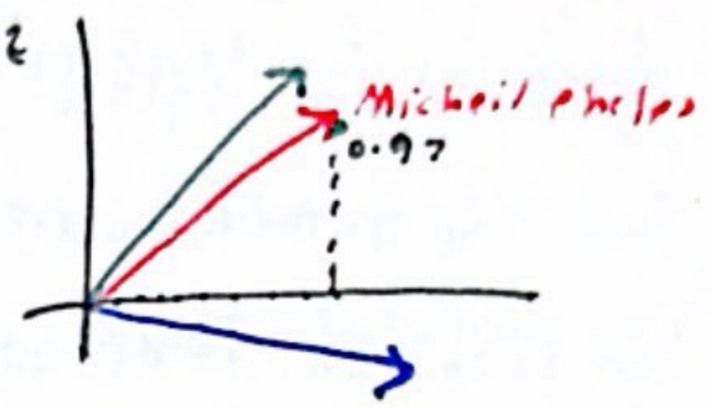
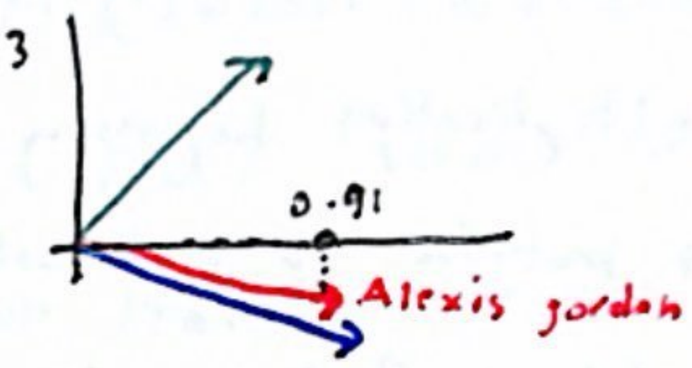
MLP (P-5) (LLM Notes)

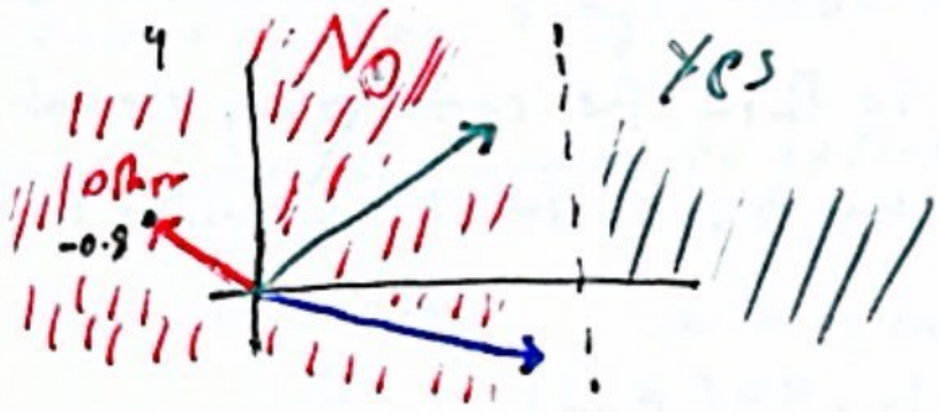
step 2 ReLu 



- A problem from the last page is that language is a non linear process while our step 1 from last page is a linear process, if the entry is high for Michael + jordan (fig 1) then it would also be triggered somewhat by Michael + phelps (fig 2) and also Alexes + jordan (fig 3) dispite those things being unrelated. what you really want is a simple yes or no for the full name. (fig 4).







- So the next step would be to pass this large intermediate vector if the result of the linear step $(WE_i + B)$ through a simple non linear function like ReLu which takes all neg values and sets them to 0 else leaves them as is. From the 3 figure on last page now the ≈ 1 it encodes "Michael jordan" and ≈ 0 otherwise is true. often models use a Gelu function (smoother)

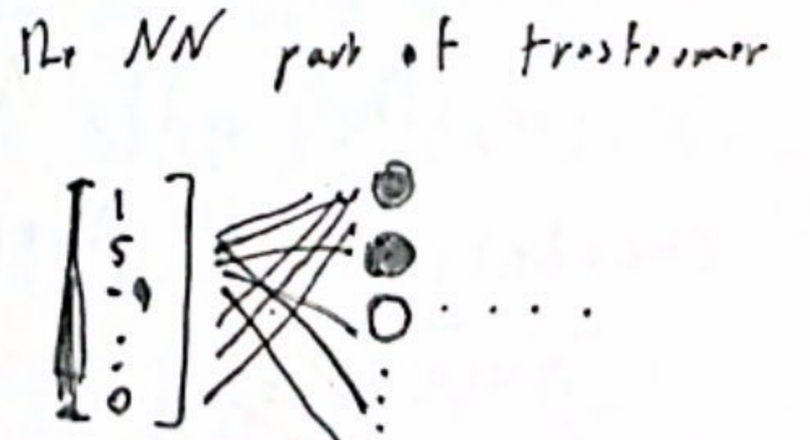
So now it looks like this this!

Neurons of a transformer

$\Rightarrow$ the NN part of transformer

$$\begin{bmatrix} 36\,8\cdots \\ 7\cdot \\ \vdots \\ \cdot \end{bmatrix} \begin{bmatrix} 5 \\ 7 \\ 3 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ -5 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -9 \\ \vdots \\ 0 \end{bmatrix} \rightarrow \text{Relu} \rightarrow \begin{bmatrix} 1 \\ 5 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

W $\quad$ $\vec{E}_{\text{Jordan}}$ $\quad$ B $\quad$ Result of linear step



Scanned with CamScanner