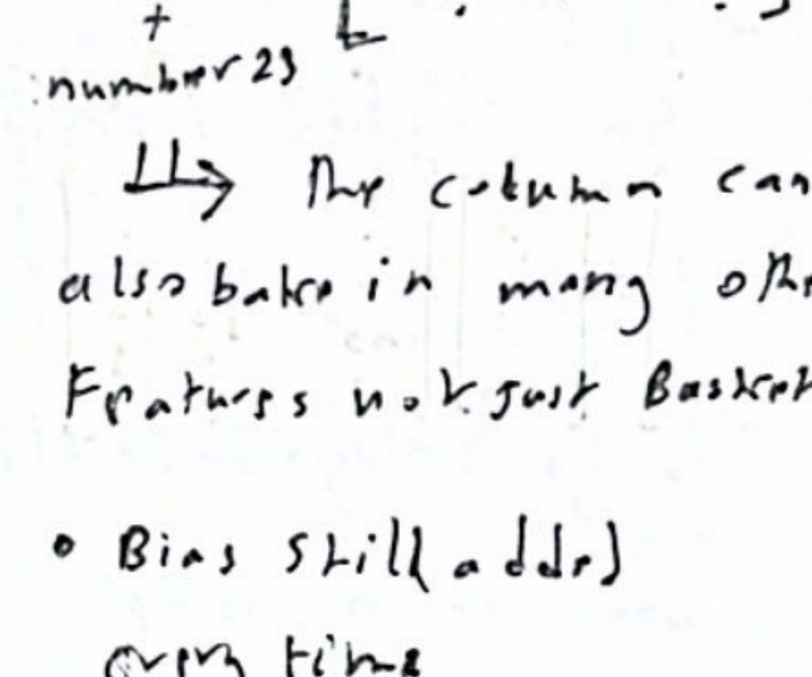
# MLP (P-6) (LLM notes)

Step 3 (linear)

- The third step (also linear) is almost identical to the first step $W\vec{E} + B \rightarrow$ relu $\rightarrow$ result... except

the weight matrix is projecting down as the first one projected up so that our result is now back down to the size of that embedding space (12,288) and so its easy to think of the multiplications happening column by column in the matrix so each column is asking a question

if Column 2 is Basketball direction Vector so then the corresponding element Eo would be 0 or 2 for if it relates to basketball or not



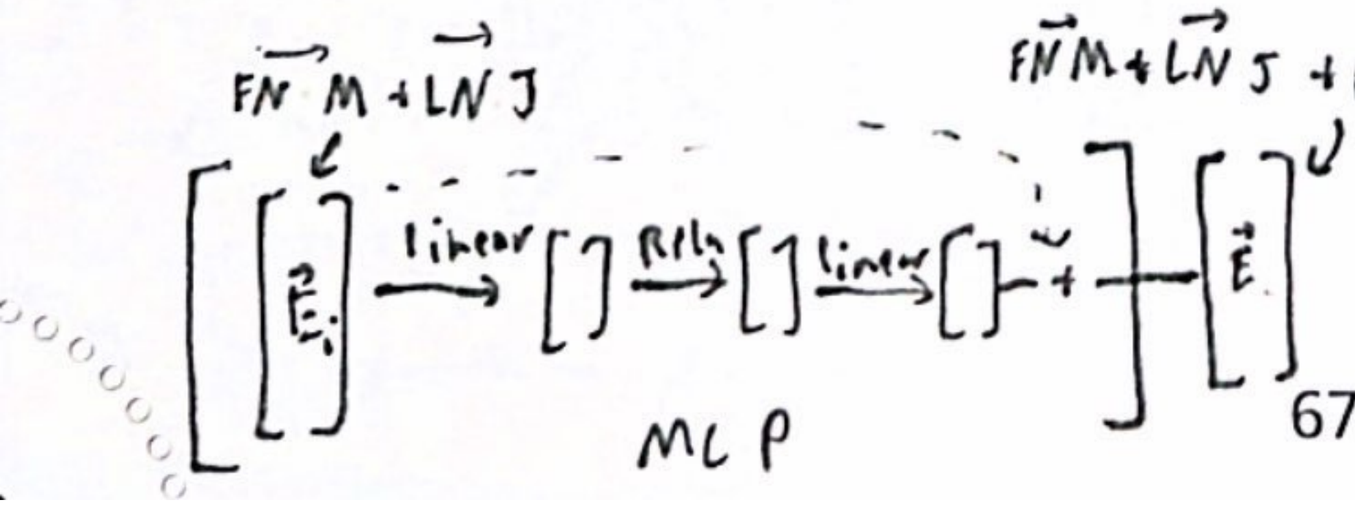

↳ The column can also bake in many other features not just Basketball

- Bias still added every time

Step 4 (Add)

- This linear result is then added back to the original vector (now of same dim) and that gives you the final result (see p 64)

- So what all this does is if input vector encoded Michael + Jordan after the MLP the output will be a new updated vector that adds on a basketball direction so this vector encodes all of those (MJ + Ball) together for more context and facts




- NOTE this process happens for all input tokens in parallel and in short adds information to vectors in a high dim embedding space