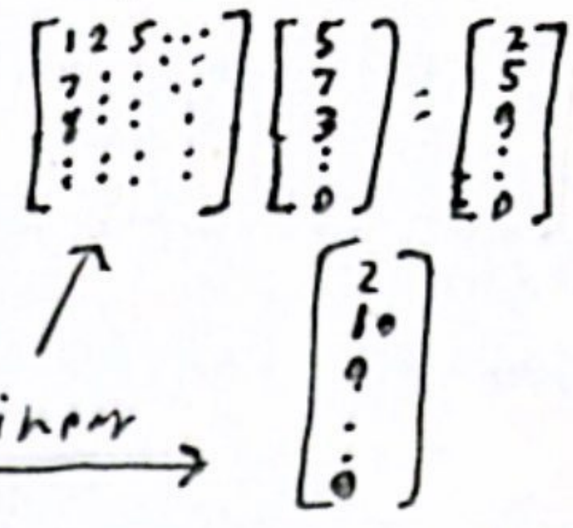
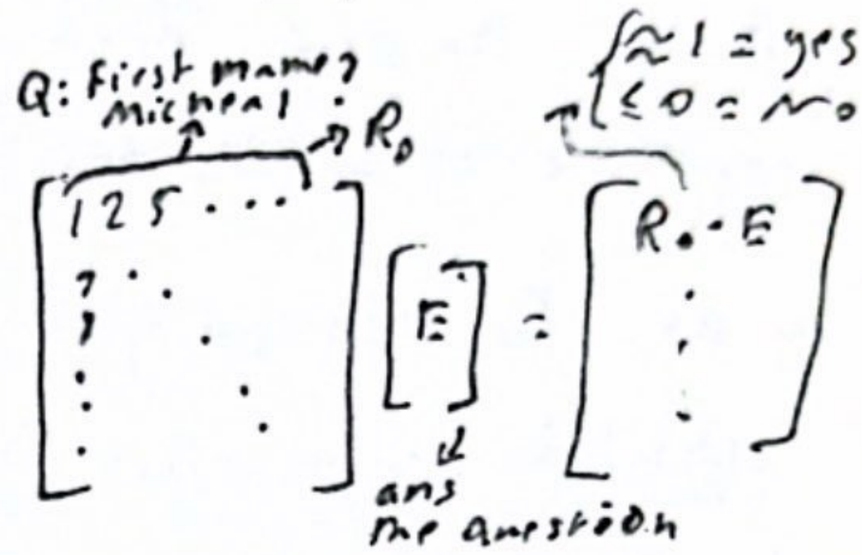
MLP (P-4) (LLM notes) 2 projects up

Step 2 (linear 2)

- The first step is multiplying [5 7 3 ... 0] by a matrix filled with model parameters (learned in training) since the multiplications are multiplying each row in the matrix by the whole vector so think of

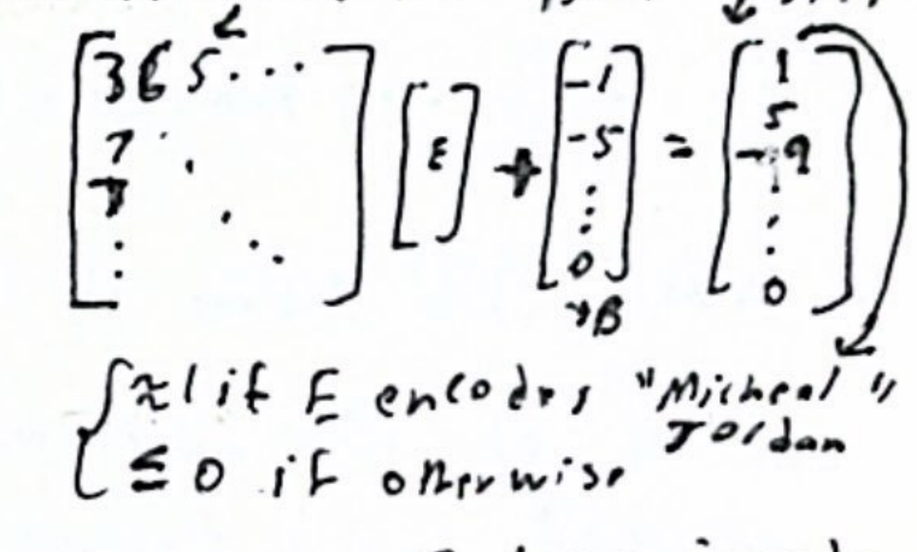


- each row as asking a question
- and this dot product multiplication it's
- ≈1 if the vector answers that question with a yes and ≤0 if not and the



Questions can be more complicated like asking 2 questions in 1 case for both Q's to be yes the R0 · E ≈ 2 = yes and ≤ 1 = NO

- Very often this step requires adding another vector to the output called the Bias which also has model parameters this Bias normalizes the dot product so in the case of R0 · E ≈ 2 for yes the bias would learn to have -1 as first element so that the result is ≈ 2 = yes ≤0 = NO



- The number of Rows in the matrix ≈ num of question asked is 50K in GPT-3 which is 4x the dimention space of embedding # choosing a clean multiple for the dimention is hardware friendly

☆ You can think of R0 in the last Ex as having vectors for M and J hence the dot product is largest if E also has directionality in the M and J direction (which it does) so think of the operation R0 · E = (M+J) · E65 once again Matrix has billions