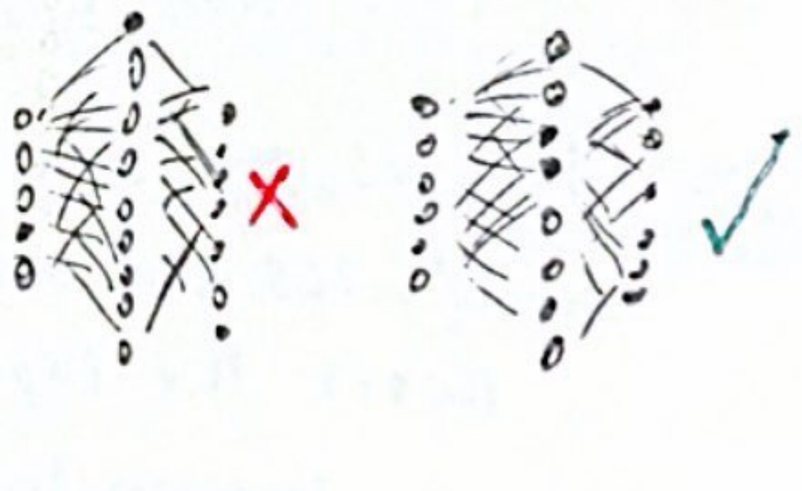
MLP (P-7) (Llm notes)

- Super position + Recap

- in real LLM facts are not stored exactly as described in the past few pages, while the rows of the Matrix can be thought as directions in embedding space, that helps get facts for context. and a neuron tells you how much a given vector aligns with some direction. and the columns of the second matrix tell you what will be added to the result if that neuron is active. But individual neurons rarely represent the single clean feature in a single row of M, does not always ask a (from single Direction like it does in our ex. rather multiple neurons come together to store a fact it ans a question. If it might take many rows to ans our Question it store some fact (Fig1)



- this happens theoretically by superposition in short it allows to store ideas > dimentions ie we had a basket ball direction but that 1 dimention can store other closely related idea Like a sport will balls for ex. hence this can explain why LLM scale so well but are still hard to interprit. and the features in the NN would have to come together to form facts not just one lit neuron (fig 2) that specific combination of neurons is a superposition.