- ## Question 1

What we expect here is that the optimal parameters would be weight of 1 for the current position and 0 elsewhere and a bias of 0 so that the output of the first layer is the vector x, the sum does its job and then we output the result of the sum directly.

- ## Question 2

Yes, because the first MLP (fc1) can output something different depending on the value of x1 and x2: sigma(x) can be different than sigma(x+1).