① output layer should n't have any activation function, because the answer is a continuous digit and because it is a regression

② only one, because we need to have 1 output.

③ The MSE should n't be fixed by 2 an on it.

④ ~~(crossed out)~~

⑥ a) The ~~Pro~~ number of nurons going up to 2 or 3 would be fine ~~(crossed out)~~ but ~~after~~ we added some more nurons model gone complex and overfitted.

⑦ activation function will change the way of learning and make the regression curves smoother if It is sigmoid is better but Relu will give us more tolerate in the ~~(crossed out)~~ answer.