

Again, the ideas are the same, although our explanation of the different layers could have been more elaborate. We did not mention how to generalize our method, but it should be pretty obvious from the diagram we drew.

Our reasoning for why the classifier would fail in practice was perhaps a bit misguided (the most important issue is the lack of generalization, the training could possibly be achieved by some other means).

Exercise 2

We want to prove that we can merge two layers into one by multiplying out the activation:

$$\begin{aligned}y &= \varphi(\varphi(X\beta_1 + b_1)\beta_2 + b_2) \\&= (X\beta_1 + b_1)\beta_2 + b_2 && \text{identities} \\&= X\beta_1\beta_2 + b_1\beta_2 + b_2 && \text{multiply out} \\&= X\underbrace{(\beta_1\beta_2)}_{\beta} + \underbrace{(b_1\beta_2 + b_2)}_b\end{aligned}$$

As before, our idea is essentially right, but our explanation could be more verbose.

Exercise 3

In its own HTML file.