

1. Questions

1.1 (2 points) **Tan, Chapter 4 Exercise 14, 15.**

14. For each of the Boolean functions given below, state whether the problem is linearly separable.

a. $A \text{ AND } B \text{ AND } C$

It is linearly separable because it contains the AND operation.

b. $\text{NOT } A \text{ AND } B$

It is linearly separable because it contains the AND and NOT operations.

c. $(A \text{ OR } B) \text{ AND } (A \text{ OR } C)$

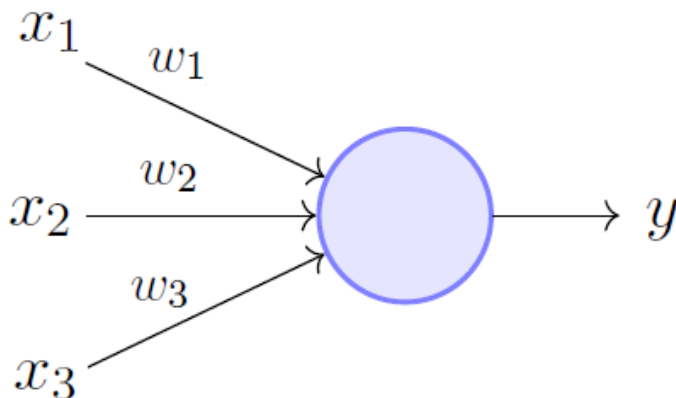
It is linearly separable because it contains the OR and AND operations.

d. $(A \text{ XOR } B) \text{ AND } (A \text{ OR } B)$

It is not linearly separable because it contains the XOR operation.

15.

a. Demonstrate how the perceptron model can be used to represent the AND and OR functions between a pair of Boolean variables.



Looking at this perceptron diagram above, let's demonstrate with only using x_1 and x_2 for the sake of keeping this example simple (I could not find a diagram with only two input arrows online).

With two Boolean variables, there are 4 possibilities of the truth values of the pair, which I denoted in the table below. For each pair, I wrote out the truth value of the output based on the two inputs and which operation it's using (AND or OR).

OPERATION: AND		
x1	x2	Output
0	0	0
0	1	0
1	0	0
1	1	1

And so, using these values with the perceptron model, it would be the following equation:

$$x_1 + x_2 - 1$$

OPERATION: OR		
x1	x2	Output
0	0	0
0	1	1
1	0	1
1	1	1

And so, using these values with the perceptron model, it would be the following equation:

$$2 * x_1 + 2 * x_2 - 1$$

- b. Comment on the disadvantage of using linear functions as activation functions for multi-layer neural networks.

One disadvantage of using linear functions is that in a multi-layer neural network with linear functions as activation functions, the model doesn't improve the error term throughout, making it redundant as a neural system. Also, we cannot use backpropagation as it has no way to refer to the information from X, and therefore, there is no way to provide a better expectation with each iteration. Another limitation is that with linear functions as the activation functions, all the layers combine into one essentially because the last layer would have a linear capacity.