Chapter 10: Theoretical Questions

Show that the MLP in Figure 10.6 of the book solves the XOR problem

zie notities

(Exercise 5 from the book) Name and draw three popular activation functions

```
sigmoid activation (s shape : 0 to 1)

tanh (hyperbolic tangent) activation (s shape : -1 to 1)

relu (rectified linear unit) activation (horizontal line: 0, linear increase to
infinity)
```

(Exercise 6 from the book) Suppose you have an MLP composed of

- input layer : 10 passthrough neurons
- one hidden layer : 50 artificial neurons
- one output layer : 3 artificial neurons.
- All artificial neurons use the ReLU activation function.

```
- What is the shape of the input matrix X?
    - (n_instances, 10)

- What are the shapes of the hidden layer’s weight matrix Wh and bias vector bh?
    - Wh: (10, 50)
    - bh: (50)

- What are the shapes of the output layer’s weight matrix Wo and bias vector bo?
    - Wo: (50, 3)
    - bo: (3)

- What is the shape of the network’s output matrix Y?
    - (n_instances, 3)

- Write the equation that computes the network’s output matrix Y as a function of
X, Wh, bh, Wo and bo.
    - Y = ReLu((ReLu(XWh + bh)) x Wo + bo)
```

(Exercise 7 from the book) How many neurons do you need in the output layer if you want to classify email into spam or ham

```
- Amount of neurons:
    - 1

- What activation function should you use in the output layer:
    - sigmoid activation
```

If instead you want to tackle MNIST:

```
- how many neurons do you need in the output layer:
    - 10

- which activation function should you use:
    - softmax activation
```

What about for getting your network to predict housing prices, as in Chapter 2:

```
- how many neurons do you need in the output layer:
    - 1

- which activation function should you use:
    - none or linear
```

Explain why activation functions are necessary in neural networks

```
- Linear combination of inputs = linear output => final output = linear => whole
model can be reduced to single layer => no complexity / can't solve complex
problems

Without activation functions, every layer in the network would output a linear
combination of the inputs, so the final output would also be a linear combination
of the inputs. In other words, the whole network could be reduced to a single
layer. If we want to solve complex problems, this is not sufficient; we need one
or more layers of non-linear neurons between the input and the output layers.
```

Suppose the logits are (−1, 0, 2) for a classification task with three classes

What are the probabilities for each class if we use the softmax activation function?

z	e^z	e^z / sum
-1	e^-1	e^-1 / sum
0	e^0	e^0 / sum
2	e^2	e^2 / sum

z	e^z	e^z / sum
sum	e^-1 + e^0 + e^2	1

Describe how to construct a neural network that is equivalent to a logistic regression model

```
-How many layers does it need?
    - 1

-What activation function should you use in the output layer?
    - sigmoid

-What loss function should you use?
    - binary cross entropy
```

Consider a TLU with weights w1 = 2, w2 = −1 and bias b = 1/2

```
-Is the example (x1, x2) = (1, 1) classified as positive or negative?
    - positive

-What about the example (x1, x2) = (−1, 0)?
    - negative

-In the (x1, x2)-plan, sketch the decision boundary of this TLU and indicate which
side is the positive class
    <=> w1x1 + w2x2 + b = 0
    <=> 2x1 - x2 + 1/2 = 0

    if x1 = 0 => x2 = 1/2 => (0, 1/2)
    if x2 = 0 => x1 = -1/4 => (-1/4, 0)

    => (0, 1/2) and (-1/4, 0) are on the decision boundary
    => positive class is above the line, negative class is below the line
```