50.039 – Theory and Practice of Deep learning

Alex

Week 04

[The following notes are compiled from various sources such as textbooks, lecture materials, Web resources and are shared for academic purposes only, intended for use by students registered for a specific course. In the interest of brevity, every source is not cited. The compiler of these notes gratefully acknowledges all such sources.]

Due: week5 Wednesday, 9pm

1 Task 1:

1. Given the sigmoid function

$$\varphi(x) = \frac{1}{1 + exp(-ax)}$$

which is limited between 0 and 1.

1. Prove that

$$\varphi'(x) = \frac{d\varphi}{dx} = a\varphi(x)[1 - \varphi(x)]$$

2. What is the value of $\varphi'(x)$ when $x = 0$

2 Task 2:

Consider a neural networks as shown in the Figure 1.
It just has one hidden layer and one output layer. Each hidden unit hsa a linear activation function $h(z) = cz$ with slope c. The sigmoid function $g(z) = \frac{1}{1+exp(-z)}$ is used as activation function at the output layer.

1. What is the output of the neural network? Write down the expression as a function of x_i, c and weights w_i

2. Consider a neural net for a binary classification. If the network would be used for binary classification with two classes, such that the prediction switches when the output of g is equal to 0.5, what is the equation for the

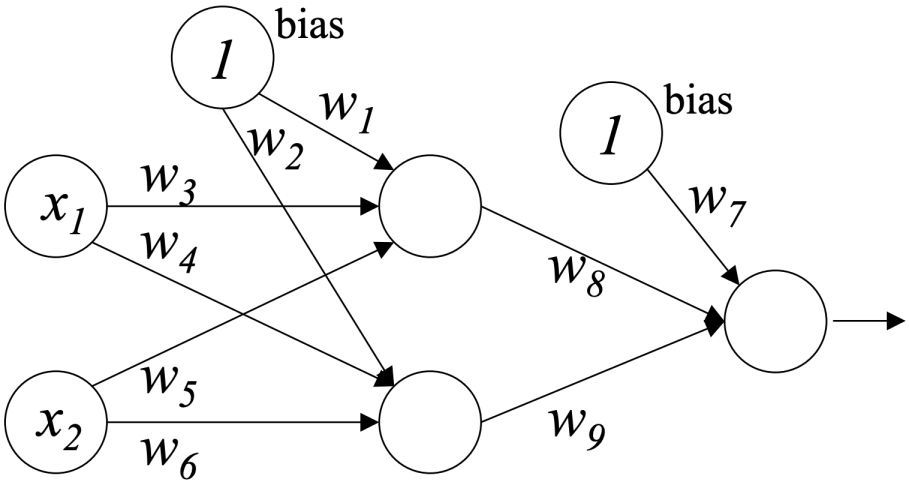

Figure 1:

decision boundary?

Write this equation in the form $h(x_1, x_2) = 0$. Hint: $g(\cdot)$ should not be part of that equation!

3. Draw a new neural network with no hidden layer which is equivalent to the given neural network.

4. what are the weights associated with the edges of the new neural network ? (Hint: write the new weights as a function of c and w_i)

3 Task 3:

Consider the following CNN:

- input size: 300×300, 3 channels
- layer 1: 30 channels, kernel size $7x7$, stride 2, no padding, then a relu
- layer 2: maxpool, kernel size $3x3$, stride 2, no padding,
- layer 3: 50 channels, kernel size $3x3$, stride 1, padding 3, then a relu
- many layers after that

1. what is the feature map size of layer 1,2,3 ? it should be something like (h, w, c)

note show your work

(FLOP). FLOP can be used to measure a computer's performance. A decent processor nowadays can perform in Giga-FLOPS, that means billions of FLOP per second.