

# 1

In a classical algorithm the instructions are specified by the programmer before run time. In a machine learning frame work, the programmer provides data the machine learning algorithm uses to train and determine most of the parameters. Based on the learned parameters, the output for new input is calculated.

# 2

# 3

# 4

## 4.a

Consider a single perceptron. Let  $\sigma$  be the activation function of the perceptron i.e.  $\sigma(x) = \mathbb{1}(x > 0)$ . Let  $w$  denote the weights and  $b$  the bias. Then the output of the perceptron for an input  $x$  is  $\sigma(wx + b)$ . Rescaling the weights and bias by  $c > 0$  is

$$\sigma(cwx + cb) = \sigma(c(wx + b)) = \mathbb{1}(c(wx + b) > 0) = \mathbb{1}(wx + b > 0) = \sigma(wx + b).$$

We used  $c > 0$ . Since this holds true for every perceptron in a perceptron network, rescaling does not change the behaviour.

## 4.b

The sigmoid function is

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

Then

$$\sigma(c(wx + b)) = \frac{1}{1 + e^{-c(wx+b)}} = \frac{1}{1 + (e^{-(wx+b)})^c}.$$

We see that for  $wx + b \neq 0$  we have  $\lim_{c \rightarrow \infty} \sigma(c(wx + b)) = \mathbb{1}(wx + b > 0)$ , which is exactly the behavior of a perceptron. For  $wx + b = 0$  we have  $\sigma(c(wx + b)) = 0.5$  for all  $c$ .

## 4.c

$$W_1 = ()$$