# Question 1

Consider the following block diagram of an artificial neuron, called the perceptron:

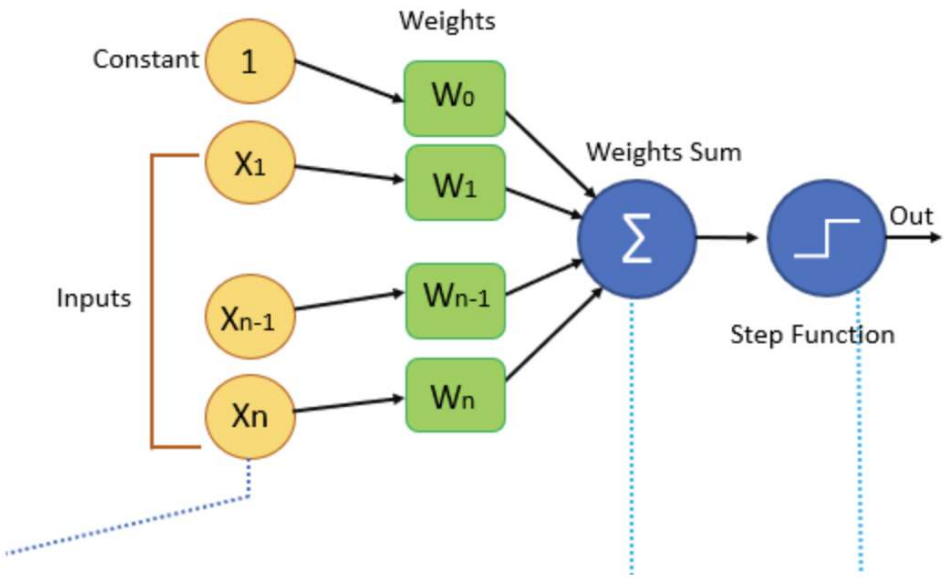


***Image:*** *Perceptron block diagram. *

The input $[x_1, x_2, x_3 \ldots . x_n]$ represents values of features of the model we want to train the neural network, where $n$ is the total number of features and $x_i$ is the value of the $i^{th}$ feature.

The weights $[w_0, w_1, w_2 \ldots . w_n]$ are learnt during the training of the model, with their values being updated after each training error is calculated.

Each input value $x_i$ will be first multiplied with the weight assigned to it $w_i$ and the sum of all the multiplied values is known as a weighted sum, denoted by $z$.

The activation function is applied on the weighted sum to converted in a numerical value, that is the output of the network. Consider as activation function the sigmoid function given by:

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

1) Write the expression that calculates $z$ in a matrix form and explain it.

2) Calculate the derivative of the output

$$o = \sigma(\sum_{i=0}^{n} x_i w_i)$$

with respect to the weight parameter $w_1$ considering all the rest as constants.

Note that the derivative of the sigmoid is given by

$$\sigma(x)' = \sigma(x)(1 - \sigma(x))$$

## Question 2

In a binary classification problem the neural network outputs are either 1 or 0. Suppose we are given a training dataset, $D = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\}$ of $N$ data points with $y_i \in \{0, 1\}$. We can then write the predicted value of the network as

$$\hat{y}_i = f(x_i|\theta)$$

where $\theta$ are the neural network parameters (weights and biases). We can then express the conditional probability $P(x_i \rightarrow y_i|\theta)$ in terms of a Bernoulli distribution as

$$P(x_i \rightarrow y_i|\theta) = \hat{y}_i^{y_i}(1 - \hat{y}_i)^{(1-y_i)}$$

1) Assuming that the observations of the training dataset are independent and each one follows the Bernoulli distribution, express the likelihood probability $P(x_1 \rightarrow y_1, \ldots, x_N \rightarrow y_N|\theta)$ of observing the data point $(x_i, y_i), i = 1, \ldots, N$ given the neural network parameters $\theta$.

2) In numerical computations instead of the likelihood probability we use its logarithm that is called the *log-likelihood*. Express the log-likelihood probability calculated in (1) and explain why we use the logarithm.

## Question 3

1) Which of the following is the correct set-builder notation for the set listed below?

$$\{-4, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

a) $\{x \in \mathbb{N} : -4 \leq x\}$

b) $\{x \in \mathbb{Z} : -4 \leq x\}$

c) $\{x \in \mathbb{R} : -4 \leq x\}$

d) All the above

2) If A = {Greek articles} and B = {Greek vowels}, then which is their intersection set $A \cap B$?

3) If A = {words in Document A}, and B = {words in Document B} and $A \subset B$, then:

a. Describe the B-A set.

b. Compare the cardinalities of sets A and B.

c. Could you make a safe conclusion about the lengths in terms of words of these two documents ?

## Question 4

Given the following observations of a random variable: 2, 4, 15, 8, 7, 7, 6, 3, 3, 3, 40

a) Calculate the mean and median values.

b) Which metric is more representative of the data, and why?

## Question 5

Given a collection of documents propose a statistical measure to estimate the importance of a word in a document.