# Understanding Long Short-Term Memory Architecture

Compiled by D. Gueorguiev 2/20/2024

## Notation

– learning rate of a network

– time unit

- initial time of an epoch

– final time of an epoch

– the set of units in the network

– (generic) units in the network;

- the set of input units, – input unit;

- the set of output units, – output unit;

– the set of non-input units, – generic or non-input unit;

- the output of unit ; the output of a unit is scalar

– the set of units with connections to a unit ; i.e. its predecessors

– the set of units with connections from a unit ; i.e. its successors

- the weight that connects the unit to unit

- the input from a unit coming from a unit

– the weighted input of the unit

- the bias of the unit

- the state of the unit

– the squashing function of unit

-the error of unit

-the error signal of unit

– the output sensitivity of the unit with respect to the weight

## Perceptron and the Delta Learning Rule

**Definition** Perceptron

Given the input vector and trained weights , the perceptron outputs , which is computed as

We refer to as the *weighted input* and to as the perceptron’s *state*. For perceptron to fire, its state must exceed the value of the threshold.

In cases of misclassification the Perceptron modifies weights accordingly. The perceptron will converge to reproduce the correct behavior provided that the training examples are linearly separable. Convergence is not assured if the training data is not linearly separable (Note: the proof is in Marvin Minsky’s book on Perceptrons).

### The Delta Learning Rule

## Literature

[Understanding Long Short-Term Memory Recurrent Neural Networks – a tutorial-like introduction, Ralf C. Staudemeyer, Eric R. Morris, 2019](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/TutorialOnLongShortTermMemory2019.pdf)