**IT159: Artificial Intelligence**

**Lab#8&9: Perceptron**

# Instructions

Read all the instructions below carefully before you start working on the assignment.

* Please submit the source code and a pdf report of your work.
* You must implement a perceptron algorithm from scratch.
* Code that does not work will not be considered.
* Please create a folder called "yourname**\_**StudentID\_Lab89" that includes all the required files and generate a zip file called "yourname\_ StudentID\_Lab89.zip".
* Please submit your work (.zip) to Blackboard.

# Practical assignment objective

This assignment is aimed at coding a perceptron from scratch in order to learn how this simple but powerful linear binary classifier works. In the following sections you can find a brief summary on what a perceptron is and how it works, and finally the text of the assignment.

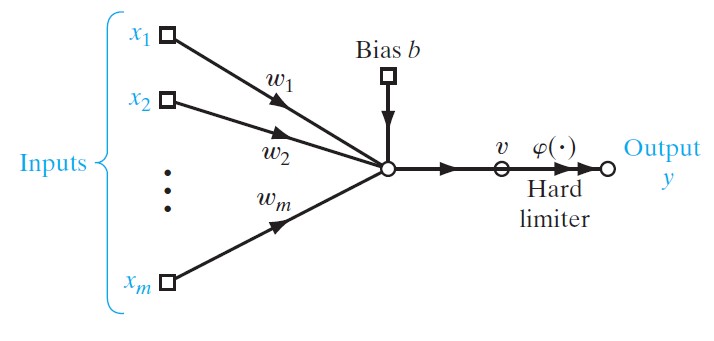
# Reminders about perceptron[[1]](#footnote-1)

## Historical introduction

The perceptron occupies a special place in the historical development of neural networks: It was the first algorithmically described neural network. Its invention by Rosenblatt, a psychologist, inspired engineers, physicists, and mathematicians alike to devote their research effort to different aspects of neural networks in the 1960s and the 1970s. The perceptron is the simplest form of a neural network used for the classification of data said to be linearly separable (i.e., data that lie on opposite sides of a hyperplane). Basically, it consists of a single neuron with adjustable synaptic weights and bias. The algorithm used to adjust the free parameters of this neural network first appeared in a learning procedure developed by Rosenblatt for his perceptron brain model. Indeed, Rosenblatt proved that if the data used to train the perceptron are drawn from two linearly separable classes, then the perceptron algorithm converges and positions the decision surface in the form of a hyperplane between the two classes. A perceptron is limited to performing pattern classification with only two classes. By expanding the output (computation) layer of the perceptron to include more than one neuron, we may correspondingly perform classification with more than two classes.

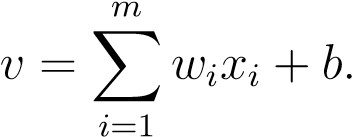
## Working with a perceptron

A perceptron consists of a linear combiner followed by a hard limiter (e.g. the sign function, the Heaviside step function or a simple function like these) (see Figure 1).



*Figure 1: Perceptron’s structure.*

The summing node of the neural model computes a linear combination of the inputs applied to its synapses, as well as incorporates an externally applied bias. The resulting sum is applied to a hard limiter. Accordingly, for example if the hard limiter is the Heaviside step function, the neuron produces an output equal to 1 if the hard limiter input is positive, and 0 if it is negative. The synaptic weights of the perceptron are denoted by *w*1, *w*2, ...,*wm*. Correspondingly, the inputs applied to the perceptron are denoted by *x*1, *x*2, ..., *xm*. The externally applied bias is denoted by *b*. From the model, we find that the hard limiter input of the neuron is:



Here is the **Python code** that models a **perceptron** as described in your paragraph. It includes:

* A **linear combiner** of inputs and weights
* An **externally applied bias**
* A **hard limiter** (Heaviside step function)

✅ **Python Code for Perceptron with Heaviside Step Function**

A screenshot of a computer program

AI-generated content may be incorrect.

🧪 Example Usage:

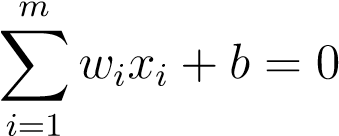
A screenshot of a computer program

AI-generated content may be incorrect.

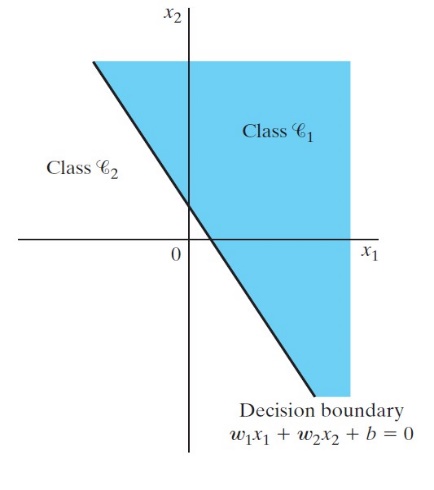
**✅ Explanation:**

* np.dot(weights, inputs) → computes the sum ∑wx​
* + bias → adds the external bias b
* The heaviside\_step() returns:
  + 1 if the result is > 0
  + 0 otherwise

The goal of the perceptron is to correctly classify the set of externally applied stimuli *x*1, *x*2, ..., *xm* into one of two classes, C1 or C2. The decision rule for the classification is to assign the sample represented by the inputs *x*1, *x*2, ..., *xm* to class C1 if the perceptron output *y* is 1 and to class C2 if it is 0. To develop insight into the behavior of a classifier, it is customary to plot a map of the decision regions in the *m*-dimensional signal space spanned by the m input variables *x*1, *x*2, ..., *xm*. In the simplest form of the perceptron, there are two decision regions separated by a hyperplane, which is defined by:

*.*

This is illustrated in Figure 2 for the case of two input variables *x*1 and *x*2, for which the decision boundary takes the form of a straight line.



*Figure 2: Example of hyperplane for a bi-dimensional, 2-classes classification problem.*

A graph of a line

AI-generated content may be incorrect.

✅ The updated plot now matches the **figure you provided**:

* The decision boundary is a **straight line** separating the plane into two classes.
* The **shaded region (Class ℂ₁)** corresponds to perceptron output **1** where:

w1x1+w2x2+b≥0w\_1 x\_1 + w\_2 x\_2 + b \geq 0w1​x1​+w2​x2​+b≥0

* The **white region (Class ℂ₂)** corresponds to output **0**.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A sample (*x*1*,x*2) that lies above the boundary line is assigned to class C1, and a sample (*x*1*,x*2) that lies below the boundary line is assigned to class C2. Note also that the effect of the bias *b* is merely to shift the decision boundary away from the origin. The synaptic weights *w*1, *w*2, ..., *wm* of the perceptron can be adapted on an iteration by iteration basis called *error-correction rule*. To derive this rule, it is more convenient to work with the modified model in Figure 3. In this model, equivalent to the one of Figure 1, the bias *b* is treated as a synaptic weight driven by a fixed input equal to +1. We may thus define the (*m* + 1)-by-1 input vector:

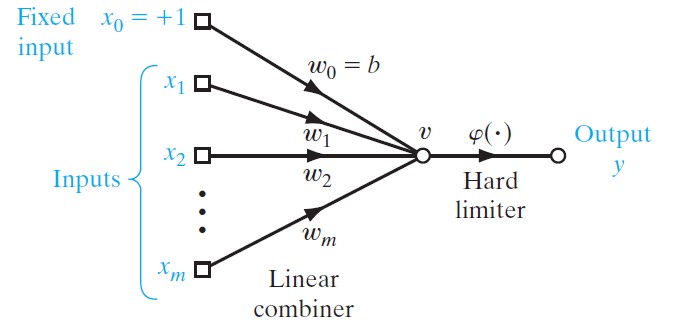
*,*

where *n* denotes the time-step in applying the algorithm, x0(*n*) = 1 and *x*(*n*) denotes the *n*th input vector of the data set {*x*(1)*, ... , x*(*N*)}. Correspondingly, we define the (*m* + 1)-by-1 weight vector as:

T,

where *w*0(*n*) = *b*(*n*) and *w*(*n*) denotes the weight vector at the *n*th iteration of the algorithm. Accordingly, the linear combiner output is written in the compact form as follows:

Suppose then that the input variables of the perceptron originate from two linearly separable classes. Let **Ҥ**1 be the subspace of training vectors that belong to class C1, and let **Ҥ**2 be the subspace of training vectors that belong to class C2. The union of **Ҥ**1 and **Ҥ**2 is the complete space denoted by **Ҥ**. Given the sets of vectors **Ҥ**1 and **Ҥ**2 to train the



*Figure 3: Modified model of perceptron*

classifier, the training process involves the adjustment of the weight vector *w* in such a way that the two classes C1 and C2 are linearly separable. That is, there exists a weight vector *w* such that we may state:

*w*T*x >* 0 for every input vector *x* belonging to class C1

*w*T*x* ≤ 0 for every input vector *x* belonging to class C2

We have arbitrarily chosen to say that the input vector *x* belongs to class C2 if *w*T*x* = 0. Given the subsets of training vectors **Ҥ**1 and **Ҥ**2, the training problem for the perceptron is then to find a weight vector *w* such that the two inequalities above are satisfied. The algorithm for adapting the weight vector of the elementary perceptron may now be formulated as follows:

1. If the *n*th member of the training set, *x*(*n*), is correctly classified by the weight vector *w*(*n*) computed at the *n*th iteration of the algorithm, no correction is made to the weight vector of the perceptron according to the rule:

*w*(*n* + 1) = *w*(*n*) if *w*T(*n*)*x*(*n*) *>* 0 and *x*(*n*) belongs to class C1,

*w*(*n* + 1) = *w*(*n*) if *w*T(*n*)*x*(*n*) ≤ 0 and *x*(*n*) belongs to class C2.

1. Otherwise, the weight vector of the perceptron is updated according to the rule:

*w*(*n* + 1) = *w*(*n*) − *η*(*n*)*x*(*n*) if *w*T(*n*)*x*(*n*) *>* 0 and *x*(*n*) belongs to class C2,

*w*(*n* + 1) = *w*(*n*) + *η*(*n*)*x*(*n*) if *w*T(*n*)*x*(*n*) ≤ 0 and *x*(*n*) belongs to class C1,

where the learning-rate parameter *η*(*n*) controls the adjustment applied to the weight vector at iteration *n*. If *η*(*n*) = *η >* 0, where *η* is a constant independent of the iteration number *n*, then we have a fixed-increment adaptation rule for the perceptron. The adaptation of the weight vector *w*(*n*) is nicely summarized in the form:

*w*(*n* + 1) = *w*(*n*) + *η* (*d*(*n*) − *y*(*n*))*x*(*n*)*,* (1)

where *d*(*n*) is the desired output for *x*(*n*).

**What Changed to Match the Figure:**

* Introduced **x₀ = 1** as the first input of each training vector.
* Treated **w₀ = b** as part of the **weight vector**.
* The model structure and learning rule now match **Figure 3**.

**Getting started!**

# Real data

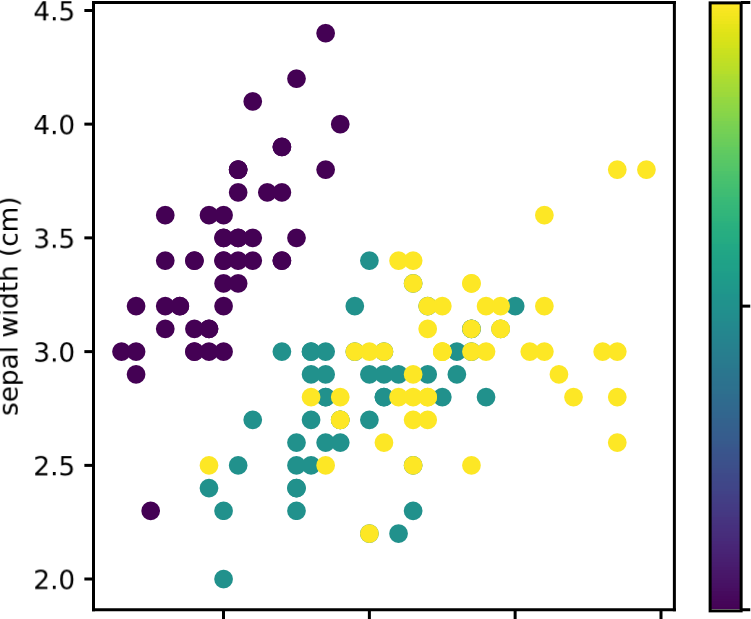
Let’s play with the real data, called the *Iris flower data set*. This data set is widely used in machine learning and can be found in <https://archive.ics.uci.edu/ml/datasets/iris> or in the machine learning package *Scikit-learn*.

**Short description:** This data set contains a set of 150 samples, which consists of 50 samples from each of three species of Iris: *setosa* (label 0), *versicolor* (label 1), and *virginica* (label 2). Each sample was measured in four features: sepal length, sepal width, petal length, and petal width.

**Data Preparation and Visualization:** Split the data set into a balanced (with respect to the labels) training and test set, containing respectively 80% and 20% of the data set.

## Application to binary classification

1. Visualize the first two features of the training set, *i.e.*, sepal length and sepal width, and their corresponding labels/classes. You should obtain a figure similar to the Fig. 4a.
2. Now consider only the data set containing two classes: *setosa* and *versicolor*, you should get the Fig. 4b. Classify the data set into two classes with the Perceptron. Report the training and test errors. Comment.



setosa

versicolor

virginica

2.0

2.5

3.0

3.5

4.0

4.5

sepal width (cm)



setosa

versicolor

5 6 7 8 4.5 5.0 5.5 6.0 6.5 7.0 sepal length (cm) sepal length (cm)

(a) (b)

*Figure 4: Visualization of two features of the Iris data set: (a) shows the data set of all 3 classes and (b) shows only the data set of the first 2 classes.*

*Figure 4a*

*Ảnh có chứa văn bản, biểu đồ, hàng, Sơ đồ

Mô tả được tạo tự động*

*Ảnh có chứa văn bản, đồ điện tử, ảnh chụp màn hình, màn hình

Mô tả được tạo tự động*

**Short Description:** This figure shows a scatter plot of the training set using the first two features of the Iris dataset: sepal length and sepal width. The data points are color-coded by class (setosa, versicolor, virginica), allowing a visual comparison of the class distributions in two-dimensional space.

*Figure 4b*

*Ảnh có chứa văn bản, biểu đồ, ảnh chụp màn hình, số

Mô tả được tạo tự động*

*Ảnh có chứa văn bản, đồ điện tử, màn hình, ảnh chụp màn hình

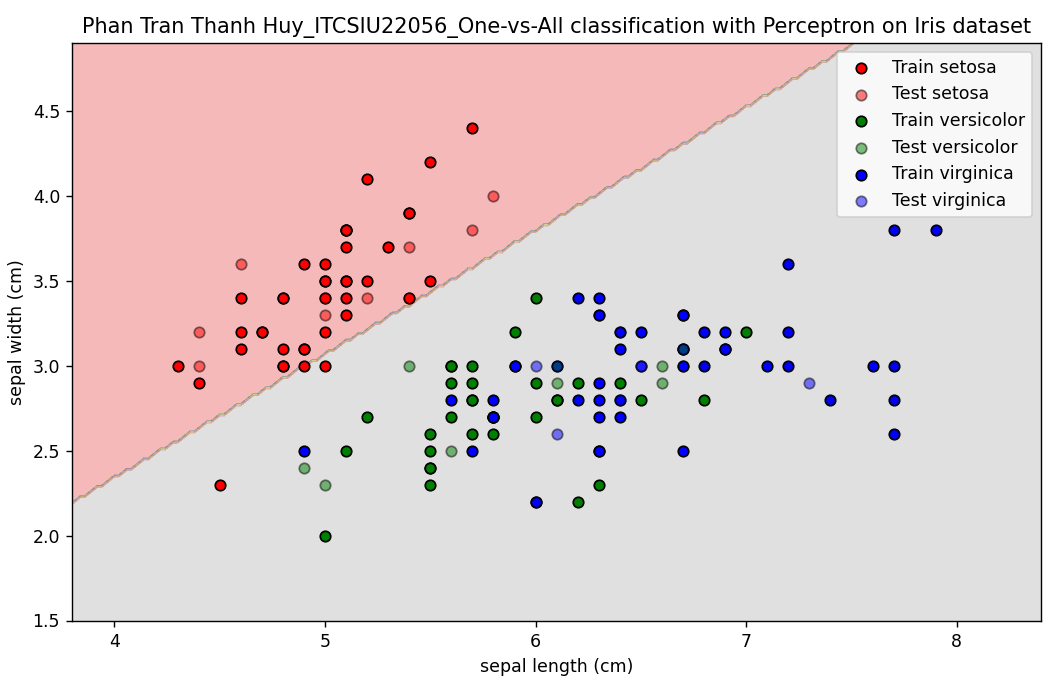
Mô tả được tạo tự động*

**Short Description:** This figure visualizes only two classes from the Iris dataset—setosa and versicolor—based on sepal length and sepal width. It shows a clearer separation between the two classes and is used as the basis for binary classification using a Perceptron.

**(Optional) Application to multi-label classification**

In this section, we take into account all 3 classes of the data set for multiclass, by implementing the *One-versus-All* method (for more information check the chapter 17 of the Book: “Understanding Machine Learning: From Theory to Algorithms”, by Shai Shalev-Shwartz and Shai Ben-David).

Propose a method to implement a three class classifier with multiple binary classifiers.



Ảnh có chứa văn bản, ảnh chụp màn hình, màn hình, phần mềm

Mô tả được tạo tự động

**Explanation:**According to the requirements, I apply the code that:

* OneVsRestClassifier from sklearn is a wrapper that automatically creates these binary classifiers for each class.
* Inside, using Perceptron as the binary classifier algorithm. Perceptron is a simple linear classifier.
* Using only the first 2 features of the Iris dataset for simplicity and visualization.
* Splitting the dataset into training and testing sets.
* Training (fit) the OvA classifier on training data.
* Predicting on training and test data, then calculate accuracy to evaluate performance.
* Finally, visualizing the decision boundaries learned by the classifier.

For OVA:

The One-Versus-All (OvA) method works by training one binary Perceptron classifier for each class in the Iris dataset (Setosa, Versicolor, Virginica), where each classifier learns to distinguish one specific class from the other two. Using OneVsRestClassifier, it fits three separate classifiers: one for Setosa vs. others, one for Versicolor vs. others, and one for Virginica vs. others. During prediction, each input is passed through all three classifiers, and the class whose classifier outputs the highest score (e.g., confidence or distance from the decision boundary) is selected as the final prediction. This approach effectively transforms a multiclass problem into multiple binary problems and is evaluated using accuracy on both training and test data.

1. The text and the pictures of this section were taken from Haykin, O., “Neural Networks and Learning Machines”, Pearson, 2009 [↑](#footnote-ref-1)