A Primer on Perceptrons

The basic constructions and operations of perceptron networks will be formulated here in terms of matrices, providing a viewpoint close to conventional matrix algebra.
Rudiments of sets, functions, vectors and matrices are the only requisites to read this primer.
Perceptrons were introduced by Warren McCulloch and Walter Pitts in 1943. For a glimpse at the vast literature on the subject, perform a Web search.

General definitions

Vectors

A *vector*, or *n-vector*, $x = (x_1, \ldots, x_n)$ is an element of n-dimensional Euclidean space, $x \in \mathbb{R}^n$.

Binary vectors

A *binary n-vector* is a vector $b = (b_1, \ldots, b_n)$ with all its coordinates binary, $b_i = 1$ or $b_i = 0$.

Forms

A *form*, or *single output form in n-variables*, is a function $f : \mathbb{R}^n \to \mathbb{R}$ given by the formula

$$\begin{aligned} f(x) &= f(x_1, \ldots, x_n) \\ &= w_0 + \textstyle\sum_{i=1}^{n} w_i\, x_i \end{aligned}$$

The coefficients w_i are called *weights*. Weight w_0 is the *bias* and can take on any value but at least one of the remaining weights w_i, $i = 1, \ldots, n$, must be non-zero; this is equivalent to f being a non-constant function. The terms 'n-variables' and 'n-inputs' are used interchangeably.

m-forms

An *m-form*, or *(multiple) m-output form in n-variables* is a product $F = f_1 \times \cdots \times f_m : \mathbb{R}^n \to \mathbb{R}^m$ of m single output forms $f_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$, that is

$$F(x) = (f_1(x), \ldots, f_m(x))$$

The form F can also be written as a composition $F = \tau \circ T$ where $T : \mathbb{R}^n \to \mathbb{R}^m$ is the linear transformation $T = F - F(0)$ and $\tau : \mathbb{R}^m \to \mathbb{R}^m$ is translation by $F(0)$, $\tau(y) = y + F(0)$. Because each f_j is non-constant the components of T are non-zero linear functions.

Heaviside function

The *Heaviside function* $h : \mathbb{R} \to \{1, 0\} \subseteq \mathbb{R}$ is defined as

$$h(t) = \begin{cases} 1 \text{ if } t \geq 0 \\ 0 \text{ if } t < 0 \end{cases}$$

m-DIMENSIONAL HEAVISIDE FUNCTION

The *m-dimensional Heaviside function* $h^m : \mathbb{R}^m \to \{1,0\}^m \subseteq \mathbb{R}^m$ consists in applying h to components: For $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$

$$h^m(y) = h(y_1) \times \overset{m}{\overbrace{\cdots}} \times h(y_m)$$

Perceptrons

UNITS

A *perceptron unit* with n inputs is a function $p : \mathbb{R}^n \to \{1,0\}$ equal to the composition of Heaviside function h with a form f in n variables

$$p = h \circ f$$

Here the small circle $\circ$ indicates the usual composition of functions. This can also be written as $p(x) = h(f(x))$.

LAYERS

A *perceptron layer* with n inputs and m outputs is a function $L : \mathbb{R}^n \to \{1,0\}^m \subseteq \mathbb{R}^m$ equal to the product of m perceptron units, $p_j = h \circ f_j : \mathbb{R}^n \to \{1,0\} \subseteq \mathbb{R}$, $j = 1, \ldots, m$, each unit being in n variables

$$\begin{aligned} L(x) &= p_1(x) \times \cdots \times p_m(x) \\ &= h \circ f_1(x) \times \cdots \times h \circ f_m(x) \end{aligned}$$

Alternatively, L equals the composition $L = h^m \circ F$ of the m-output form $F = f_1 \times \cdots \times f_m : \mathbb{R}^n \to \mathbb{R}^m$ with the m-dimensional Heaviside function.

NETWORKS

A *perceptron network* with k layers, n_0 inputs and n_k outputs is a function $P = \mathbb{R}^{n_0} \to \{1,0\}^{n_k} \subseteq \mathbb{R}^{n_k}$ equal to the composition of k perceptron layers. In more detail, P is a k-layer perceptron network if there exist perceptron layers $L_i : \mathbb{R}^{n_{i-1}} \to \{1,0\}^{n_i} \subseteq \mathbb{R}^{n_i}$, $i = 1, \ldots, k$, the range of each contained in the domain of the next, such that

$$P = L_k \circ \cdots \circ L_1$$

Network P can also be written as

$$P(x) = L_k(L_{k-1}(\cdots(L_1(x))\cdots))$$

Or inductively, let $y_1 = L_1(x)$, $y_{j+1} = L_{j+1}(y_j)$, $j \geq 1$, then $P(x) = y_k$.
The expression of P as a composition of layers is not unique. It can be proved in general that, for all $k > 0$, any k-layer perceptron network with n_0 inputs and

n_k outputs, say $P = L_k \circ \cdots \circ L_1$ can be expressed as a composition of three layers, $P = L'_3 \circ L'_2 \circ L'_1$. Additionally, one can take $L'_1 = L_1$.

Representation and evaluation

According to the previous definitions units, layers and networks are real valued functions of several variables. These objects and the procedures for their effective calculation can be represented numerically in terms of matrices and Heaviside functions. Details follow.

BORDERED PRODUCTS

The *bordered product* of an $m \times (n+1)$ matrix and an $n \times 1$ matrix, denoted with an asterisk $*$ as shown below, is by definition an $m \times 1$ matrix defined by the expression

$$\begin{bmatrix} a_{1\,0} & a_{1\,1} & \cdots & a_{1\,n} \\ a_{2\,0} & a_{2\,1} & \cdots & a_{2\,n} \\ \vdots & \vdots & & \vdots \\ a_{m\,0} & a_{m\,1} & \cdots & a_{m\,n} \end{bmatrix} * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1\,0} + \sum_{i=1}^{n} a_{1\,i}\, x_i \\ a_{2\,0} + \sum_{i=1}^{n} a_{2\,i}\, x_i \\ \vdots \\ a_{m\,0} + \sum_{i=1}^{n} a_{m\,i}\, x_i \end{bmatrix}$$

Bordered products are commonly used for non-homogeneous linear equations and to represent non-homogeneous linear transformations. The term ‘bordered’ refers to the left border, or first column, of the matrix. These are independent terms in the formulas.

VECTORS AND MATRICES

A n-vector $x \in \mathbb{R}^n$ is represented as an $n \times 1$, or single column height n matrix with entries equal to the coordinates

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

FORMS AND MATRICES

A single output form f in n variables is represented as an $1 \times (n+1)$, or single row, matrix with entries equal to the weights

$$f = [w_0 \; w_1 \; \cdots \; w_n]$$

VALUES OF FORMS

The value of a single output form f in n variables, on an n-vector x, can be calculated as the bordered product of the respective single row and single column matrices

$$\begin{aligned} f(x) &= [w_0 \; w_1 \; \cdots \; w_n] * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= w_0 + \textstyle\sum_{i=1}^{n} w_i\, x_i \end{aligned}$$

m-FORMS AND MATRICES

An m-output form in n variables $F = f_1 \times \cdots \times f_m$, where $f_j = [w_{j\,0}\, w_{j\,1}\, \ldots\, w_{j\,n}]$, $j = 1, \ldots, m$, is to be represented by a corresponding $m \times (n+1)$ matrix having entries equal to the weights

$$F = \begin{bmatrix} w_{1\,0} & w_{1\,1} & \cdots & w_{1\,n} \\ w_{2\,0} & w_{2\,1} & \cdots & w_{2\,n} \\ \vdots & \vdots & & \vdots \\ w_{m\,0} & w_{m\,1} & \cdots & w_{m\,n} \end{bmatrix}$$

VALUES OF m-FORMS

The value of an m-output form F in n variables on an n-vector x can be calculated as the bordered product of the respective matrices

$$\begin{aligned} F(x) &= \begin{bmatrix} w_{1\,0} & w_{1\,1} & \cdots & w_{1\,n} \\ w_{2\,0} & w_{2\,1} & \cdots & w_{2\,n} \\ \vdots & \vdots & & \vdots \\ w_{m\,0} & w_{m\,1} & \cdots & w_{m\,n} \end{bmatrix} * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \\ &= \begin{bmatrix} w_{1\,0} + \sum_{i=1}^{n} w_{1\,i}\, x_i \\ w_{2\,0} + \sum_{i=1}^{n} w_{2\,i}\, x_i \\ \vdots \\ w_{m\,0} + \sum_{i=1}^{n} w_{m\,i}\, x_i \end{bmatrix} \end{aligned}$$

HEAVISIDE TRANSFORM AND SYMBOL

Will use h as an exponent to indicate composition on the left with a Heaviside function. Thus, if X is a set and $G : X \to \mathbb{R}^m$ is a function then, by definition,

$$G^h = h^m \circ G$$

Function $G^h : X \to \{1, 0\}^m$ is the *Heaviside transform* of G. In particular if $m = 1$ then $G^h = h \circ G$.
The correspondence $G \to G^h$ is a non-linear operator transforming m-vector valued functions G into binary m-vector valued functions G^h.
It will be convenient to use the term *Heaviside symbol* for the exponent h.

UNITS AND MATRICES

A perceptron unit in n variables, $p = f^h$, will be represented by the $1 \times (n+1)$ matrix of f and Heaviside symbol

$$\begin{aligned} p &= h \circ f \\ &= f^h \\ &= [w_0\ w_1\ \ldots\ w_n]^h \end{aligned}$$

VALUES OF PERCEPTRON UNITS

The value of a perceptron unit $p = h \circ f$ in n variables, on an n-vector x, can be expressed in terms of the Heaviside function and a bordered product

$$\begin{aligned} p(x) &= [w_0\ w_1\ \ldots, w_n]^h \left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) \\ &= h \left([w_0\ w_1\ \ldots w_n] * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) \\ &= h(w_0 + \textstyle\sum_{i=1}^n w_i\, x_i) \end{aligned}$$

Layers and matrices

An m-output perceptron layer $L = h^m \circ F$ in n variables is represented by the $m \times (n+1)$ matrix of F affected by Heaviside symbol

$$L = \begin{bmatrix} w_{1\,0} & w_{1\,1} & \cdots & w_{1\,n} \\ w_{2\,0} & w_{2\,1} & \cdots & w_{2\,n} \\ \vdots & \vdots & & \vdots \\ w_{m\,0} & w_{m\,1} & \cdots & w_{m\,n} \end{bmatrix}^h$$

Values of perceptron layers

The value of an m-output perceptron layer $L = h^m \circ F$ in n variables, on an n-vector x, is a binary vector that can be expressed in terms of bordered products of matrices and Heaviside function:

$$\begin{aligned} L(x) &= \begin{bmatrix} w_{1\,0} & w_{1\,1} & \cdots & w_{1\,n} \\ w_{2\,0} & w_{2\,1} & \cdots & w_{2\,n} \\ \vdots & \vdots & & \vdots \\ w_{m\,0} & w_{m\,1} & \cdots & w_{m\,n} \end{bmatrix}^h \left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) \\ &= h^m \left(\begin{bmatrix} w_{1\,0} & w_{1\,1} & \cdots & w_{1\,n} \\ w_{2\,0} & w_{2\,1} & \cdots & w_{2\,n} \\ \vdots & \vdots & & \vdots \\ w_{m\,0} & w_{m\,1} & \cdots & w_{m\,n} \end{bmatrix} * \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) \end{aligned}$$

Perceptron Networks and matrices

A k-layer, n_0-input, n_k-output perceptron network $P = L_k \circ \cdots \circ L_1$ with

$$L_j = \begin{bmatrix} w^{(j)}_{1\,0} & w^{(j)}_{1\,1} & \cdots & w^{(j)}_{1\,n_{j-1}} \\ w^{(j)}_{2\,0} & w^{(j)}_{2\,1} & \cdots & w^{(j)}_{2\,n_{j-1}} \\ \vdots & \vdots & & \vdots \\ w^{(j)}_{n_j\,0} & w^{(j)}_{n_j\,1} & \cdots & w^{(j)}_{n_j\,n_{j-1}} \end{bmatrix}^h \qquad j = 1, \ldots, k$$

can be represented in terms of the matrices corresponding to the layers as

$$P = \begin{bmatrix} w^{(k)}_{1\,0} & w^{(k)}_{1\,1} & \cdots & w^{(k)}_{1\,n_{k-1}} \\ w^{(k)}_{2\,0} & w^{(k)}_{2\,1} & \cdots & w^{(k)}_{2\,n_{k-1}} \\ \vdots & \vdots & & \vdots \\ w^{(k)}_{n_k\,0} & w^{(k)}_{n_k\,1} & \cdots & w^{(k)}_{n_k\,n_{k-1}} \end{bmatrix}^h \circ \cdots \circ \begin{bmatrix} w^{(1)}_{1\,0} & w^{(1)}_{1\,1} & \cdots & w^{(1)}_{1\,n_0} \\ w^{(1)}_{2\,0} & w^{(1)}_{2\,1} & \cdots & w^{(1)}_{2\,n_0} \\ \vdots & \vdots & & \vdots \\ w^{(1)}_{n_1\,0} & w^{(1)}_{n_1\,1} & \cdots & w^{(1)}_{n_1\,n_0} \end{bmatrix}^h$$

where the small circles $\circ$ indicate composition of functions.

Values of perceptron networks

The value of the above perceptron network on an n_0-vector x can be expressed in terms of bordered products of matrices and Heaviside functions as

$$P(x) = \begin{bmatrix} w^{(k)}_{1\,0} & w^{(k)}_{1\,1} & \cdots & w^{(k)}_{1\,n_{k-1}} \\ w^{(k)}_{2\,0} & w^{(k)}_{2\,1} & \cdots & w^{(k)}_{2\,n_{k-1}} \\ \vdots & \vdots & & \vdots \\ w^{(k)}_{n_k\,0} & w^{(k)}_{n_k\,1} & \cdots & w^{(k)}_{n_k\,n_{k-1}} \end{bmatrix}^h \left(\cdots \left(\begin{bmatrix} w^{(1)}_{1\,0} & w^{(1)}_{1\,1} & \cdots & w^{(1)}_{1\,n_0} \\ w^{(1)}_{2\,0} & w^{(1)}_{2\,1} & \cdots & w^{(1)}_{2\,n_0} \\ \vdots & \vdots & & \vdots \\ w^{(1)}_{n_1\,0} & w^{(1)}_{n_1\,1} & \cdots & w^{(1)}_{n_1\,n_0} \end{bmatrix}^h \begin{bmatrix} x_1 \\ \vdots \\ x_{n_0} \end{bmatrix} \right) \cdots \right)$$

For an equivalent inductive formulation let

$$\begin{bmatrix} y^{(1)}_1 \\ \vdots \\ y^{(1)}_{n_1} \end{bmatrix} = \begin{bmatrix} w^{(1)}_{1\,0} & w^{(1)}_{1\,1} & \cdots & w^{(1)}_{1\,n_0} \\ w^{(1)}_{2\,0} & w^{(1)}_{2\,1} & \cdots & w^{(1)}_{2\,n_0} \\ \vdots & \vdots & & \vdots \\ w^{(1)}_{n_1\,0} & w^{(1)}_{n_1\,1} & \cdots & w^{(1)}_{n_1\,n_0} \end{bmatrix}^h \begin{bmatrix} x_1 \\ \vdots \\ x_{n_0} \end{bmatrix}$$

and for $j \geq 1$ take

$$\begin{bmatrix} y^{(j+1)}_1 \\ \vdots \\ y^{(j+1)}_{n_{j+1}} \end{bmatrix} = \begin{bmatrix} w^{(j+1)}_{1\,0} & w^{(j+1)}_{1\,1} & \cdots & w^{(j+1)}_{1\,n_j} \\ w^{(j+1)}_{2\,0} & w^{(j+1)}_{2\,1} & \cdots & w^{(j+1)}_{2\,n_j} \\ \vdots & \vdots & & \vdots \\ w^{(j+1)}_{n_{j+1}\,0} & w^{(j+1)}_{n_{j+1}\,1} & \cdots & w^{(j+1)}_{n_{j+1}\,n_j} \end{bmatrix}^h \begin{bmatrix} y^{(j)}_1 \\ \vdots \\ y^{(j)}_{n_j} \end{bmatrix}$$

which is always a binary vector. Then

$$P(x) = \begin{bmatrix} y^{(k)}_1 \\ \vdots \\ y^{(k)}_{n_k} \end{bmatrix}$$

Variants

Sigmoids
The *sigmoid* or *logistic function* is the function $\sigma : \mathbb{R} \to (0,1)$ given by

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

The m-dimensional sigmoid is then $\sigma^m(y_1, \ldots, y_m) = (\sigma(y_1), \ldots, \sigma(y_m))$.
Replace now Heaviside function h with the sigmoid σ to obtain the *sigmoid transform*, $G \to \sigma^m \circ G$, with corresponding *sigmoid symbol*, $G^\sigma = \sigma^m \circ G$ as well as *sigmoid units* in n variables $s : R^n \to (0,1)$,

$$\begin{array}{rcl} s & = & \sigma \circ f \\ & = & f^\sigma \end{array}$$

m-output sigmoid layers $L : R^n \to (0,1)^m$,

$$\begin{array}{rcl} L & = & \sigma^m \circ F \\ & = & F^\sigma \end{array}$$

and *k-layer sigmoid networks* $S : R^{n_0} \to (0,1)^{n_k}$,

$$\begin{array}{rcl} S & = & (\sigma^{n_k} \circ F_k) \circ \cdots \circ (\sigma^{n_1} \circ F_1) \\ & = & F_k^\sigma \circ \cdots \circ F_1^\sigma \\ & = & L_k \circ \cdots \circ L_1 \end{array}$$

Similarly to perceptrons, all these functions and their values can be represented by means of matrices, with sigmoid symbols replacing Heaviside symbols.
Sigmoid networks are differentiable functions of their weights.

April 9, 2005.