CAS ETH Machine Learning in Finance and Insurance. Mini-exercises - Lecture 3.

Andrea Ferrario

Version of March 21, 2024

Activation functions

1. (Linear activation). Plot the linear activation function $f(x) = x$, $x \in \mathbb{R}$.

2. (Heaviside function). Plot the Heaviside function $H(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$.

3. (Sigmoid functions). Consider the sigmoid functions $\sigma_c(x) = \frac{1}{1+e^{-cx}}$, where $c > 0$ and $x \in \mathbb{R}$. Compute $\sigma_c(0)$ and plot $\sigma_1(x)$. What is the name of $\sigma_1(x)$? Finally, prove the following results:

   (a) $\lim_{c \to +\infty} \sigma_c(x) = H(x)$ $x \neq 0$.

   (b) $\sigma_c(-x) = 1 - \sigma_c(x)$.

   (c) $\frac{d\sigma_c}{dx} = c\sigma_c(1 - \sigma_c)$.

4. (Bipolar step function). Plot the bipolar step function $S(x) = \begin{cases} -1, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0 \end{cases}$. Prove that $S(x) = 2H(x) - 1$, for all $x \in \mathbb{R}$.

5. (Hockey stick functions). The hockey stick functions have graphs that resemble a hockey stick or an increasing L-shaped curve. There exists different types of hockey stick functions. Here, consider the class of *parametric rectified linear unit functions*

$$\mathrm{PReLu}_\alpha(x) = \begin{cases} \alpha x, & \text{if } x < 0 \\ x, & \text{if } x \geq 0 \end{cases}$$

   and $\alpha \geq 0$. Plot $\mathrm{PReLu}_\alpha(x)$, for different values of α. The function $\mathrm{PReLu}_0 := \mathrm{ReLu}$ is called the *rectified linear unit function.* Then, consider the following exercises:

   (a) Prove that $\mathrm{ReLu}(x) = \frac{1}{2}x(S(x) + 1)$.

   (b) Compute the first derivative of $\mathrm{ReLu}(x)$ (where is it defined?). Express it in terms of the Heaviside function $H(x)$.

6. (Softmax function). Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. $\|z\|_1$ denotes the L-1 norm of any vector z in $\mathbb{R}^n$, i.e., $\|z\|_1 = \sum_{i=1}^n |z_i|$. The function

$$\mathrm{softmax}(x) = \begin{pmatrix} \frac{e^{x_1}}{\|e^x\|_1} \\ \vdots \\ \frac{e^{x_n}}{\|e^x\|_1} \end{pmatrix}$$

is called the $n-$valued softmax function. Here, $e^x = \begin{pmatrix} e^{x_1} \\ \vdots \\ e^{x_n} \end{pmatrix}$. Then, consider the following exercises:

(a) Compute $\|\mathrm{softmax(x)}\|_1, \ \forall x \in \mathbb{R}^n$.

(b) Show that the softmax function is invariant with respect to addition of constant vectors $c = \begin{pmatrix} c_1 \\ \vdots \\ c_1 \end{pmatrix} \in \mathbb{R}^n$, i.e.,

$$\mathrm{softmax}(x + c) = \mathrm{softmax}(x), \text{ for all } x \in \mathbb{R}^n.$$

Remark: this property is used in classification problems with $c = \begin{pmatrix} \max\limits_{i=1,\ldots,n} x_i \\ \vdots \\ \max\limits_{i=1,\ldots,n} x_i \end{pmatrix}$ to obtain a more stable numerically variant of the softmax function.

Abstract neurons

1. Explain the concept of an abstract neuron and its role in artificial neural networks.
2. Describe the mathematical representation of a perceptron, sigmoid and linear neurons.
3. Prove that the number of k-ary Boolean functions is 2^{2^k}, $k \geq 1$.
4. Recall that $\neg x$ is the negation of the Boolean variable x. In formulae: $\neg x = \begin{cases} 1, & \text{if } x = 0 \\ 0, & \text{if } x = 1 \end{cases}$.

   (a) Show that a perceptron can learn the Boolean function $f(x) = \neg x$, with $x \in \{0, 1\}$. Can a linear abstract neuron learn the same function?

   (b) Show that a perceptron can learn the Boolean function $f(x_1, x_2) = x_1 \wedge \neg x_2$, with $x_1, x_2 \in \{0, 1\}$.

(c) Show that a perceptron can learn the Boolean function $f(x_1, x_2) = x_1 \vee \neg x_2$, with $x_1, x_2 \in \{0, 1\}$.

(d) Prove graphically that a perceptron cannot learn the XOR function. (Remember that $\text{XOR}(x_1, x_2) = \neg(x_1 \wedge x_2) \wedge (x_1 \vee x_2)$)

(e) Prove analytically that a perceptron cannot learn the XOR function.

5. Consider the function $f(x_1, x_2, x_3) = (x_1 \wedge x_2) \wedge x_3$, with $x_1, x_2, x_3 \in \{0, 1\}$.

(a) Prove that $\wedge$ is associative, namely, $f(x_1, x_2, x_3) = (x_1 \wedge x_2) \wedge x_3 = x_1 \wedge (x_2 \wedge x_3)$ for all $x_1, x_2, x_3 \in \{0, 1\}$. Hints: (1) introduce the Boolean variables $z_{12} := x_1 \wedge x_2$ and $z_{23} := x_2 \wedge x_3$. (2) Using the definition of $\wedge$, fill-out the table below for all combinations (x_1, x_2, x_3). Prove that the two rightmost columns are equal.

x_1	x_2	x_3	z_{12}	z_{23}	$z_{12} \wedge x_3$	$x_1 \wedge z_{23}$
0	0	0	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

(b) Prove graphically that there exists a perceptron that can learn the function f. Hints: (1) draw the set of all combinations (x_1, x_2, x_3) in $\mathbb{R}^3$, with $x_1, x_2, x_3 \in \{0, 1\}$. (2) Remember that a *plane* in $\mathbb{R}^3$ is a set $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_0 = 0\}$ for some parameters $\theta_0, \ldots, \theta_2 \in \mathbb{R}$. (3) Use the table in (a) to determine if a plane exists that can separate the triples (x_1, x_2, x_3) such that $f(x_1, x_2, x_3) = 0$ from those satisfying $f(x_1, x_2, x_3) = 1$.