CentraleSupélec

2nd Year
Elective Course

2022 – 2023

# Artificial Intelligence – Tutorial Session 1
# TS 1 : Machine learning

## 1 Q1. Machine Learning: Potpourri

(a) It is possible for the perceptron algorithm to never terminate on a dataset that is linearly separable in its feature space.

- ☐ True
- ☐ False

(b) If the perceptron algorithm terminates, then it is guaranteed to find a max-margin separating decision boundary.

- ☐ True
- ☐ False

(c) In binary perceptron where the initial weight vector is $\vec{0}$, the final weight vector can be written as a linear combination of the training data feature vectors.

- ☐ True
- ☐ False

(d) For binary class classification, logistic regression produces a linear decision boundary.

- ☐ True
- ☐ False

(e) In the binary classification case, logistic regression is exactly equivalent to a single-layer neural network with a sigmoid activation and the cross-entropy loss function.

- ☐ True
- ☐ False

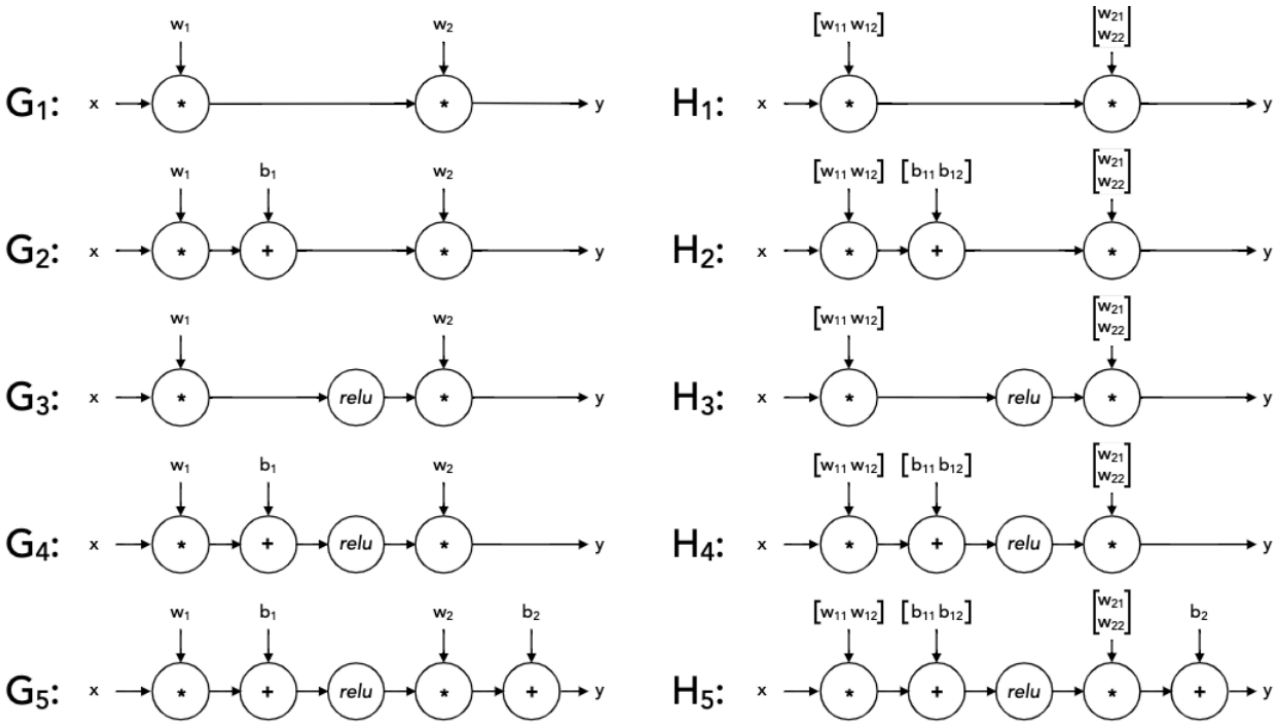
(f) You train a linear classifier on 1,000 training points and discover that the training accuracy is only 50%. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

- ☐ Add novel features
- ☐ Train on more data

(g) You now try training a neural network but you find that the training accuracy is still very low. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

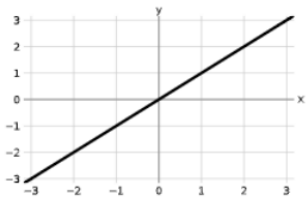
- ☐ Add more hidden layers
- ☐ Add more units to the hidden layers

## 2 Neural networks representation

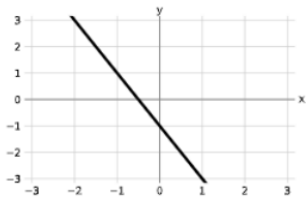



For each of the piecewise linear functions below, mark all networks from the list above that can represent the function exactly on the range $x \in [-\infty, +\infty]$. In the neworks above, *relu* denotes the element-wise ReLU (rectified linear unit) non-linearity: $relu(z) = \max(0, z)$. The neworks $G_i$ use 1-dimensional layers, while the networks $H_i$ have some 2-dimensional intermediate layers.

**(a)**

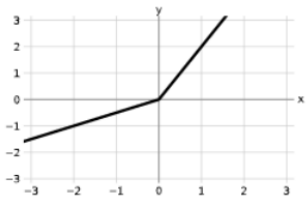


☑ $G_1$ ☑ $H_1$
☑ $G_2$ ☑ $H_2$
☐ $G_3$ ☑ $H_3$
☐ $G_4$ ☑ $H_4$
☐ $G_5$ ☑ $H_5$
○ None of the above

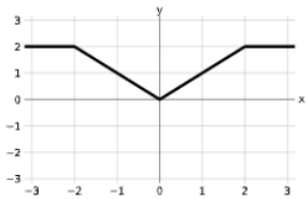


☐ $G_1$ ☐ $H_1$
☑ $G_2$ ☑ $H_2$
☐ $G_3$ ☐ $H_3$
☐ $G_4$ ☑ $H_4$
☐ $G_5$ ☑ $H_5$
○ None of the above

**(b)**



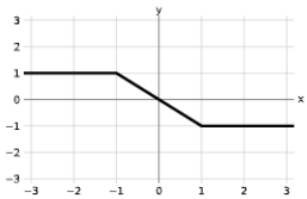
☐ $G_1$ ☐ $H_1$
☐ $G_2$ ☐ $H_2$
☑ $G_3$ ☑ $H_3$
☐ $G_4$ ☑ $H_4$
☐ $G_5$ ☑ $H_5$
○ None of the above



☐ $G_1$ ☐ $H_1$
☐ $G_2$ ☐ $H_2$
☐ $G_3$ ☐ $H_3$
☐ $G_4$ ☐ $H_4$
☐ $G_5$ ☐ $H_5$
○ None of the above

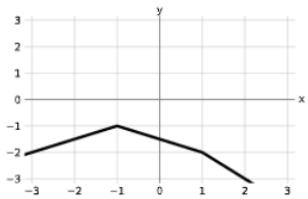
Trop de points d'inflexion (il faut 3 relu)

**(c)**



☐ $G_1$ ☐ $H_1$
☐ $G_2$ ☐ $H_2$
☐ $G_3$ ☐ $H_3$
☐ $G_4$ ☐ $H_4$
☐ $G_5$ ☑ $H_5$
○ None of the above

Il faut décaler selon l'axe y donc il faut un biais après le relu

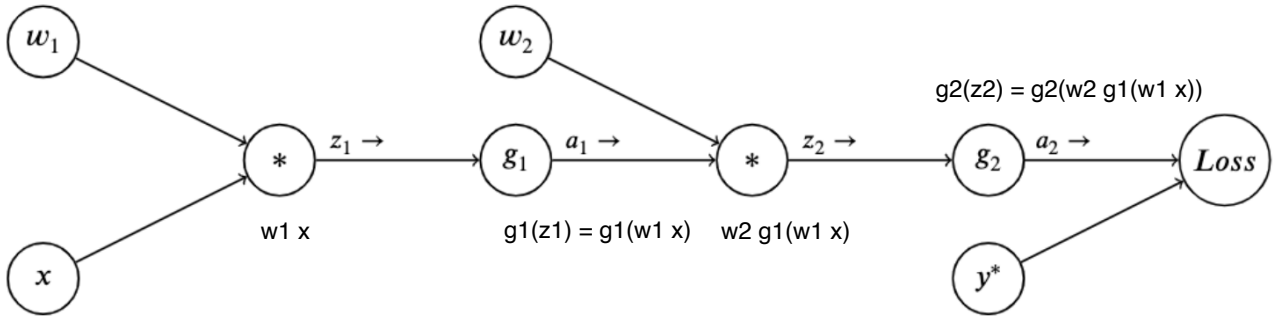


☐ $G_1$ ☐ $H_1$
☐ $G_2$ ☐ $H_2$
☐ $G_3$ ☐ $H_3$
☐ $G_4$ ☑ $H_4$
☐ $G_5$ ☑ $H_5$
○ None of the above

On peut continuer les relu et voir que seul le décalage sur l'axe x sont nécessaires (donc biais avant relu)

# 3 Neural Nets - Computation Graphs

Consider the following computation graph for a simple neural network for binary classification. Here $x$ is a single real-valued input feature with an associated class $y^*$(0 or 1). There are two weight parameters $w_1$ and $w_2$, and non-linearity functions $g_1$ and $g_2$ (to be defined later, below). The network will output a value $a_2$ between 0 and 1, representing the probability of being in class 1 . We will be using a loss function Loss (to be defined later, below), to compare the prediction $a_2$ with the true class $y^*$.



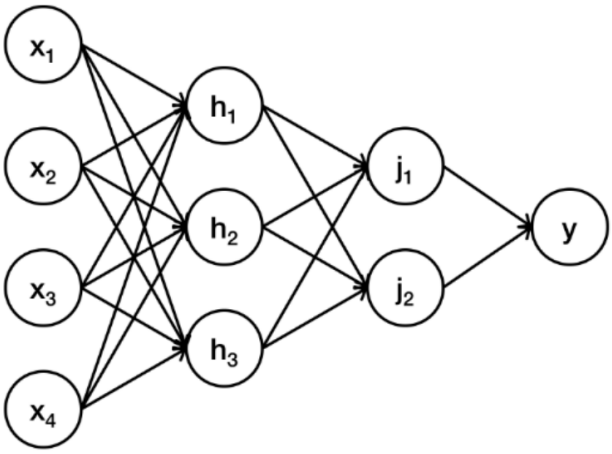

1. Perform the forward pass on this network, writing the output values for each node $z_1, a_1, z_2$ and $a_2$ in terms of the node's input values:

2. Compute the loss Loss $(a_2, y^*)$ in terms of the input $x$, weights $w_i$, and activation functions $g_i$ :
   Loss(a2, y*) = Loss(g2(w2 g1(w1 x)), y*)
3. Now we will work through parts of the backward pass, incrementally. Use the chain rule to derive $\frac{\partial Loss}{\partial w_2}$. Write your expression as a product of partial derivatives at each node: i.e. the partial derivative of the node's output with respect to its inputs. (Hint: the series of expressions you wrote in part 1 will be helpful; you may use any of those variables.) d/dw2 Loss(a2, y*) = d/da2 Loss(a2, y*) * d/dz2 a2 * d/dw2 z2
4. Suppose the loss function is quadratic, Loss $(a_2, y^*) = \frac{1}{2}(a_2 - y^*)^2$, and $g_1$ and $g_2$ are both sigmoid functions $g(z) = \frac{1}{1+e^{-z}}$ (note: it's typically better to use a different type of loss, cross-entropy, for classification problems, but we'll use this to make the math easier).

   Using the chain rule from Part 3 , and the fact that $\frac{\partial g(z)}{\partial z} = g(z)(1 - g(z))$ for the sigmoid function, write $\frac{\partial Loss}{\partial w_2}$ in terms of the values from the forward pass, $y^*, a_1$, and $a_2$ :
   d/dw2 Loss(a2, y*) = (a2-y*) a2 (1-a2) a1
5. Now use the chain rule to derive $\frac{\partial Loss}{\partial w_1}$ as a product of partial derivatives at each node used in the chain rule:
   d/dw1 Loss(a2, y*) = d/da2 Loss(a2, y*) * d/dz2 a2 * d/da1 z2 * d/dz1 a1 * d/dw1 z1
6. Finally, write $\frac{\partial Loss}{\partial w_1}$ in terms of $x, y^*, w_i, a_i, z_i$ : The partial derivatives at each node (in addition to the ones we computed in Part 4) are: d/dw1 Loss(a2, y*) = (a2-y*) a2 (1-a2) w2 a1(1-a1) x
7. What is the gradient descent update for $w_1$ with step-size $\alpha$ in terms of the values computed above?
   w1 <- w1 - alpha * d/dw1 Loss(a2, y*)

# 4 Neural Network Data Sufficiency

The next few problems use the below neural network as a reference. Neurons $h_{1-3}$ and $j_{1-2}$ all use ReLU activation functions. Neuron $y$ uses the identity activation function: $f(x) = x$. In the questions below, let $w_{a,b}$ denote the weight that connects neurons $a$ and $b$. Also, let $o_a$ denote the value that neuron $a$ outputs to its next layer.




Given this network, in the following few problems, you have to decide whether the data given are sufficient for answering the question.

(a) Given the above neural network, what is the value of $o_y$ ?

Data item 1: the values of all weights in the network and the values $o_{h_1}, o_{h_2}, o_{h_3}$

Data item 2: the values of all weights in the network and the values $o_{j_1}, o_{j_2}$

- ☐ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☐ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☑ Each data item alone is sufficient to answer the question.
- ☐ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(b) Given the above neural network, what is the value of $o_{h_1}$ ?

- ☐ Data item 1: the neuron input values, i.e., $o_{x_1}$ through $o_{x_4}$
- ☐ Data item 2: the values $o_{j_1}, o_{j_2}$
- ☐ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☐ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☐ Each data item alone is sufficient to answer the question.
- ☑ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(c) Given the above neural network, what is the value of $o_{j_1}$ ?

- ☐ Data item 1: the values of all weights connecting neurons $h_1, h_2, h_3$ to $j_1, j_2$
- ☐ Data item 2: the values $o_{h_1}, o_{h_2}, o_{h_3}$
- ☐ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☑ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☐ Each data item alone is sufficient to answer the question.
- ☐ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(d) Given the above neural network, what is the value of $\partial o_y / \partial w_{j_2,y}$ ?

- ☐ Data item 1: the value of $o_{j_2}$
- ☐ Data item 2: all weights in the network and the neuron input values, i.e., $o_{x_1}$ through $o_{x_4}$
- ☐ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☐ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☑ Each data item alone is sufficient to answer the question.
- ☐ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(e) Given the above neural network, what is the value of $\partial o_y / \partial w_{h_2,j_2}$ ?

- ☐ Data item 1: the value of $w_{j_2,y}$
- ☐ Data item 2: the value of $\partial o_{j_2} / \partial w_{h_2,j_2}$
- ☐ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☐ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☐ Each data item alone is sufficient to answer the question.
- ☐ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

(f) Given the above neural network, what is the value of $\partial o_y / \partial w_{x_1,h_3}$ ?

- ☐ Data item 1: the value of all weights in the network and the neuron input values, i.e., $o_{x_1}$ through $o_{x_4}$
Data item 2: the value of $w_{x_1,h_3}$
- ☑ Data item (1) alone is sufficient, but data item (2) alone is not sufficient to answer the question.
- ☐ Data item (2) alone is sufficient, but data item (1) alone is not sufficient to answer the question.
- ☐ Both statements taken together are sufficient, but neither data item alone is sufficient.
- ☐ Each data item alone is sufficient to answer the question.
- ☐ Statements (1) and (2) together are not sufficient, and additional data is needed to answer the question.

# 5 Dataset train/test split

In 2017, a team led by Andrew Ng published a paper showing off a Deep Learning model to detect pneumonia. Andrew is one of the most recognized researchers in the world, and the paper showed excellent results. But there was a big problem with their results. Can you spot the issue ?

> **3.1. Training**
>
> We use the ChestX-ray14 dataset released by Wang et al. (2017) which contains 112,120 frontal-view X-ray images of 30,805 unique patients. Wang et al. (2017) annotate each image with up to 14 different thoracic pathology labels using automatic extraction methods on radiology reports. We label images that have pneumonia as one of the annotated pathologies as positive examples and label all other images as negative examples for the pneumonia detection task. We randomly split the entire dataset into 80% training, and 20% validation.