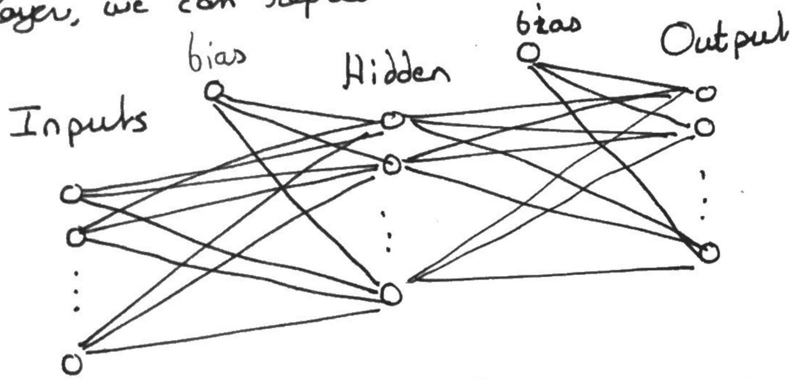
Louis François
ljf 2137
Quentin Glangetas
qag 2000

Deep Learning - HW1

<u>Problem 1</u>: Let's show that any feedforward neural network (FNN) with linear activation function is equivalent to a FNN with no hidden layers.

Let's proceed by induction. For the case $k = 1$, where k is the number of hidden layer, we can represent the FNN as:



Let's define:

- X any vector of inputs
- H the vector of the hidden layer's values from X
- Y the vector of outputs' values from X
- $W_{i \to h}$ the vector of weights from the input layer to the hidden layer
- $B_{i \to h}$ the vector of biases from the input layer to the hidden layer
- F_h the activation function of the hidden layer.
- $W_{h \to o}$ the vector of weights from the hidden layer to the output layer
- $B_{h \to o}$ the vector of biases from the hidden layer to the output layer
- F_o the activation function of the output layer.

With these notations, we are able to write:

$$\begin{cases} H = F_h(W_{i \to h} X + B_{i \to h}) \\ Y = F_o(W_{h \to o} H + B_{h \to o}) \end{cases}$$

We further assumed that F_h and F_o are linear. So we can define their coefficients as:

$$\left| \begin{array}{l} F_h(U) = \alpha_h U + \beta_h \\ F_o(U) = \alpha_o U + \beta_o \end{array} \right.$$

In that case:

$$Y = F_o\left(W_{h\to o} H + B_{h\to o}\right)$$
$$= \alpha_o W_{h\to o} F_h(W_{i\to h} X + B_{i\to h}) + \alpha_o B_{h\to o} + \beta_o$$
$$= \alpha_o W_{h\to o}\left[\alpha_h W_{i\to h} X + \alpha_h B_{i\to h} + \beta_h\right] + \alpha_o B_{h\to o} + \beta_o$$
$$= \left(\alpha_o W_{h\to o} \alpha_h W_{i\to h}\right) X + \left(\alpha_o W_{h\to o} \alpha_h B_{i\to h} + \alpha_o W_{h\to o} \beta_h + \alpha_o B_{h\to o} + \beta_o\right)$$

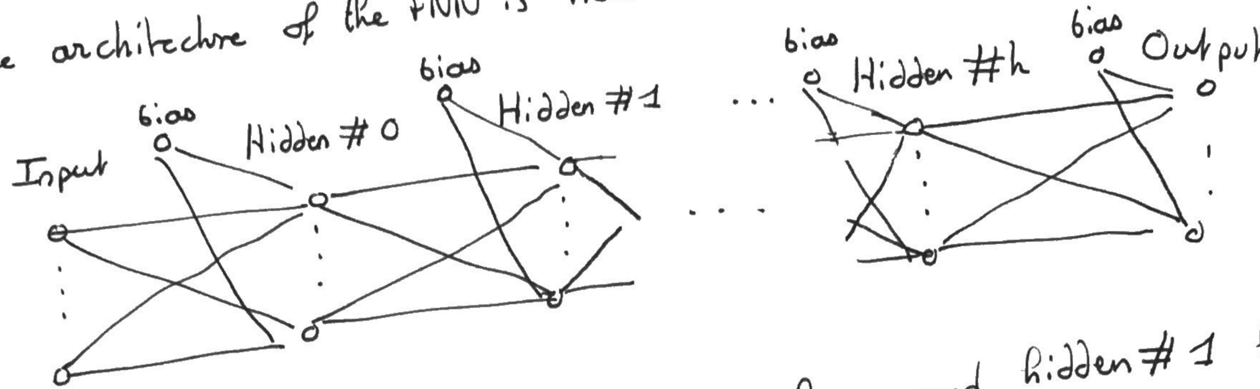
w.l.o.g. we can define $W = \alpha_o W_{h\to o} + \alpha_h W_{i\to h}$
$B = \alpha_o W_{h\to o} \alpha_h B_{i\to h} + \alpha_o W_{h\to o} \beta_h + \alpha_o B_{h\to o} + \beta_o$

so that $\boxed{Y = WX + B}$, this is the equation of a FNN with no hidden layer (and the identity as the activation function).

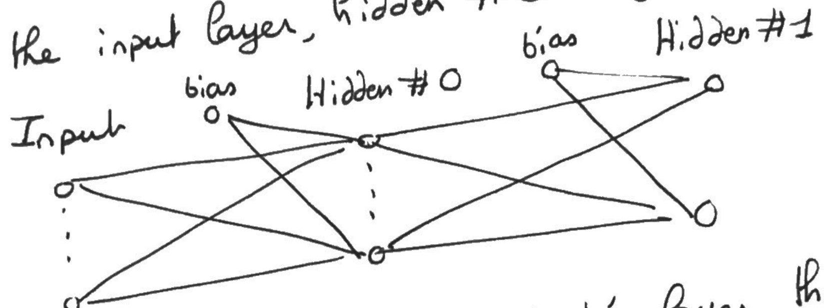
$\boxed{\text{So the result holds true for } k = 1}$

Let's assume the result holds true for $k \in \mathbb{N}^*$, let's prove it remains true for $(k+1)$ hidden layers with linear activation function.

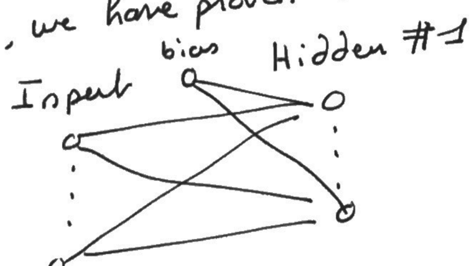
The architechre of the FNN is now :



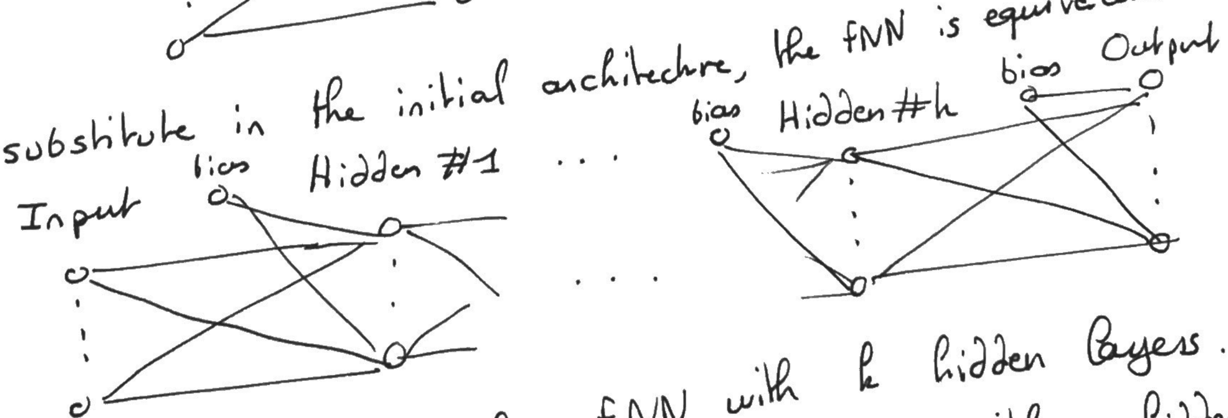
Let's consider the input layer, hidden #0 layer and hidden #1 layer in a vacuum :.



If we consider the Hidden #1 as an 'output' layer, this is exactly the case $k = 1$, we have proven than it is equivalent to :



If we substitute in the initial architechre, the FNN is equivalent to :



which go back to the case of a FNN with k hidden layers. As we assumed in out induction, it is equivalent to a FNN with no hidden layers.

So we have proven that any FNN with linear activation function is equivalent to a FNN with no hidden layer