Machine Learning II: Deep Learning

Jesús Fernández-Villaverde

University of Pennsylvania

December 27, 2016

Deep learning

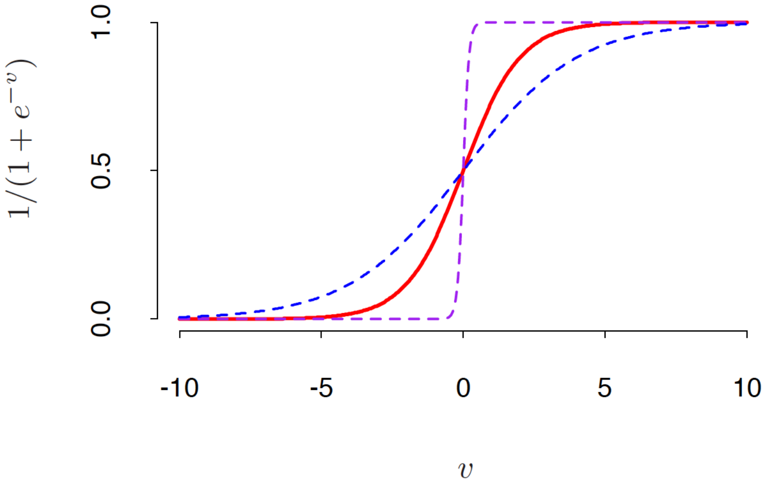
- Some of the most active research right now.
- Deal with very abstract set-ups.
- AlphaGo vs. Lee Sedol in March 2016.
- Multilayer neural networks (combinations of generalized linear models).
- Backpropagation through gradient descent.

Neural networks

- Non-linear statistical model.
- Much hype around them and over-emphasis of biological interpretation.
- We will follow a much sober formal treatment.
- In particular, we will highlight connections with econometrics.
- We will start describing the simplest possible neural network.

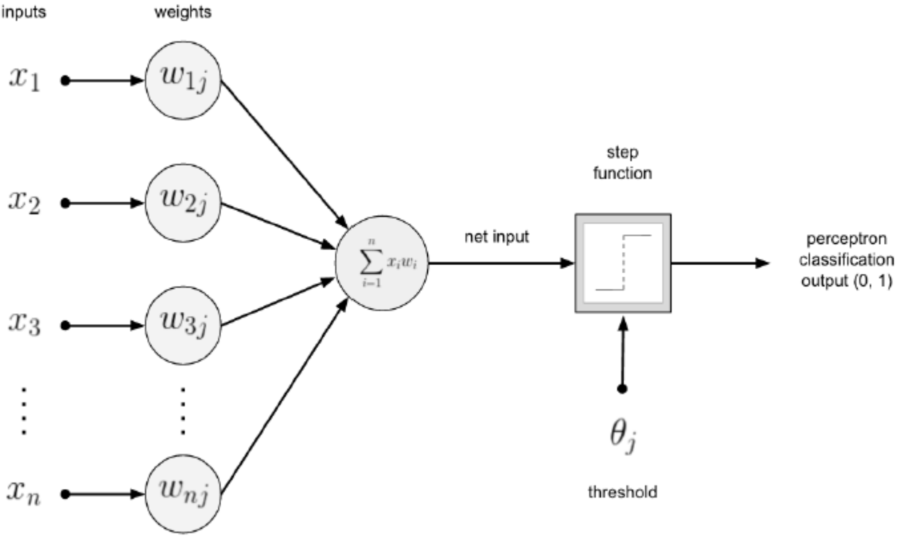
Structure I

- $N+1$ observables: x_0 (a constant), x_1, x_2,...,x_N.
- Coefficients (or weights): θ_0, θ_1, θ_2,...,θ_N.
- Linear combination:
$$z = \sum_{i=0}^{N} \theta_i x_i$$
- Activation function:
$$y = g\left(s\left(z - z_0\right)\right)$$
where $g\left(\cdot\right)$ is often a sigmoidal function:
$$g\left(v\right) = \frac{1}{1+e^{-v}}$$


1.0
0.5
0.0
$1/(1+e^{-v})$
-10
-5
0
5
10
v

Structure II

- s controls the activation rate (the higher s, the harder the activation).
- z_0 controls the activation threshold.
- Potential Identification problem with θ_i's.
- But in practice θ_i's do not have a structural interpretation and it is convenient to control activation rate and threshold separately.
- Other activation functions are possible.
- Closely resembles single index models in econometrics.


inputs
weights
x_1
x_2
x_3
x_n
w_{1j}
w_{2j}
w_{3j}
w_{nj}
$\sum_{i=1}^{n} x_i w_i$
net input
step
function
perceptron
classification
output (0, 1)
θ_j
threshold

A hidden layer

- A single hidden layer back-propagation network.
- We build M linear combinations:

$$z_{1,m} = \sum_{i=0}^{N} \theta_{1,i}^{m} x_i, \text{ for } m = 1, ..., M$$

- Then, we can recombine the z_m

$$z_2 = \sum_{j=0}^{M} \theta_{2,j}^{m} z_{1.j}$$

  and use another activation function:

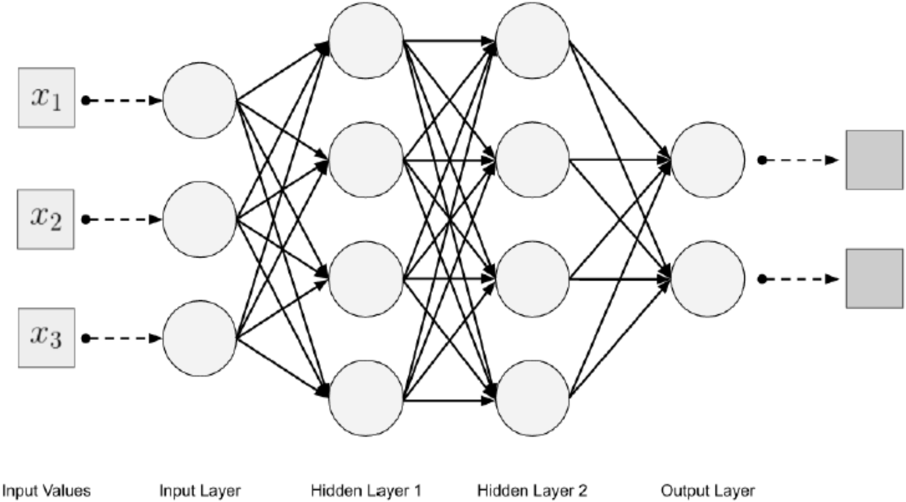
$$y = g_2 \left(s \left(z_2 - z_{2,0} \right) \right)$$

  (either the same or a different one as before).

- Why do we want to introduce hidden layers?

Multiple layers

- The hidden layers can be multiplied without limit.
- We can also add different outputs.


x_1
x_2
x_3
Input Values
Input Layer
Hidden Layer 1
Hidden Layer 2
Output Layer