Introduction to neural networks

Deep Neural Networks

Juan Irving Vásquez
jivg.org

Centro de Innovación y Desarrollo Tecnológico en Cómputo
Instituto Politécnico Nacional

October 2, 2023©

Contents

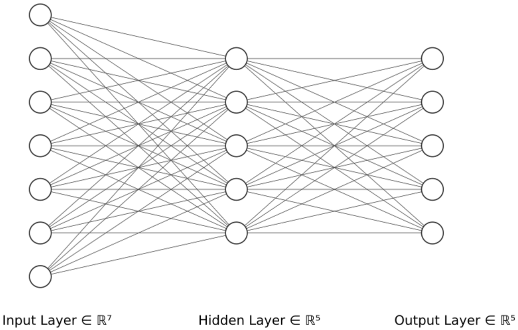
Roadmap

1. Deep fully connected networks

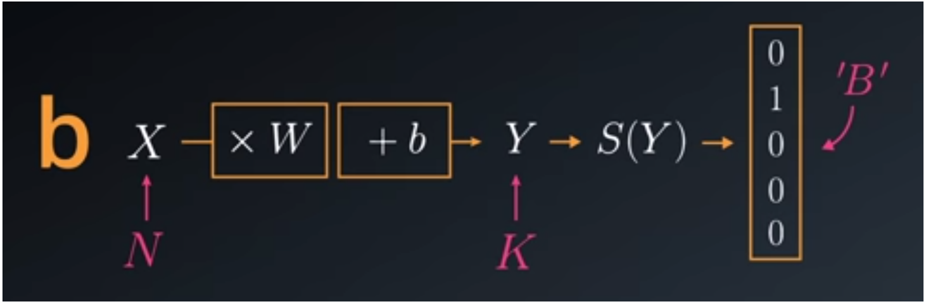
Linear model

- Multinomial regression model

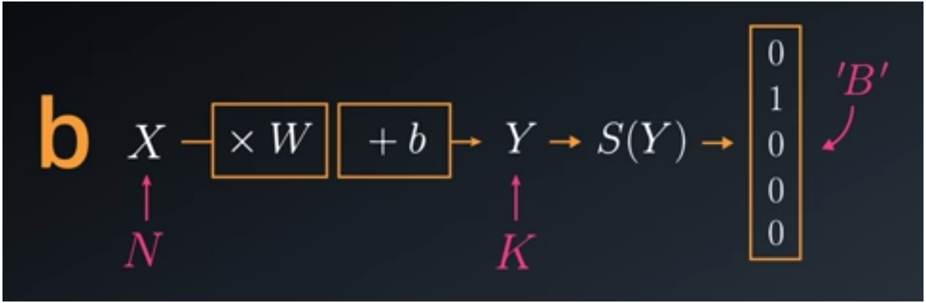
$$X \rightarrow XW + b \rightarrow f(h) \rightarrow Y \rightarrow S(Y) \rightarrow P \tag{1}$$



Linear model complexity



Linear model complexity



- Number of parameters in each layer:

$$f(n, k) = (n + 1)k \tag{2}$$

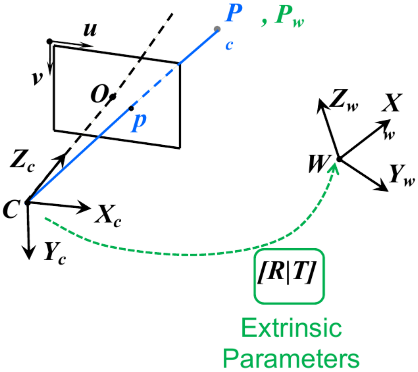
where n is the number of inputs and k the number of outputs.

Linear models advantages

- Efficiency
- Hardware available

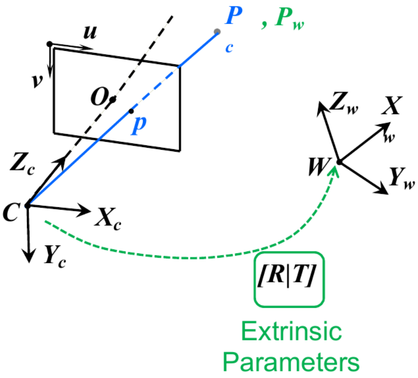
GPU

Why are so effective?



GPU

Why are so effective?



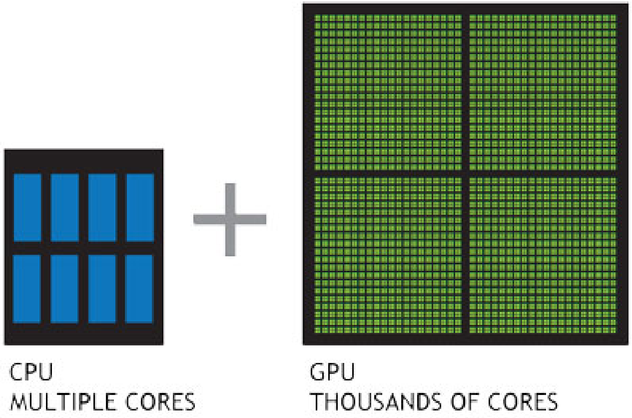
Projection Matrix

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boxed{K\left[R|T\right]} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

GPU

Why are so effective?



GPU

Linear models disadvantages

- Linear models are restricted to linear combinations!

$$Y = x_1 + x_2 \tag{3}$$

Linear models disadvantages

- Linear models are restricted to linear combinations!

$$Y = x_1 + x_2 \tag{3}$$

- How do we improve them?

Linear models disadvantages

- Linear models are restricted to linear combinations!

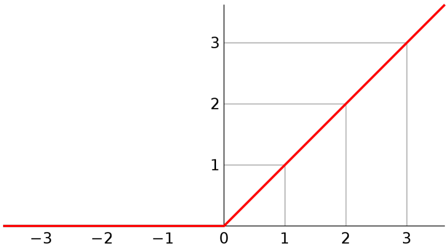
$$Y = x_1 + x_2 \tag{3}$$

- How do we improve them?
  Activation functions!

ReLU

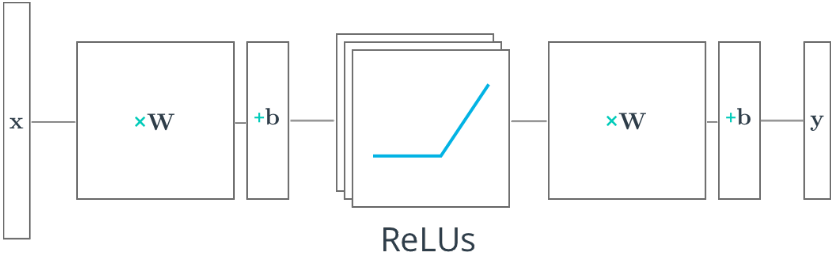
- Rectified linear unit

$$y = max(0, x)$$



- Derivative?

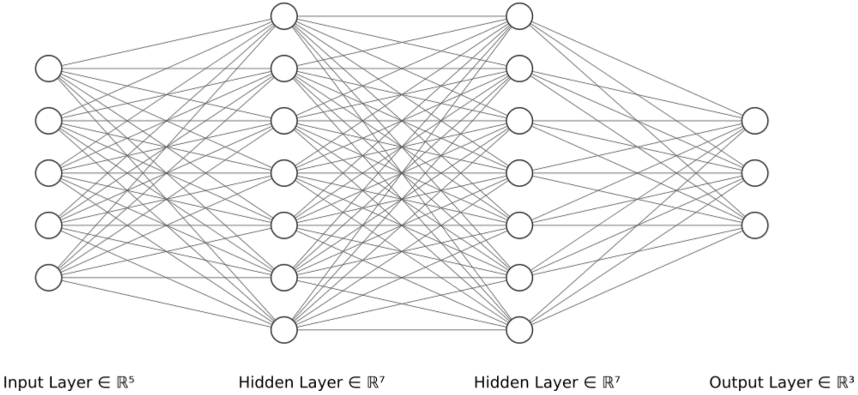
Multilayer Neural Network



- Layer 1: the set of W and b applied to X and passed through ReLUs. The output is fed to the next one, but is not observable outside the network (hidden layer).
- Layer 2: the weights and biases applied to these intermediate outputs, followed by $S(y)$.

Excercises

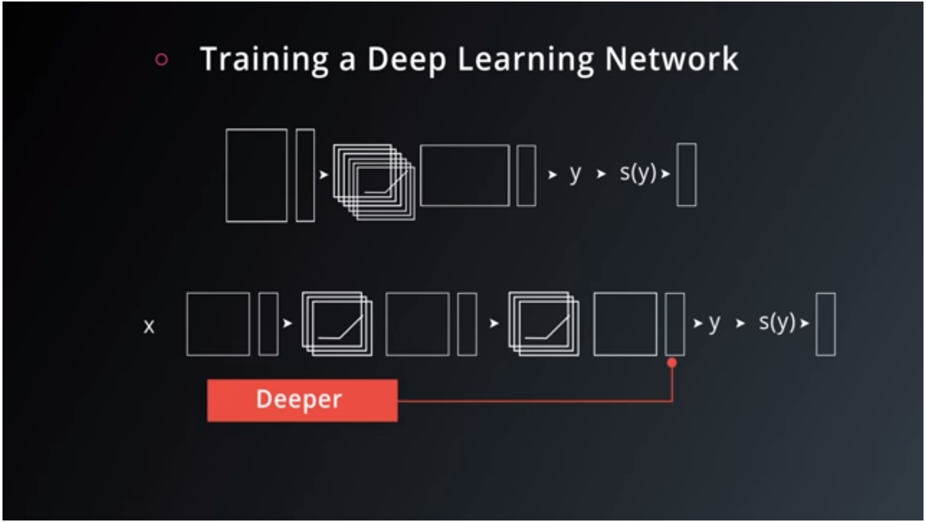
- Determine the amount of weights required in the following network:



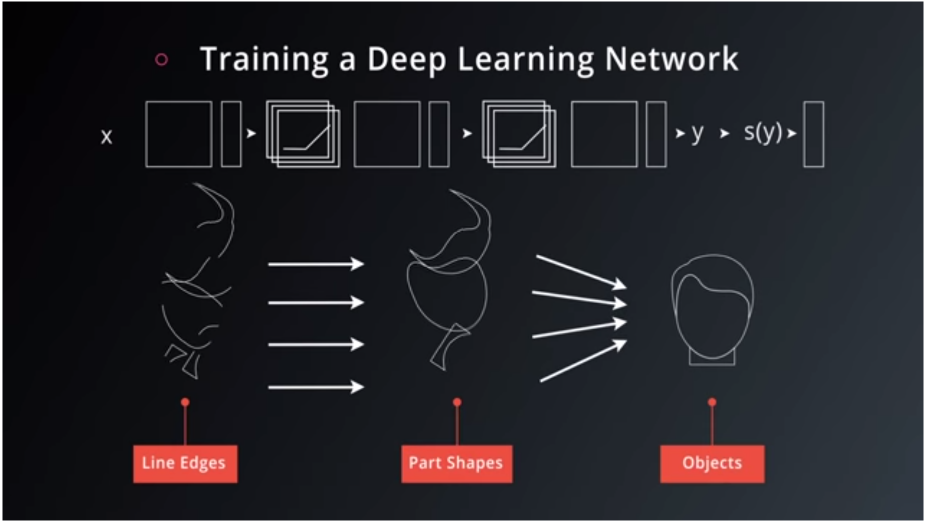
- The chain rule

$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x} \tag{4}$$

Training a DNN



Training a DNN



References

dfellow Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.

Sucar15 Sucar, Luis Enrique. Probabilistic Graphical Models: Principles and Applications. Springer, 2015.

Smola Alex Smola and S.V.N. Vishwanathan, Introduction to Machine Learning, Cambirdge University Press.

Udacity Self Driving Car Nanodegree

vastava Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.