Supervised learning III

Liangjie Cao

1 May 2018

The paper says since the 1960s we have known that linear classifiers can only carve their input space into very simple regions, namely half-spaces separated by a hyperplane.Actually image and speech recognition require the input–output function to be insensitive to irrelevant variations of the input, such as variations in position, orientation or illumination of an object, or variations in the pitch or accent of speech. Meanwhile it is very sensitive to particular minute variations. [1]

If we take this to the pixel level, images of two Samoyeds in different poses and in different environments may be very different from each other, though two images of a Samoyed and a wolf in the same position and on similar backgrounds may be very similar to each other. This is linear classifier, or called any other 'shallow' classifier operating on Figure 1
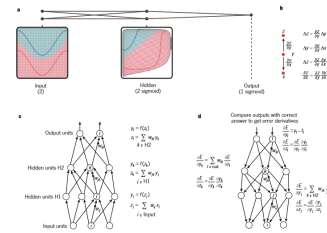

Figure 1: **Multilayer neural networks and backpropagation**

The counting process can be seen in c. They first compute the total input z to each unit, which is a weighted sum of the outputs of the units in the layer below. Then using a non-linear function f(.), f(z) = (exp(z) − exp(−z))/(exp(z) + exp(−z)) and logistic function logistic, f(z) = 1/(1 + exp(−z)). To be honest, I haven't understood these two function in Table 1, maybe this is just an output and follow the rule. I will continue learning.

hyberbolic tangent	f(z) = (exp(z) − exp(−z))/(exp(z) + exp(−z))
logistic function logistic	f(z) = 1/(1 + exp(−z))

Table 1: **Classifition of output funtion**

References

[1] Yoshua Bengio& Geoffrey Hinton Yann LeCun. Deep learning. *Nature*, 521(28):9, 2015.