4 Chapter 4

We now move to briefly discuss Theorem 2.

Theorem 4.2. *It considers a function $f \in W_m^{2,n}$, the class of all compositional functions f of n variables with a binary tree architecture and constituent functions h in W_m^2 and a deep network with a compositional architecture.*
The activation function $\sigma : \mathbb{R} \to \mathbb{R}$ in this context is also infinitely differentiable and not a polynomial. The complexity of the network necessary to provide an approximation with accuracy at least ε is

$$N = O((n-1)\varepsilon^{-2/m})$$