Solution to Homework 24

Question. Explain why, in our derivation of the ReLU activation function, we used the possibility to change the measuring unit, but not the possibility to change the starting point.

Answer. If we require invariance of the activation function $s(x)$ with respect to the change of the starting point – which means a transformation of each value x into $x + x_0$ – this would be means that if we have $y = s(x)$, then we should have
$$y + x_0 = s(x + x_0)$$
for all x and x_0. Since $y = f(x)$, this means that for all x and x_0, we should have
$$s(x) + x_0 = s(x + x_0).$$

In particular, for $x = 0$, we should conclude that every x_0, we should have $s(x_0) = s(0) + x_0$. Thus, we will get a linear activation function – but the main purpose of activation functions is to deal with non-linear situations.