- Rectified linear unit (ReLU)
- Leaky ReLU
- Maxout

Let's briefly examine them.

Sigmoid

The sigmoid function illustrated in Figure 29-7 is a non-linear function that brings (or squashes) the activations to fall within a range of 0 and 1. This brings large negative and positive numbers to 0 and 1, respectively. The neurons typically begin firing when the function output is above a threshold of 0.5.

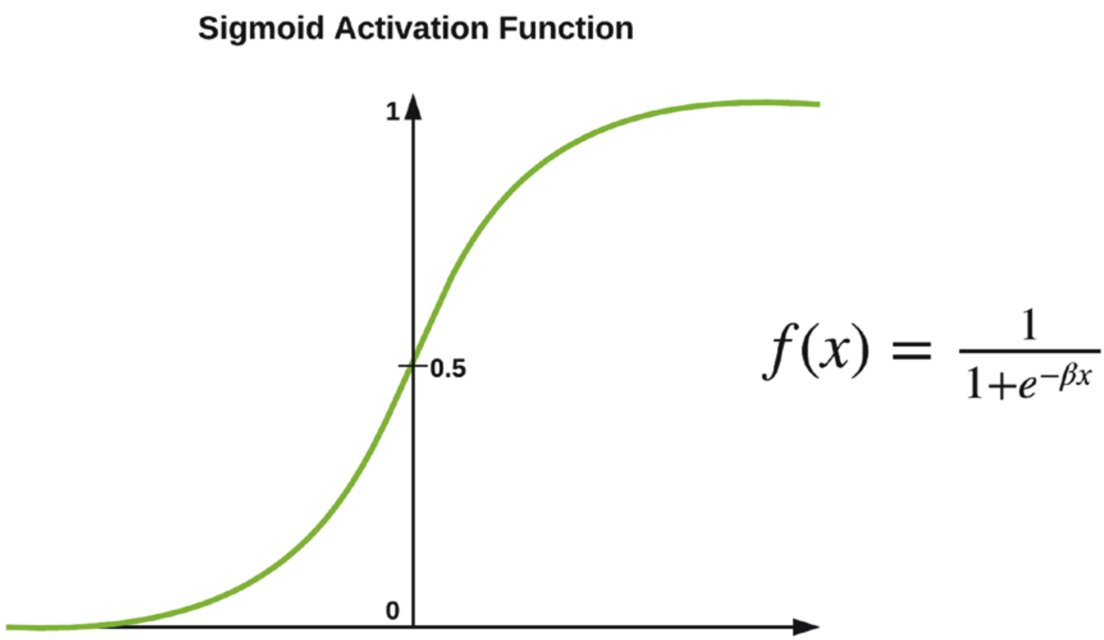

Figure 29-7. *Sigmoid activation function*

However, a significant drawback of the sigmoid function is its susceptibility to a phenomenon called exploding and vanishing gradients. In the process of optimizing the weights of the network during backpropagation, the gradients can become disproportionately small or large with their activations concentrated at either 0 or 1.

When this happens, we say that the gradients have saturated. Hence, further multiplication via backpropagation causes the gradient to either vanish or explode; and as a result, the affected neurons become dead and transfer no information across the network, thus negatively affecting training.

Another drawback is that the outputs of the function are not zero-centered. As a consequence, during backpropagation, the gradients can either become all positive or all negative. This has a negative effect in minimizing the function objective (i.e., the cost function).

Hyperbolic Tangent (tanh)

The hyperbolic tangent illustrated in Figure 29-8 improves on the sigmoid function by bordering its output within a range of –1 and 1. So, while it still suffers from the exploding and vanishing gradient problem, its outputs are now zero-centered. From the formula, the reader will observe that tanh is merely a scaled sigmoid function.

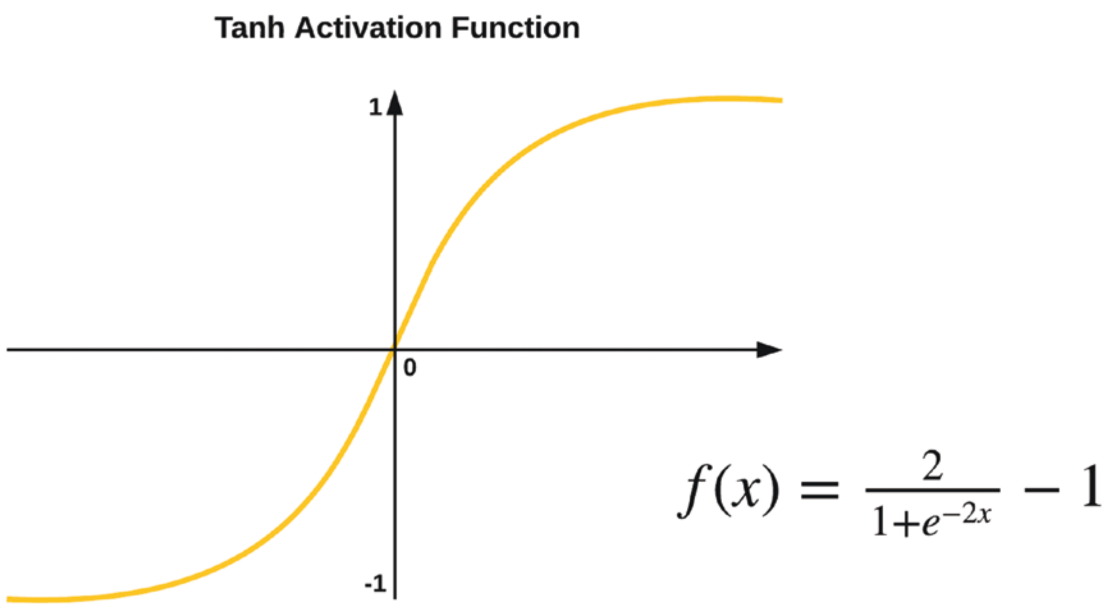

Figure 29-8. *The hyperbolic tangent activation function*