Activation Functions

$$f(x) = sigmoid(x)$$

$$z_2 = XW_1$$

$$a_2 = f(z_2)$$

$$z_3 = a_2W_2$$

$$\hat{y} = a_3 = f(z_3)$$

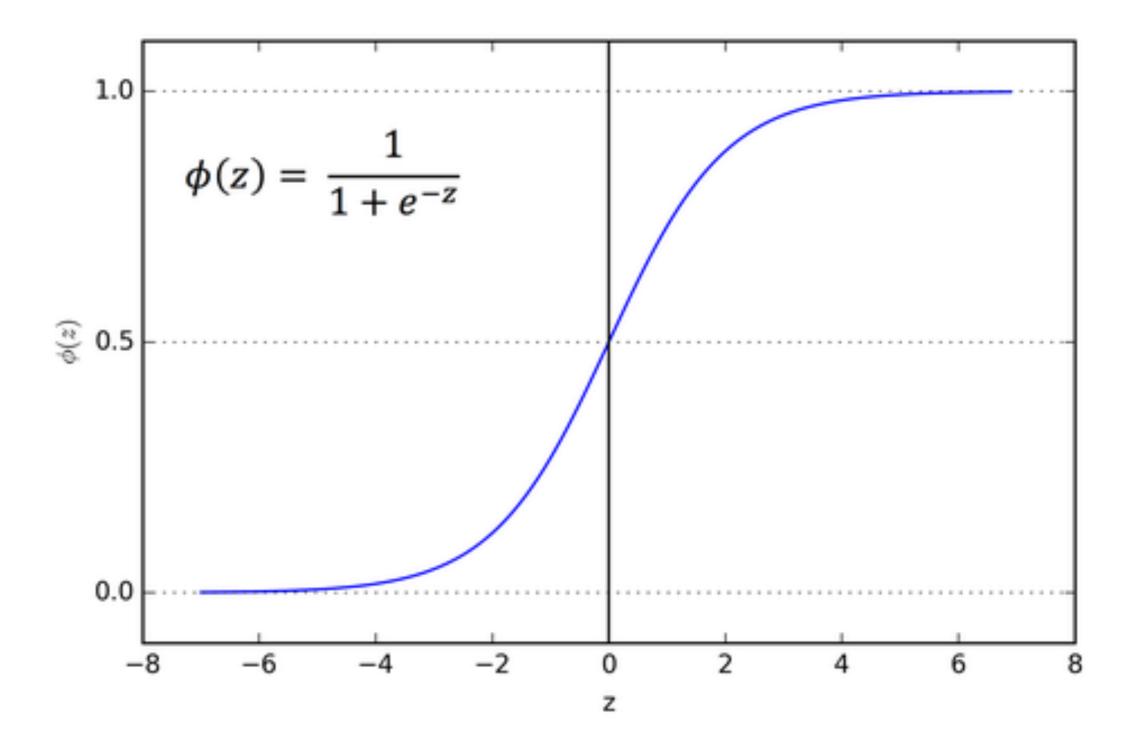


Image Credit: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

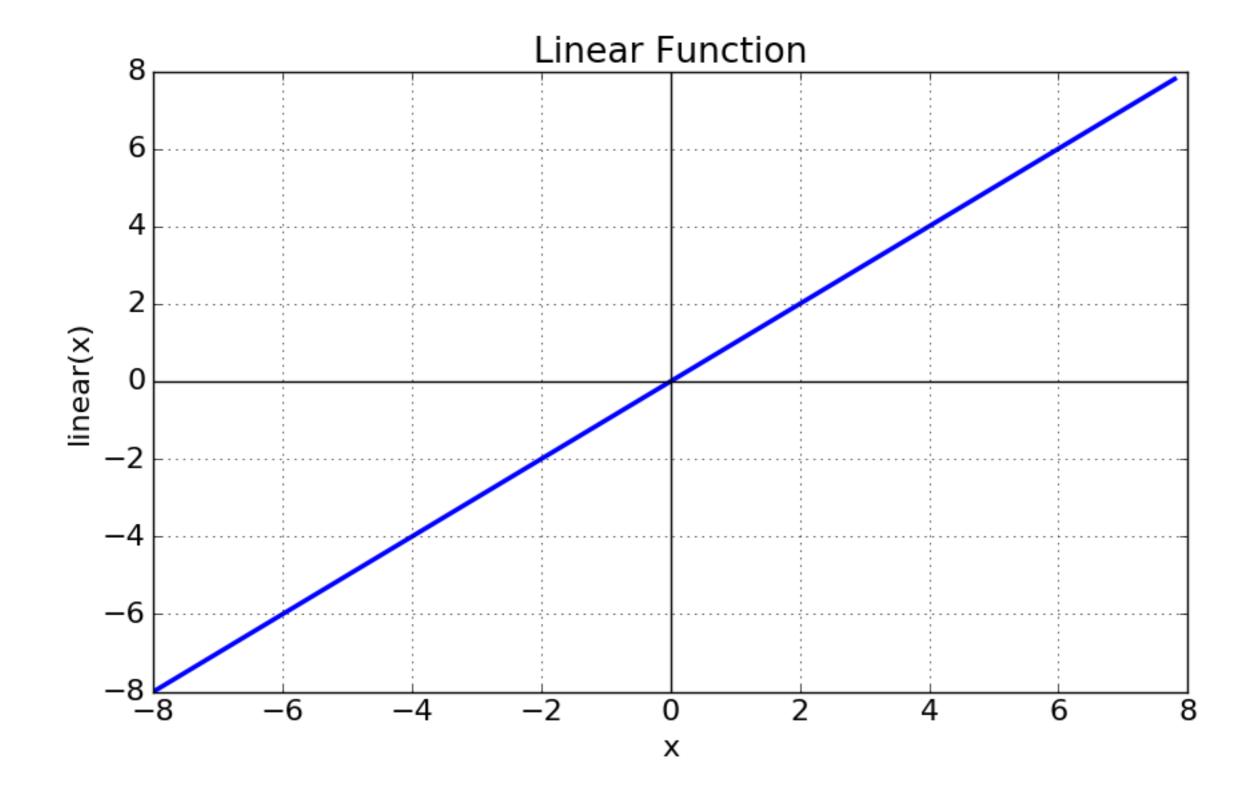


Image Credit: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

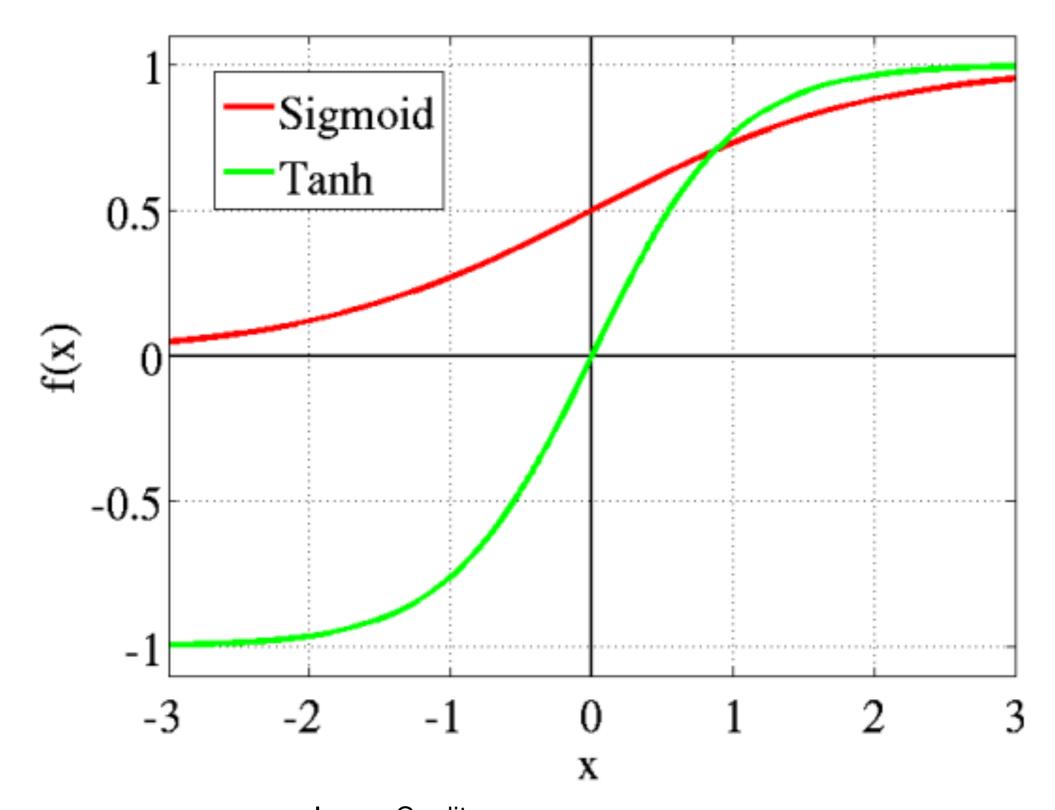


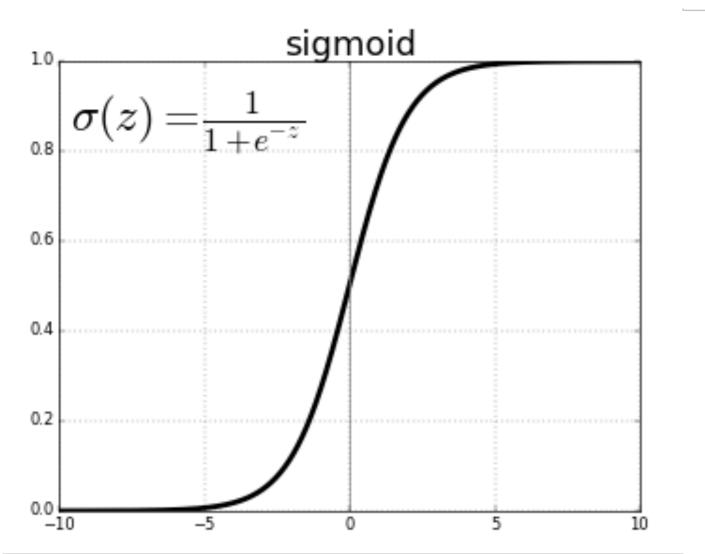
Image Credit: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

$$\sigma: \mathbb{R}^K o [0,1]^K$$
 $\sigma(\mathbf{z})_j = rac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad ext{for } j=1,..., extit{ extit{K}}.$

$$egin{align} \sigma: \mathbb{R}^K &
ightarrow [0,1]^K \ \sigma(\mathbf{z})_j &= rac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} & ext{for } j=1,..., extit{ extit{ extit{K}}}. \end{aligned}$$

$$egin{aligned} \sigma: \mathbb{R}^K &
ightarrow [0,1]^K \ \sigma(\mathbf{z})_j &= rac{e^{(z_j)}}{\sum_{k=1}^K e^{z_k}} & ext{for } j=1,...,K. \end{aligned}$$

$$egin{aligned} \sigma: \mathbb{R}^K &
ightarrow [0,1]^K \ \sigma(\mathbf{z})_j &= rac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \end{aligned} ext{ for } j=1,..., extit{ K}.$$



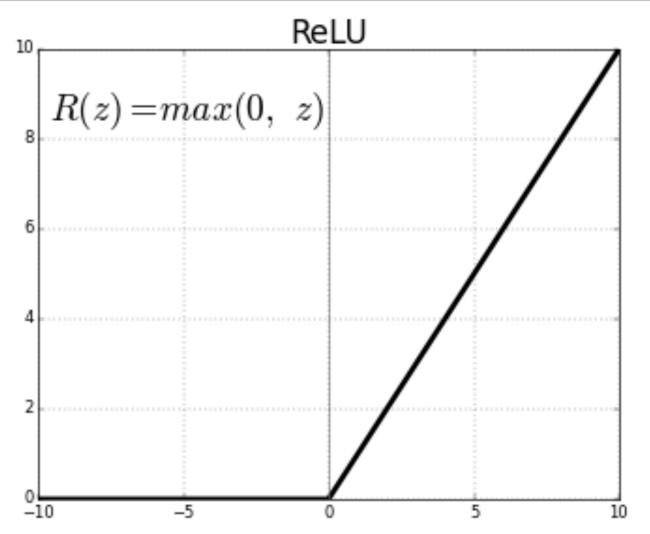
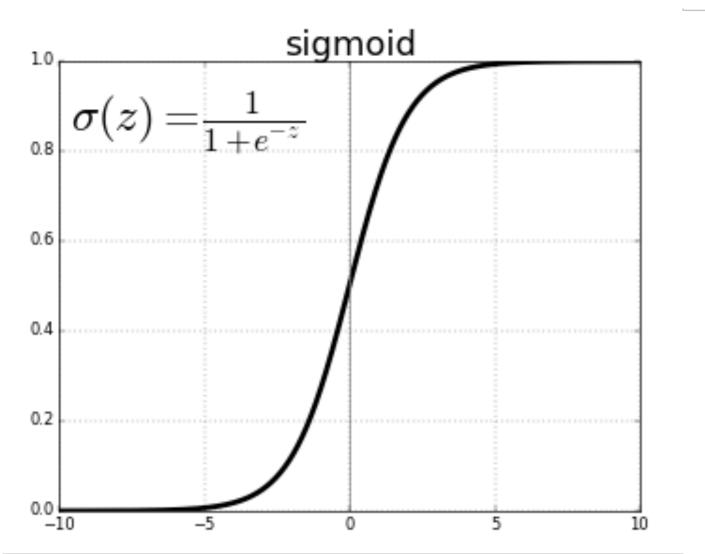


Image Credit:

https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6



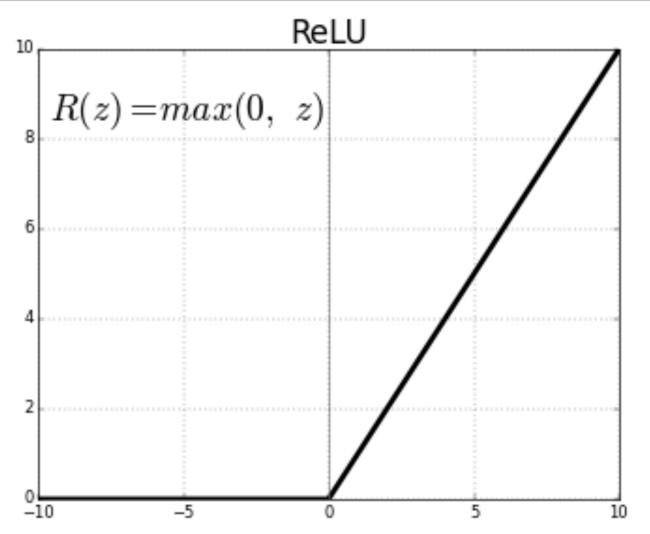


Image Credit:

https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

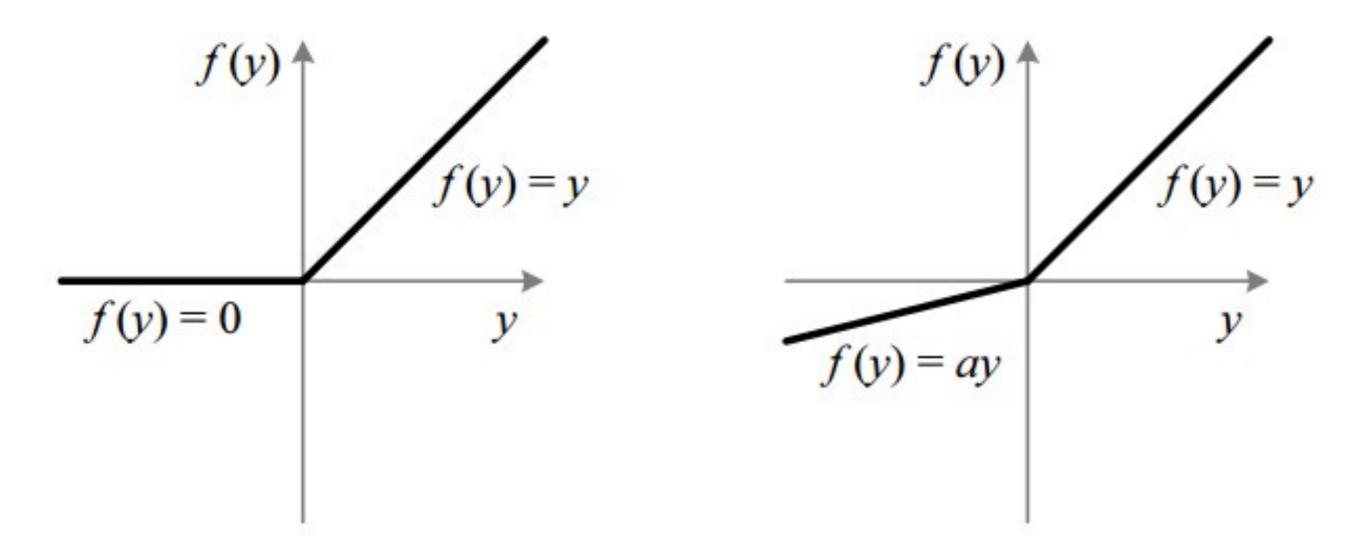


Image Credit:

https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

Summary

Bias-Variance Tradeoff