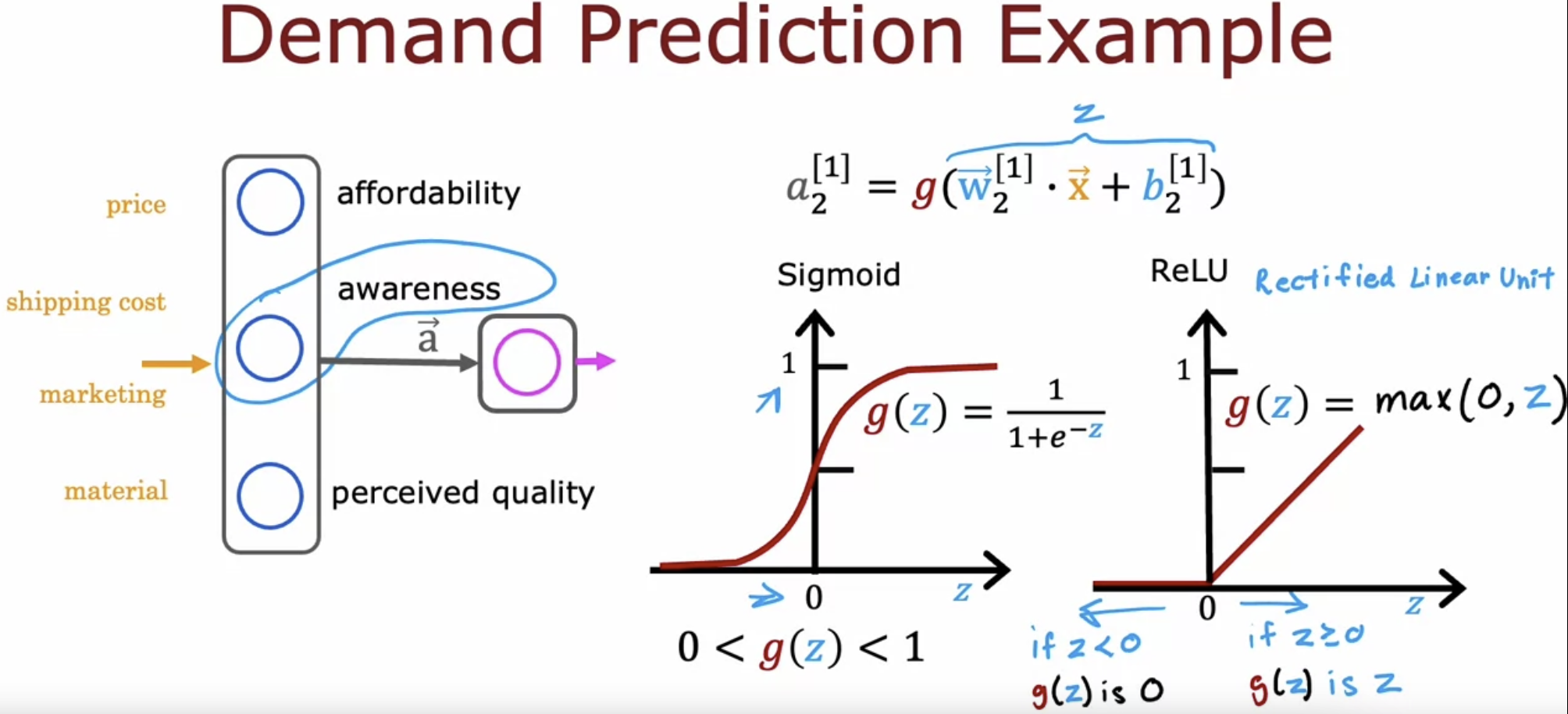
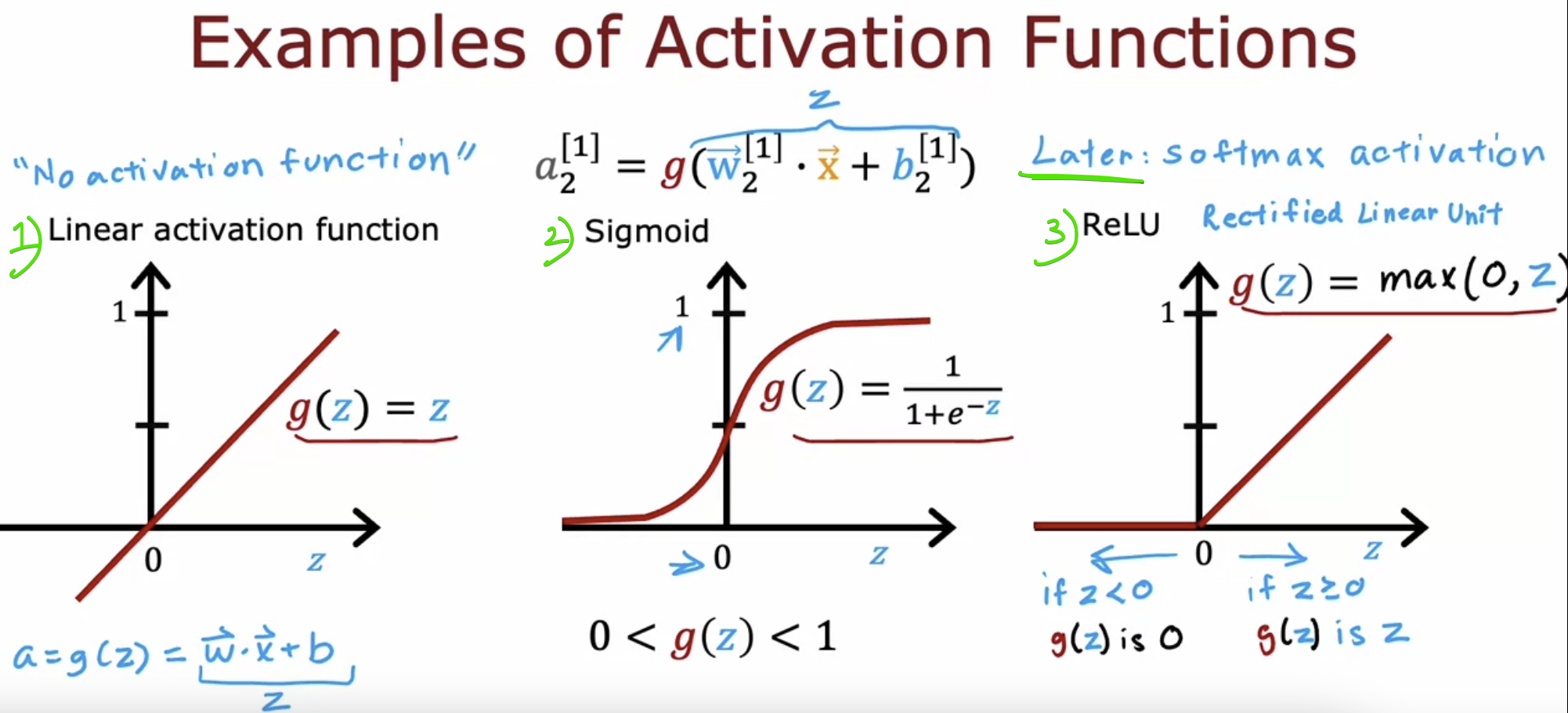
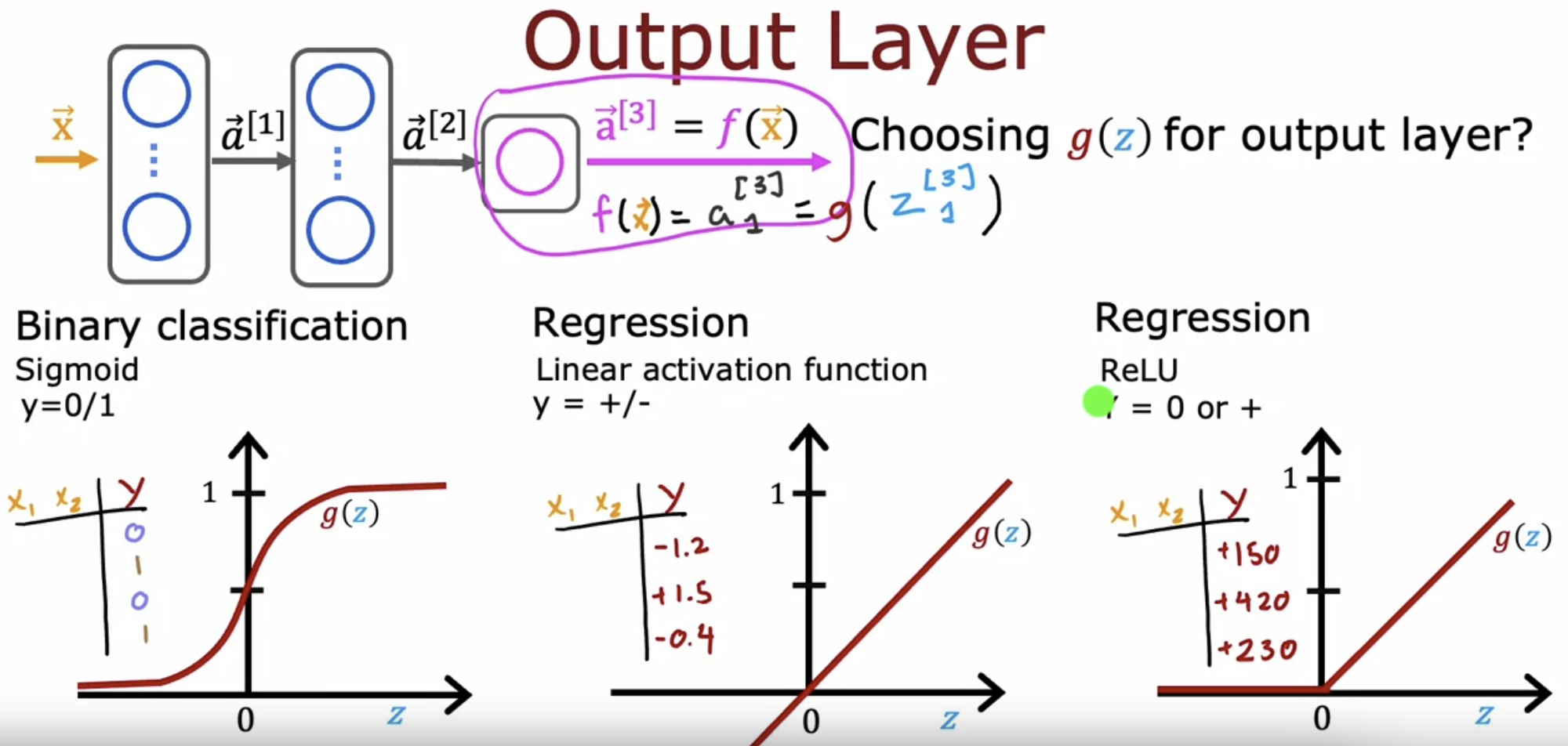
# **Alternatives to the sigmoid activation**





# **Choosing the Activation Function**

If we use gradient descent to train the neural network, a function that is flat in lot of places gradient descent would be really slow.



A picture containing diagram

Description automatically generated

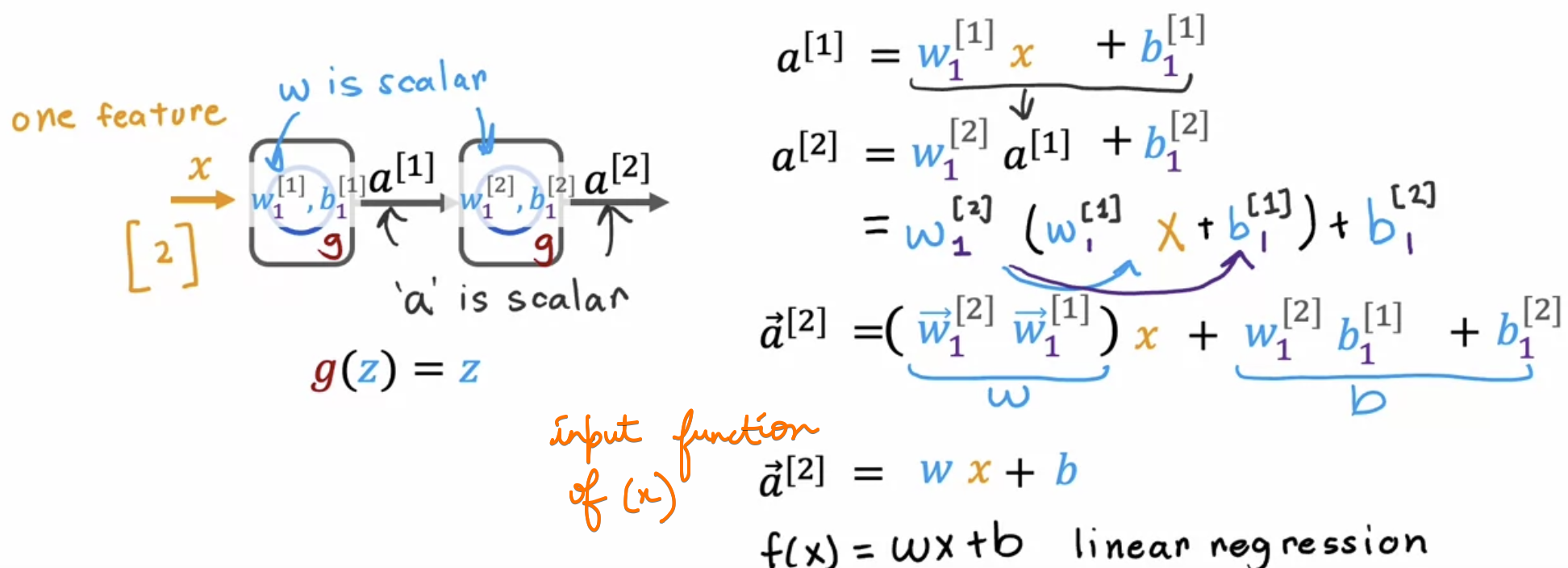
Text

Description automatically generated

# **Why do we need activation functions ??**

Diagram

Description automatically generated



* Rather than using a neural network with one hidden layer and one output layer,   
  we might as well have just used a linear regression model.
* If you're familiar with linear algebra, this result comes from the fact that   
  a linear function of a linear function is itself a linear function.
* This is why having multiple layers in a neural network doesn't let the neural network compute any more complex features or learn anything more complex than just a linear function.

Diagram, text

Description automatically generated

Chart, bubble chart

Description automatically generated

# **References**

<https://www.v7labs.com/blog/neural-networks-activation-functions#h2>