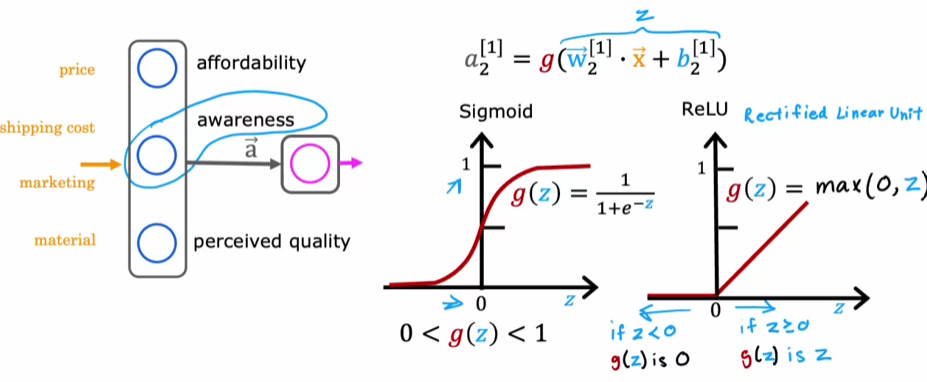
**ACTIVATION FUNCTIONS**

**ALTERNATIVES TO THE SIGMOID ACTIVATION**

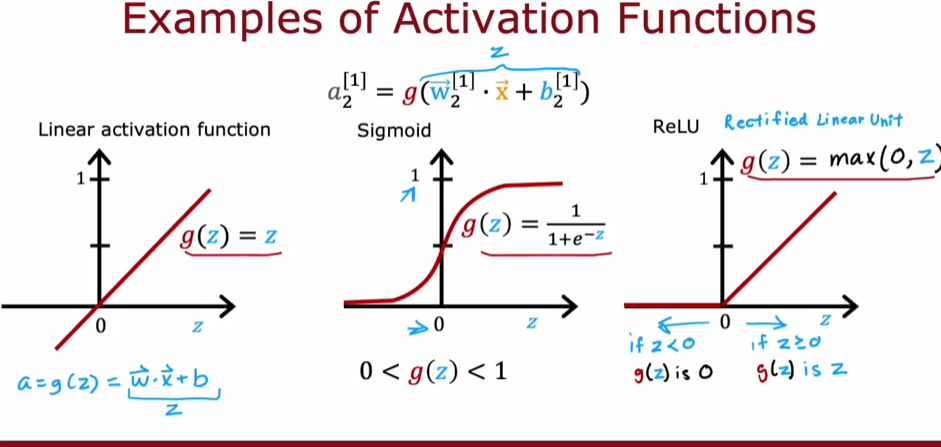
**Understanding Activation Functions**

* **The sigmoid activation function has been commonly used in hidden and output layers, but exploring other functions can significantly improve neural network capabilities.**
* **The ReLU (Rectified Linear Unit) activation function is introduced as a powerful alternative, allowing for non-negative values and enabling the model to capture more complex patterns.**

****

**Common Activation Functions**

* **The sigmoid function outputs values between 0 and 1, making it suitable for binary classification tasks.**
* **The ReLU function, defined as g(z) = max(0, z), allows for values greater than or equal to zero, which can lead to better performance in deep learning models.**
* **The linear activation function, g(z) = z, is sometimes referred to as having no activation function, as it does not transform the input.**

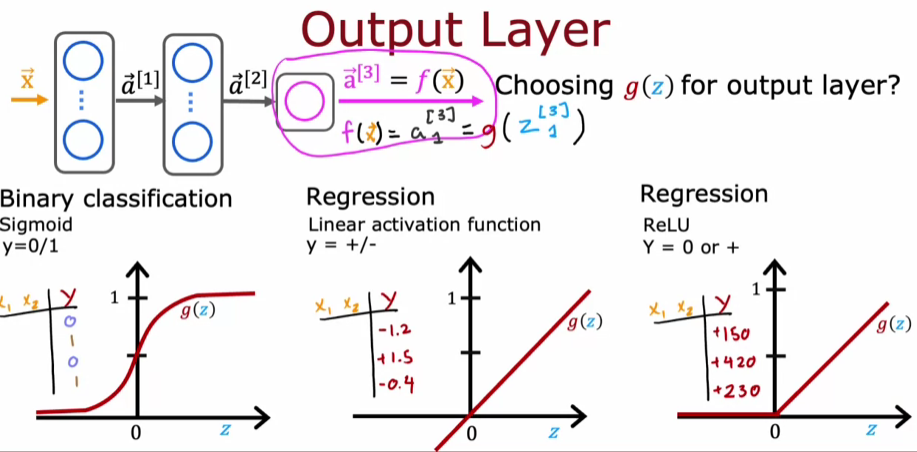
****

**CHOOSING ACTIVATION FUNCTIONS**

**The choice of activation function can impact the performance of the neural network, and understanding the characteristics of each function is crucial for effective model building.**

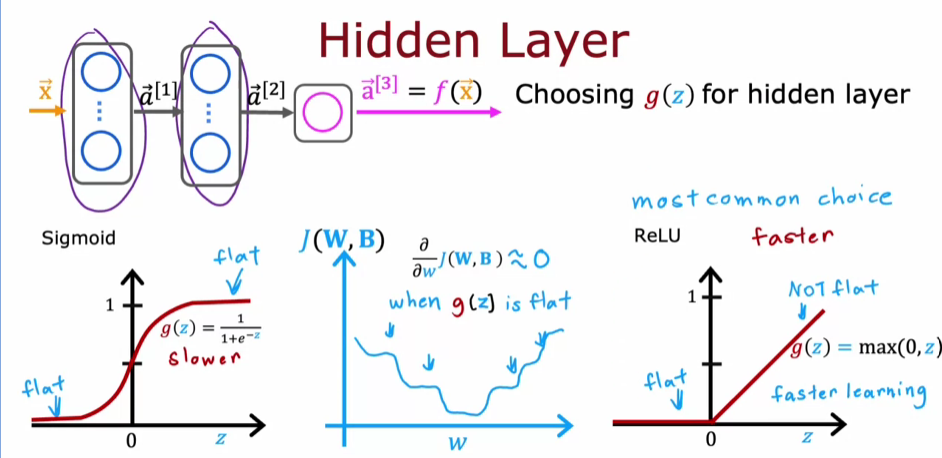
**Choosing activation functions for the output layer**

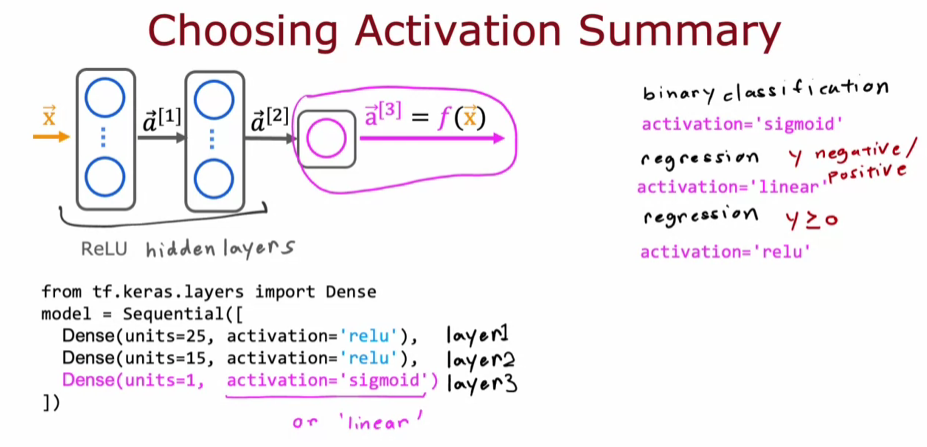
* **For binary classification problems (where the target label is either 0 or 1), the sigmoid activation function is recommended, as it predicts the probability of the label being 1.**
* **For regression problems, if the target can take on both positive and negative values, a linear activation function is suitable; if the target can only be non-negative, the ReLU activation function is the best choice.**

****

**Choosing activation functions for hidden layers**

* **The ReLU activation function is the most common choice for hidden layers due to its efficiency and faster learning capabilities compared to the sigmoid function.**
* **The ReLU function is preferred because it only goes flat in one region, reducing the chances of slow learning during gradient descent.**

****

****

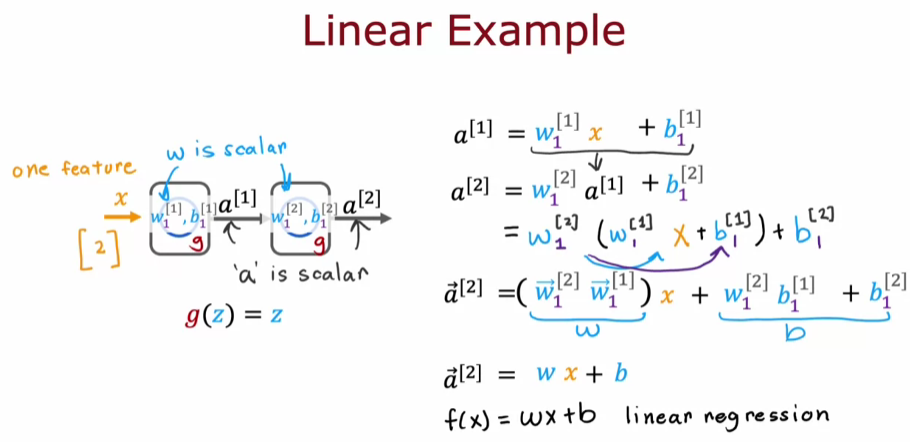
**Reinforcing learning and exploring other options**

* **While ReLU is widely used, other activation functions like LeakyReLU or tanh may be explored for specific cases, but the ones discussed are generally sufficient for most applications.**
* **Understanding the role of activation functions is crucial, as they significantly impact the performance of neural networks. Engaging in practice labs can help reinforce these concepts.**

**WHY DO WE NEED ACTIVATION FUNCTIONS**

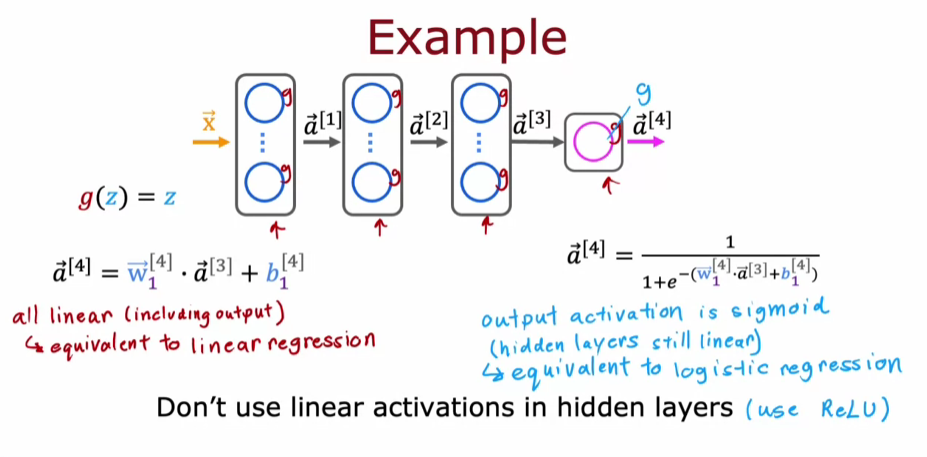
**Understanding Activation Functions**

* **Using a linear activation function throughout a neural network results in a model that behaves like linear regression, which cannot capture complex patterns in data.**
* **A neural network with multiple layers and linear activation functions does not provide any advantage over simpler models, as it can only compute linear functions.**

****

**The Role of Non-Linear Activation Functions**

* **To enable neural networks to learn complex features, it's essential to use non-linear activation functions, such as ReLU, especially in hidden layers.**
* **For example, if a neural network uses linear activation functions in hidden layers but a logistic activation function in the output layer, it behaves like logistic regression.**

****

**Key Takeaways**

* **Avoid using linear activation functions in hidden layers to ensure the neural network can learn complex relationships.**
* **Understanding the role of activation functions is crucial for building effective neural networks for various classification and regression tasks.**