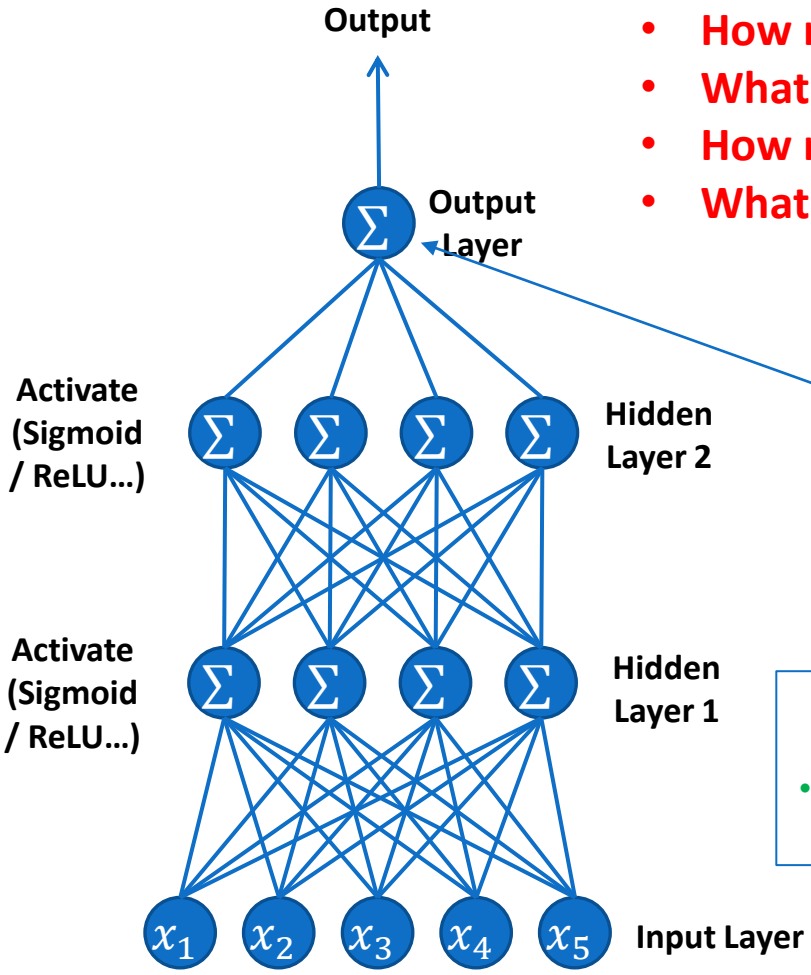
Review

MULTI-LAYER PERCEPTRON (MLP)

MLP to Deep Neural Network

- How many inputs?
- How many neurons in each hidden layer?
- How many hidden layers?
- What activation function to use for hidden layer neurons?
- How many neurons in output layer?
- What activation function to use for output layer?

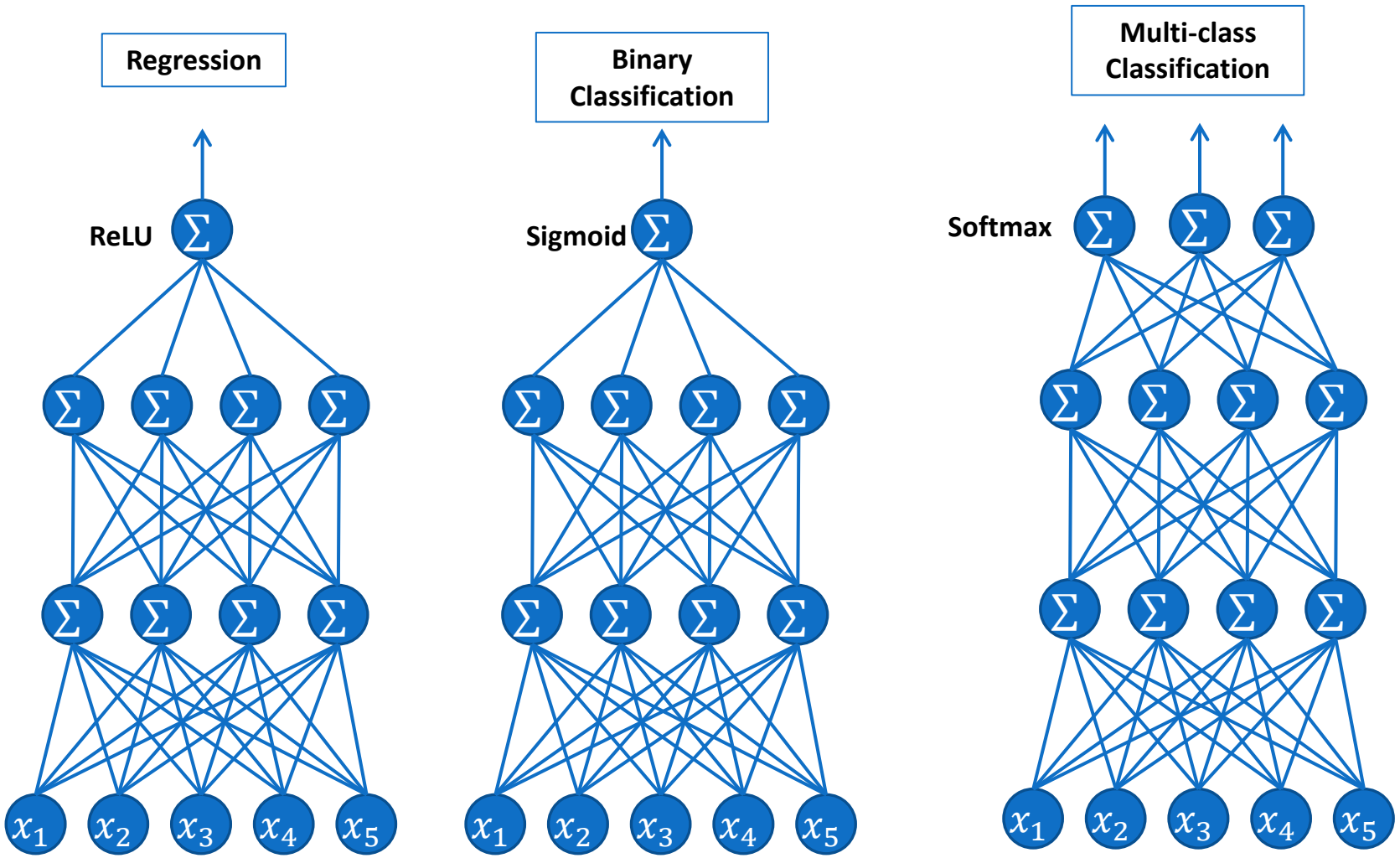


What activation function should be used at output layer?

It depends!!

- None if you want a regression output.
- ReLU if you want regression output but only positive values.
- Sigmoid if you want a probability $(between\ 0-1)$ for a class (i.e., binary classification.

Deep Neural Network For Different Tasks



Summary

- If your problem is a **regression** problem, you should use a **linear activation** function (i.e., multiply sum by 1 aka no activation).
  - **Regression:** One output node, linear activation.

- If your problem is a **classification** problem, then there are three main types of classification problems and each may use a different activation function.
  - If there are two mutually exclusive classes (binary classification).
    - **Binary Classification:** One output node, sigmoid activation.
  - If there are more than two mutually exclusive classes (multiclass classification).
    - **Multiclass Classification:** One output node per class, softmax activation.
  - If there are two or more mutually inclusive classes (multilabel classification),
    - **Multilabel Classification:** One output node per class, sigmoid activation.

What is multilabel classification?

Credit: https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/

Multilabel Classification

- The goal is to predict ***Color*** and ***Type*** of clothing.
- **How to achieve that?**

Credit: https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

Approach 1 – Separate Models

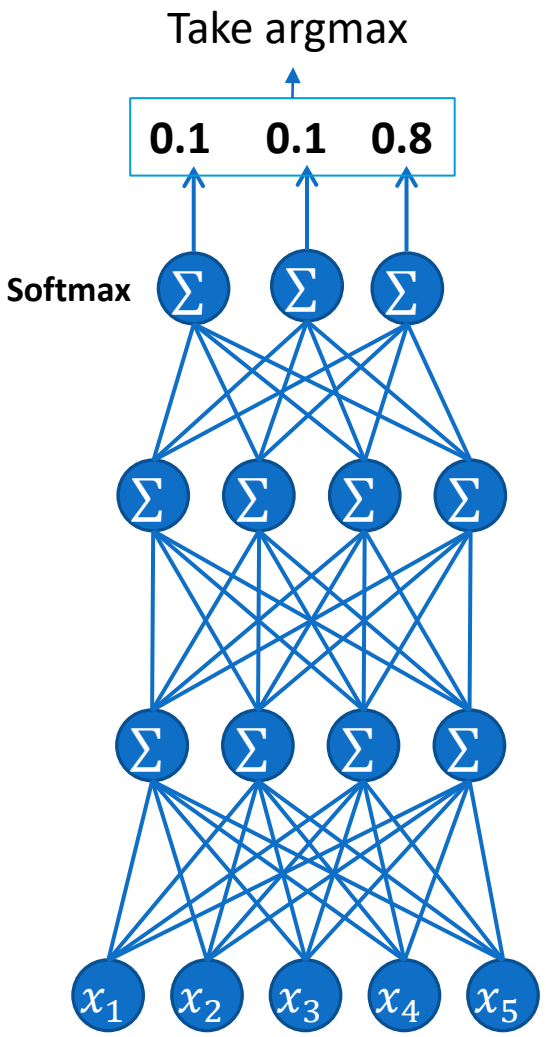
{ “Jeans”: 0, “Dress”: 1, “Shirt”: 2}

This input image belongs to class 2.

One hot encoded label: [0, 0, 1]

This input image belongs to class 1.

One hot encoded label: [0, 1, 0]



Credit: https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

Approach 1 – Separate Models

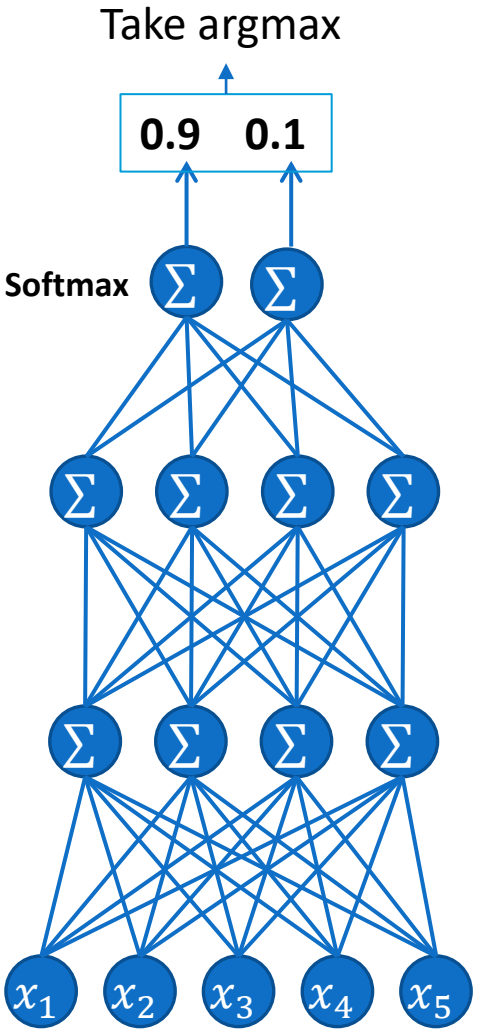
{ “Red”: 0, “Black”: 1}

This input image belongs to class 0.

One hot encoded label: [1, 0]

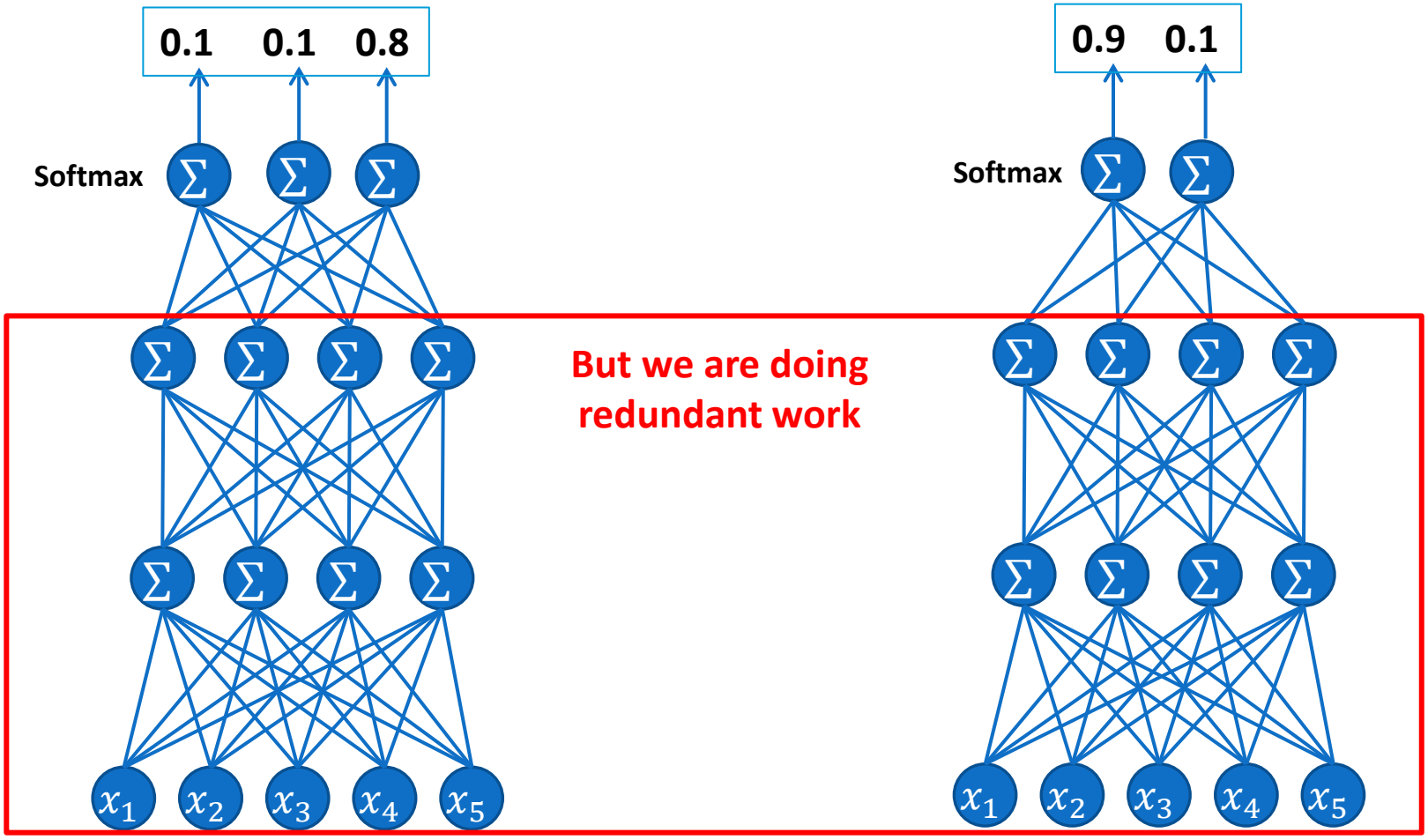
This input image belongs to class 1.

One hot encoded label: [0, 1]



Credit: https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

We have two separate models and we can take two predictions and present to user.



Credit: https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

Approach 2 – Multilabel Models

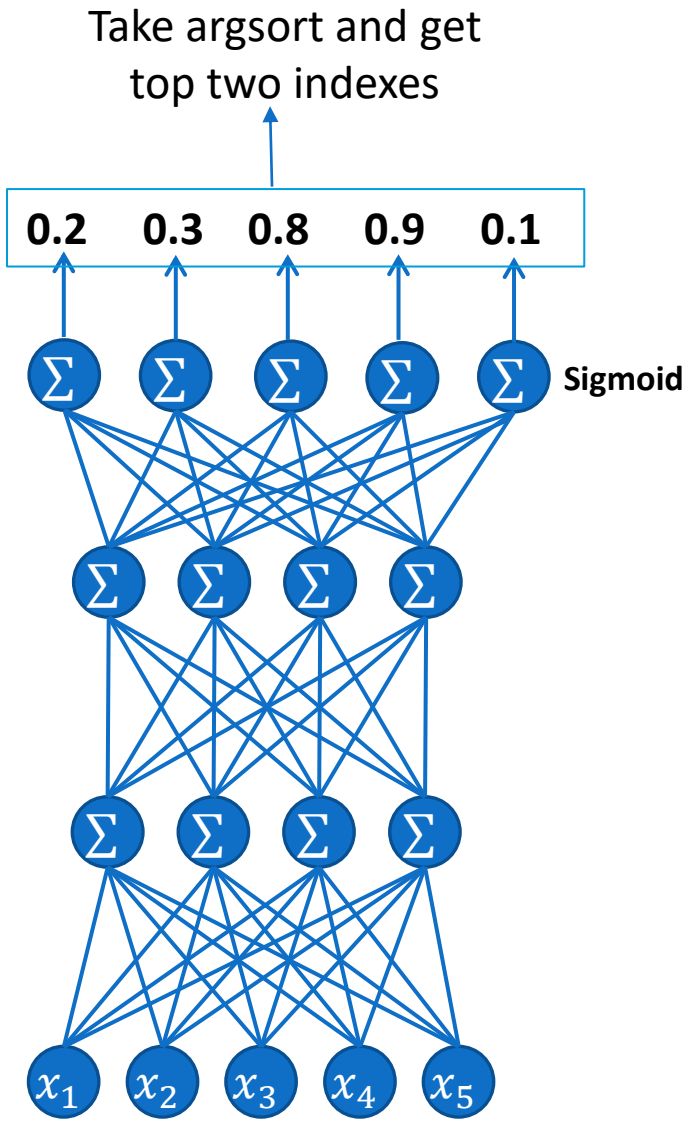
{ “Jeans”: 0, “Dress”: 1, “Shirt”: 2, “Red”:3, “Black”:4}

This input image belongs to class 2 and 3.

One hot encoded label: [0, 0, 1, 1, 0]

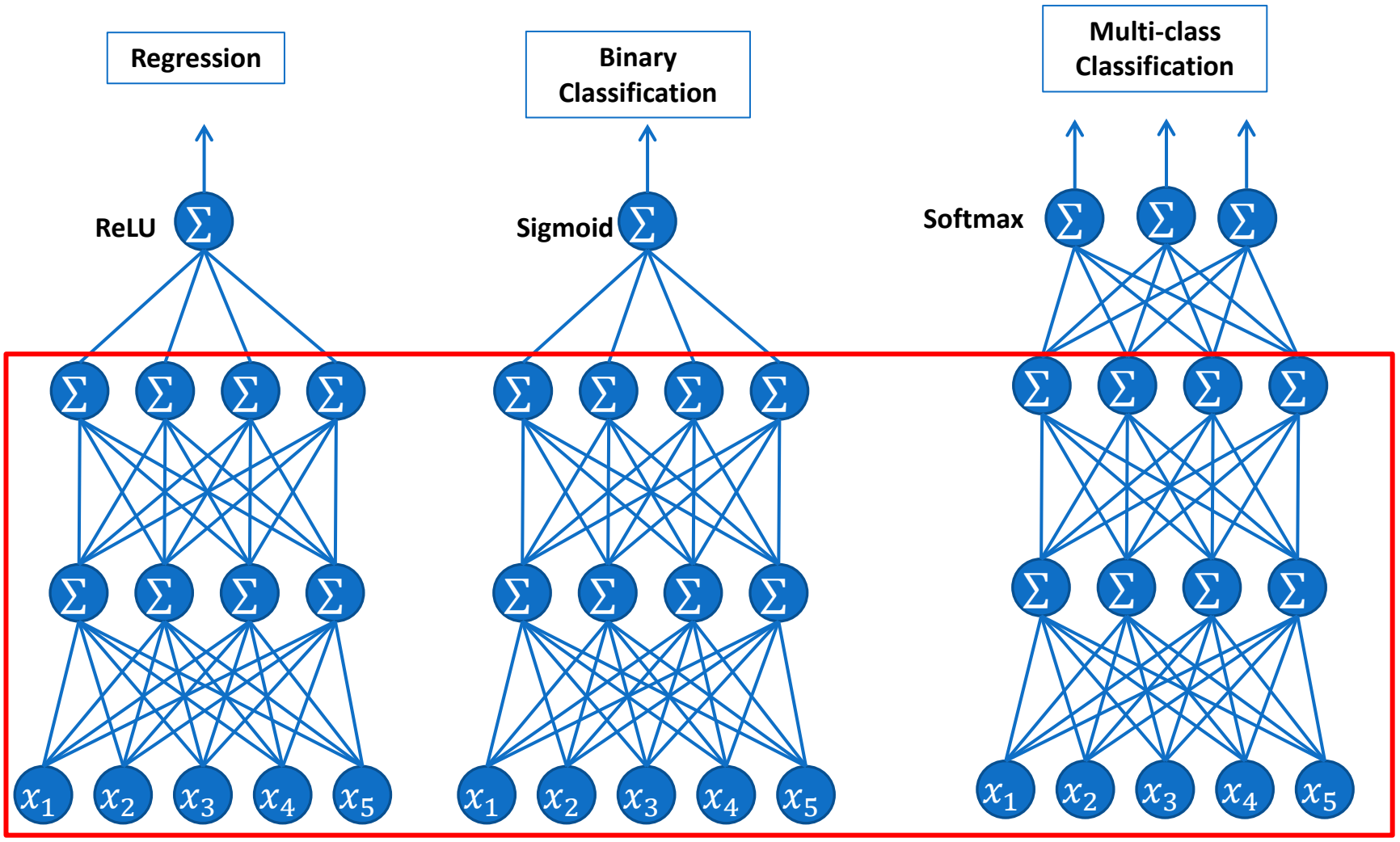
This input image belongs to class 1 and 4.

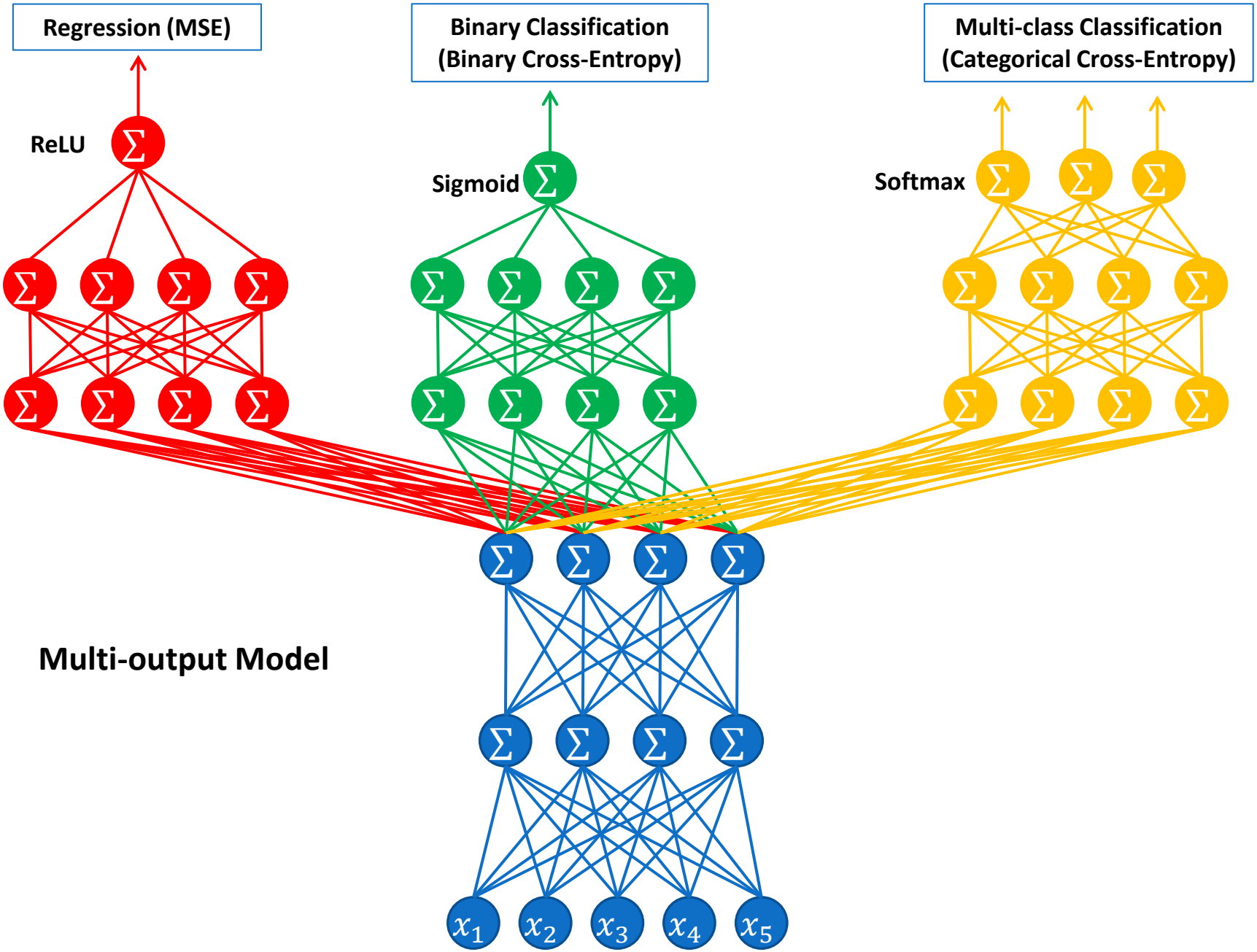
One hot encoded label: [0, 1, 0, 0, 1]



Credit: https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

Considering our age, emotion, and me/not me dataset, are we are not doing redundant work here also, if we train three different models?




Regression (MSE)
Binary Classification
(Binary Cross-Entropy)
Multi-class Classification
(Categorical Cross-Entropy)
ReLU
Sigmoid
Softmax
Multi-output Model
x_1
x_2
x_3
x_4
x_5

Relationship of Logistic Regression and Multilayer NN

- Neural Network is like logistic regression but with many hidden layers!

- We add layer after layer by combining multiple logistic regression units (so to speak)

- But for inputs, instead of handcrafted features, we give raw values as input and let the NN make sense of it! (Recall XOR example)
  - Prior layers of the network **induce the feature representation!**

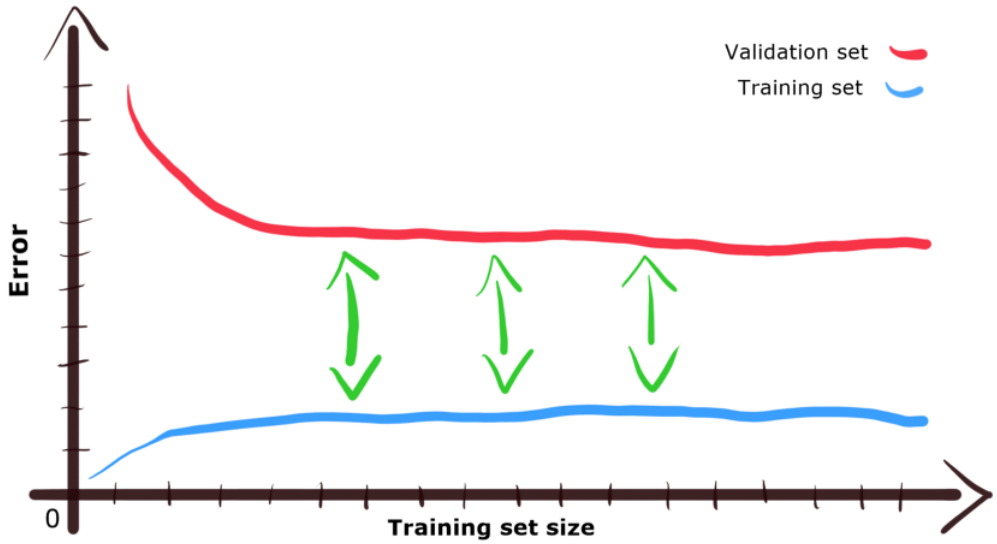
Roll of Bias in NNs

Roll of Bias in Neural Networks

- **Suppose you are training a neural network…**
- **But the model cannot learn the patterns from the data!**
- **Or it overfits the data!**

Roll of Bias in Neural Networks

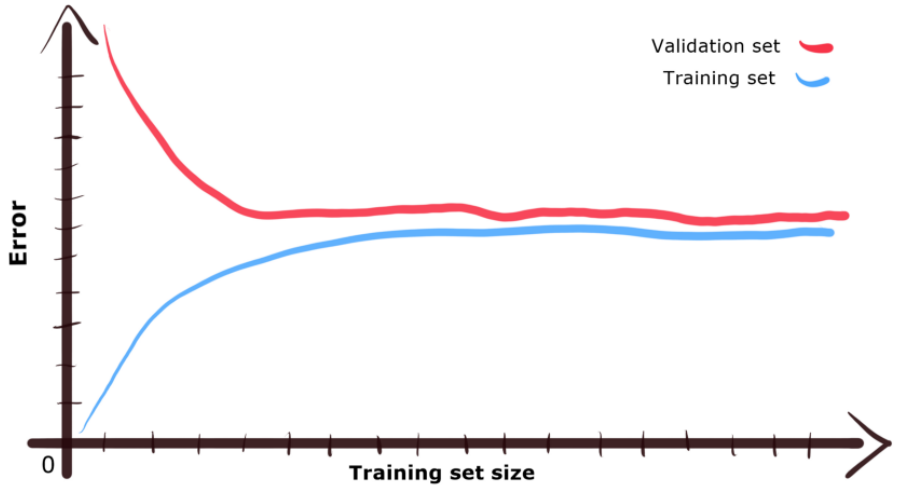
- **Low error for training** but **high for validation/test set** means we have a **high variance model**
- Big gap between these two errors signify **overfitting**.



Credit: https://towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98

Roll of Bias in Neural Networks

- But on the other hand, if **training and validation losses are very high but closer to each other**:
- This essentially means the **model is not learning** what we are trying to teach it.
  - Either something is wrong with the data or the model



Credit: https://towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98

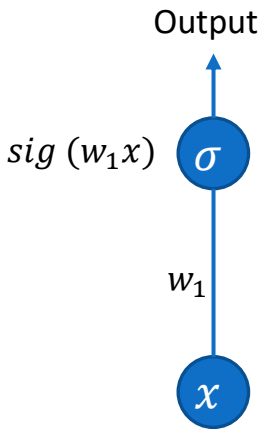
Roll of Bias in Neural Networks

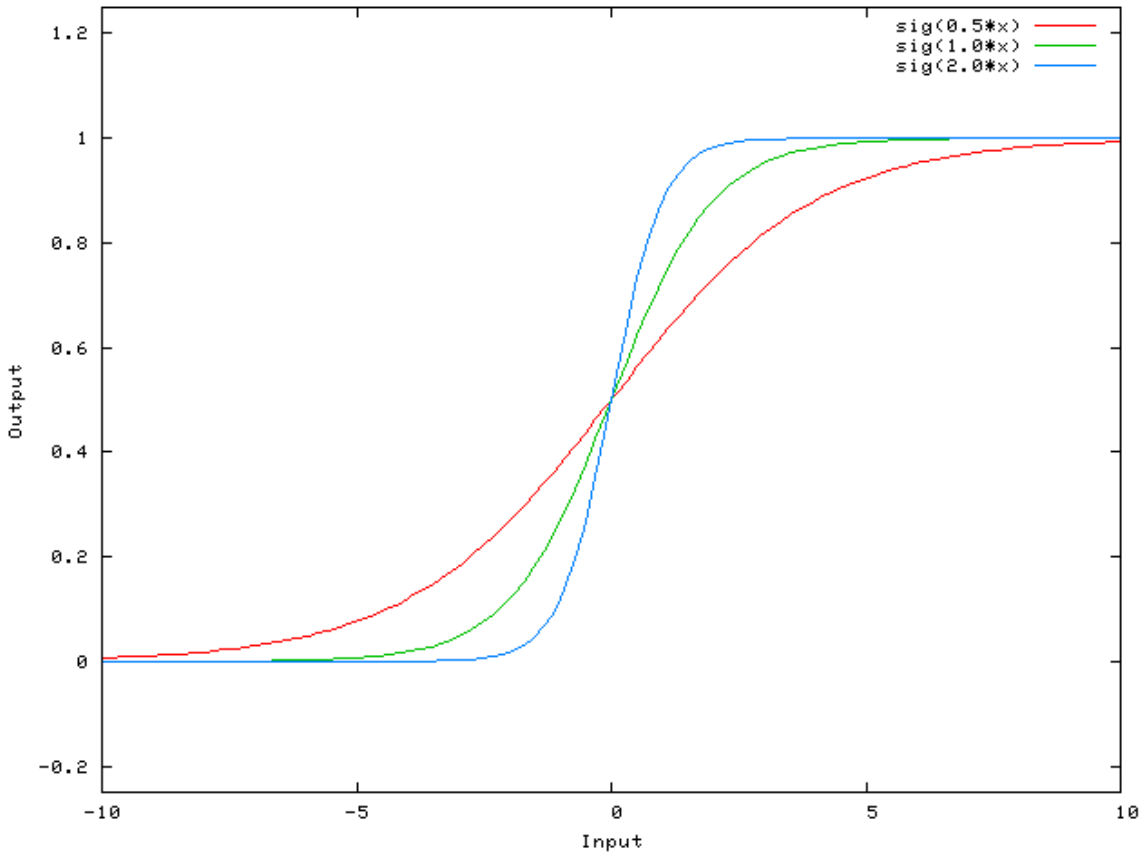
- **The bias acts like an extra feature that *"tells the model what to learn"*.**

Kind of!!

- **A bias value allows the shift of the activation function to the left or right, which may be critical for successful learning!**

Roll of Bias in Neural Networks





Changing the weight only changes the **"steepness"** of the sigmoid!

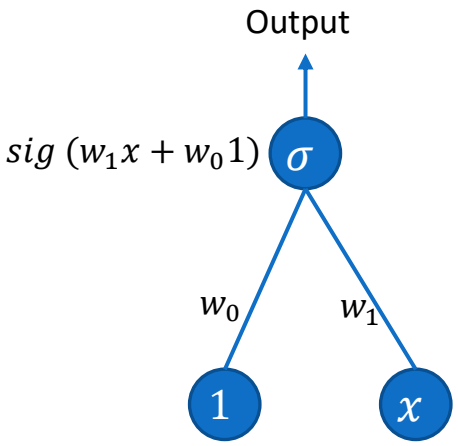
But what if we want the network to output 0 when $x = 2$ or $x = 5$

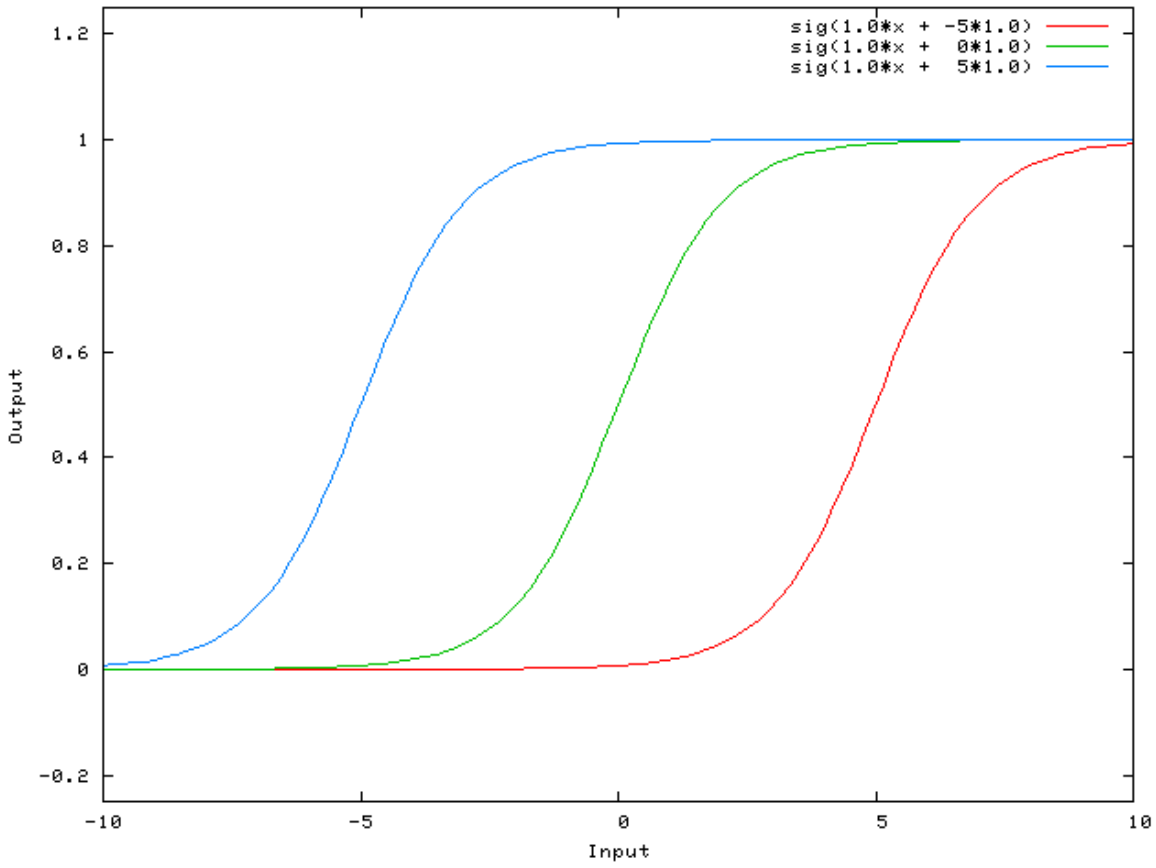
This won't work no matter how much you change the "steepness"

You need to shift the whole curve to the right

Credit: https://stackoverflow.com/questions/2480650/what-is-the-role-of-the-bias-in-neural-networks

Roll of Bias in Neural Networks





Bias value to the rescue!

As bias is also learnable, the network "*learns*" how much curve has to be shifted to left or right to produce correct output!!

Credit: https://stackoverflow.com/questions/2480650/what-is-the-role-of-the-bias-in-neural-networks

Take Away….

- **Always use bias!!!**
  - **If it is not needed at all, the network will make it's contribution in the output 0 by learning it's corresponding wight!**

- **Multilayer network is better than single layer network (Perceptron) for learning non-linear decision boundaries**

More Details:

https://www.geeksforgeeks.org/effect-of-bias-in-neural-network/
https://towardsdatascience.com/why-we-need-bias-in-neural-networks-db8f7e07cb98

What should be the activation function for a regression model?

- For predicting “age”
  - ReLU
- For predicting “price” for a house
  - ReLU
- For predicting by how much a stock price would go up or down
  - Linear (aka no activation function)
- For assigning a score to an essay where the score value is between $0 - 1$
  - ReLU
  - Sigmoid (more suitable)

Quiz 4

- Thursday 11th May 2023
- Syllabus
  - Lectures 17, 18, 19

Book Reading

- Murphy – Chapter 8
- Jurafsky – Chapter 5, Chapter 4, Chapter 7
- Tom Mitchel – Chapter 4