

Introduction to Deep Learning

CSI 5180 - Machine Learning for Bioinformatics

Marcel Turcotte

Version: Mar 10, 2025 10:54

Preamble

Quote of the Day

Summary

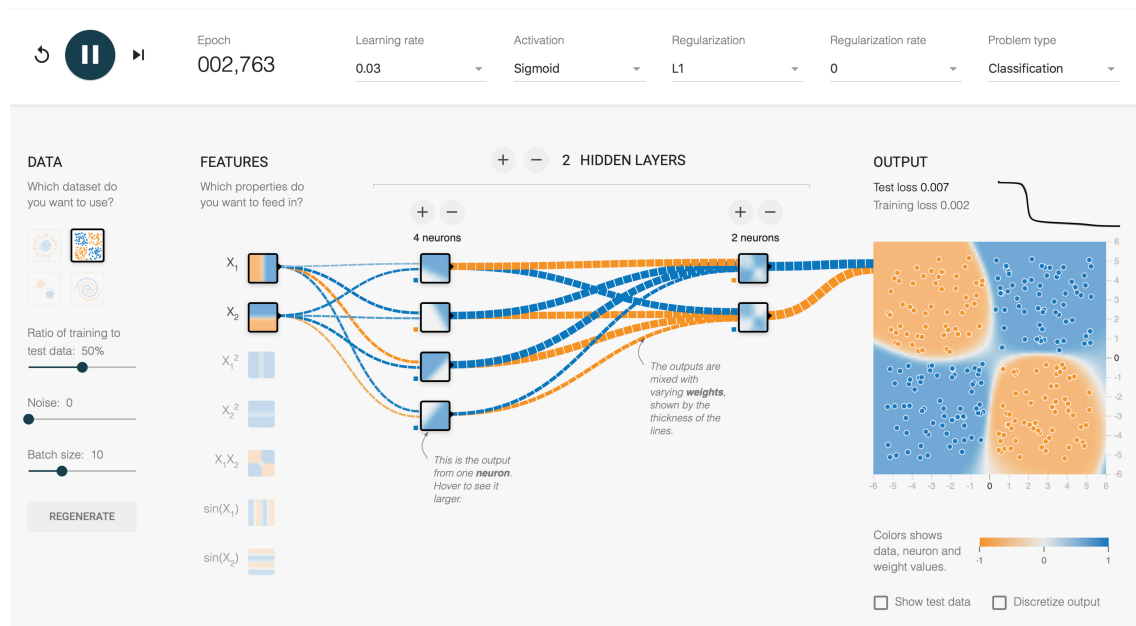
Neural networks evolved from simple, biologically inspired perceptrons to **deep, multilayer** architectures that rely on **nonlinear activation functions** for learning complex patterns. The **universal approximation theorem** underpins their ability to approximate any continuous function, and modern frameworks like **PyTorch**, **TensorFlow**, and **Keras** enable practical deep learning applications.

Learning Objectives

- **Explain** basic neural network models (perceptrons and MLPs) and their computational foundations.
- **Appreciate** the limitations of single-layer networks and the necessity for hidden layers.
- **Describe** the role and impact of nonlinear activation functions (sigmoid, tanh, ReLU) in learning.
- **Articulate** the universal approximation theorem and its significance.
- **Implement** and evaluate deep learning models using modern frameworks such as TensorFlow and Keras.

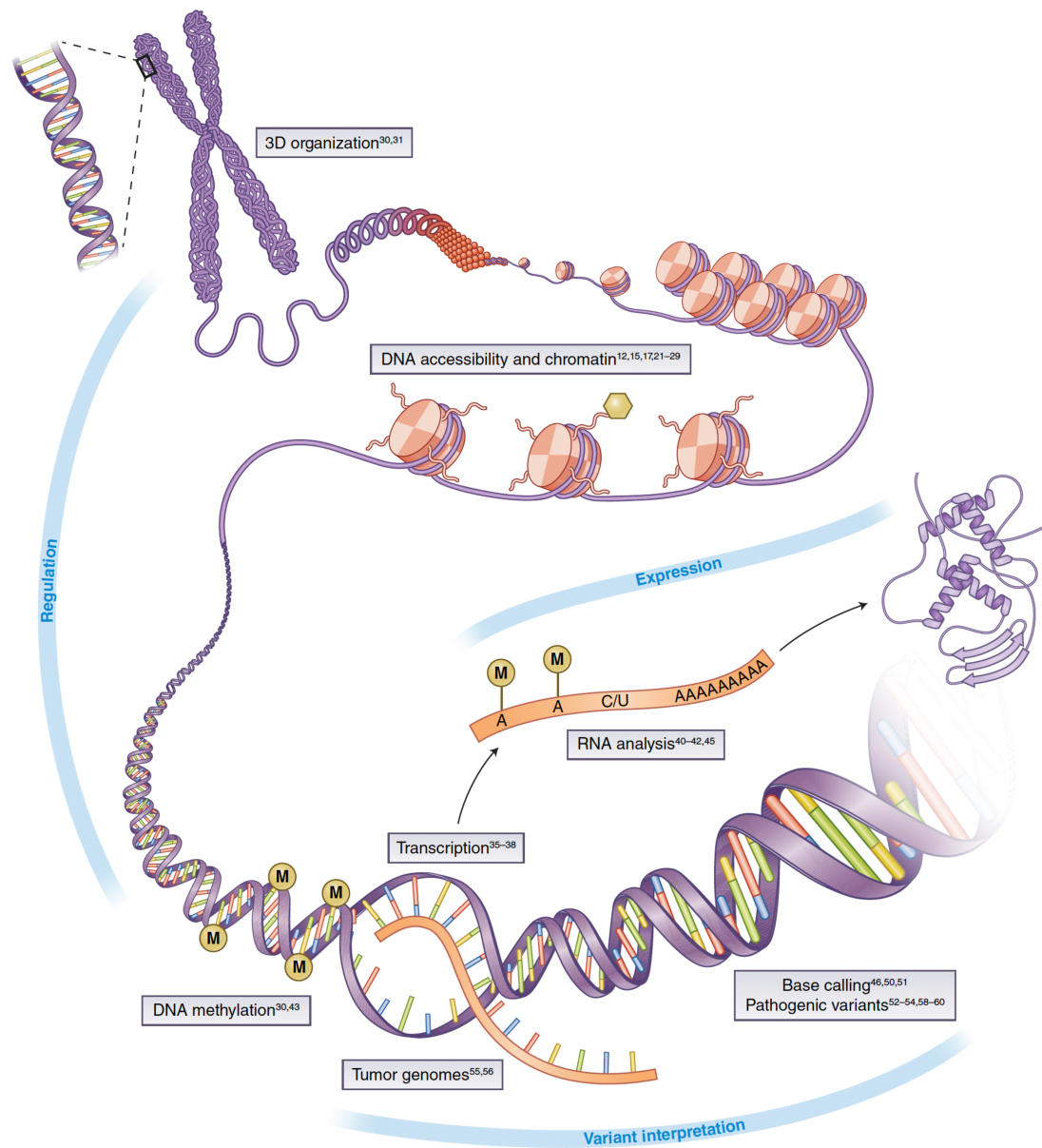
Introduction

TensorFlow Playground



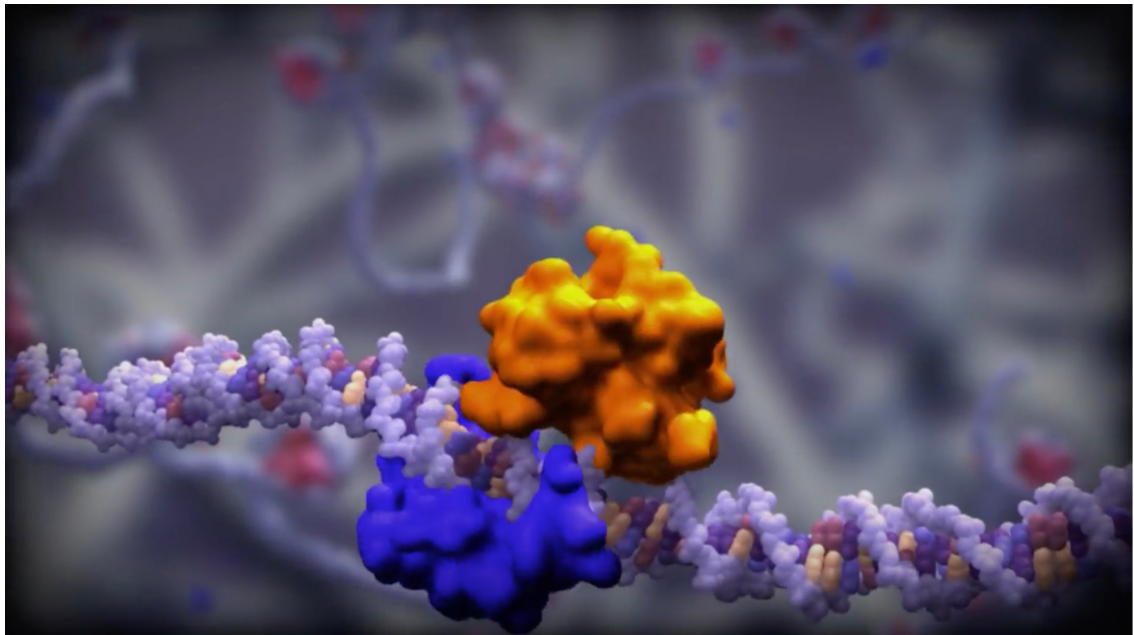
playground.tensorflow.org

Primer on Deep Learning in Genomics



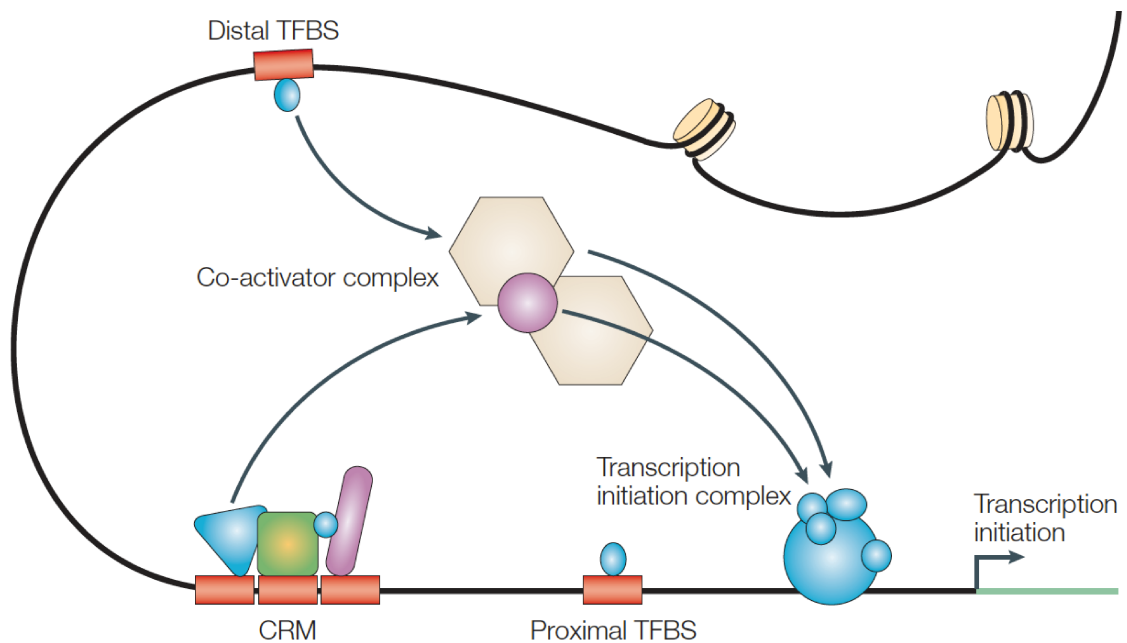
Zou et al. (2019), Figure 2

Transcription Factors



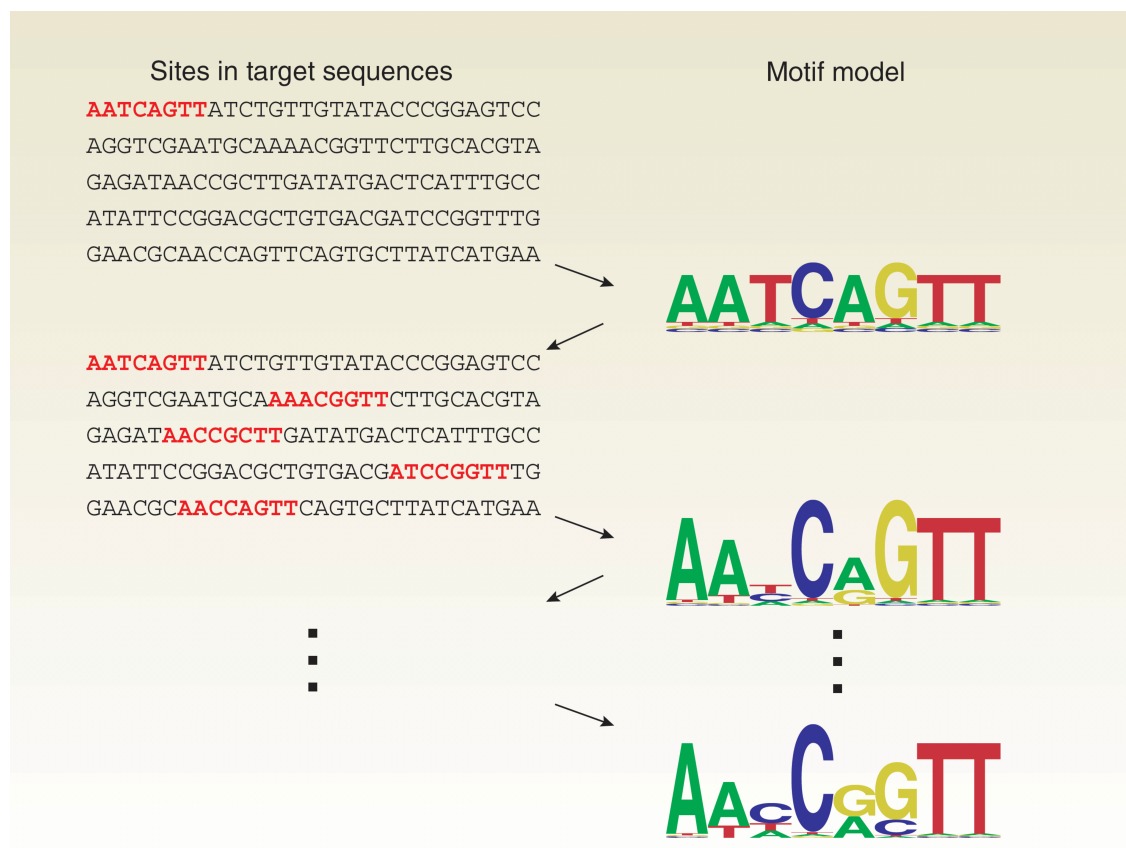
See: <https://youtu.be/MkUgkDLp2iE>

Gene Regulation



Wasserman and Sandelin (2004)

DNA Sequence Motif Discovery



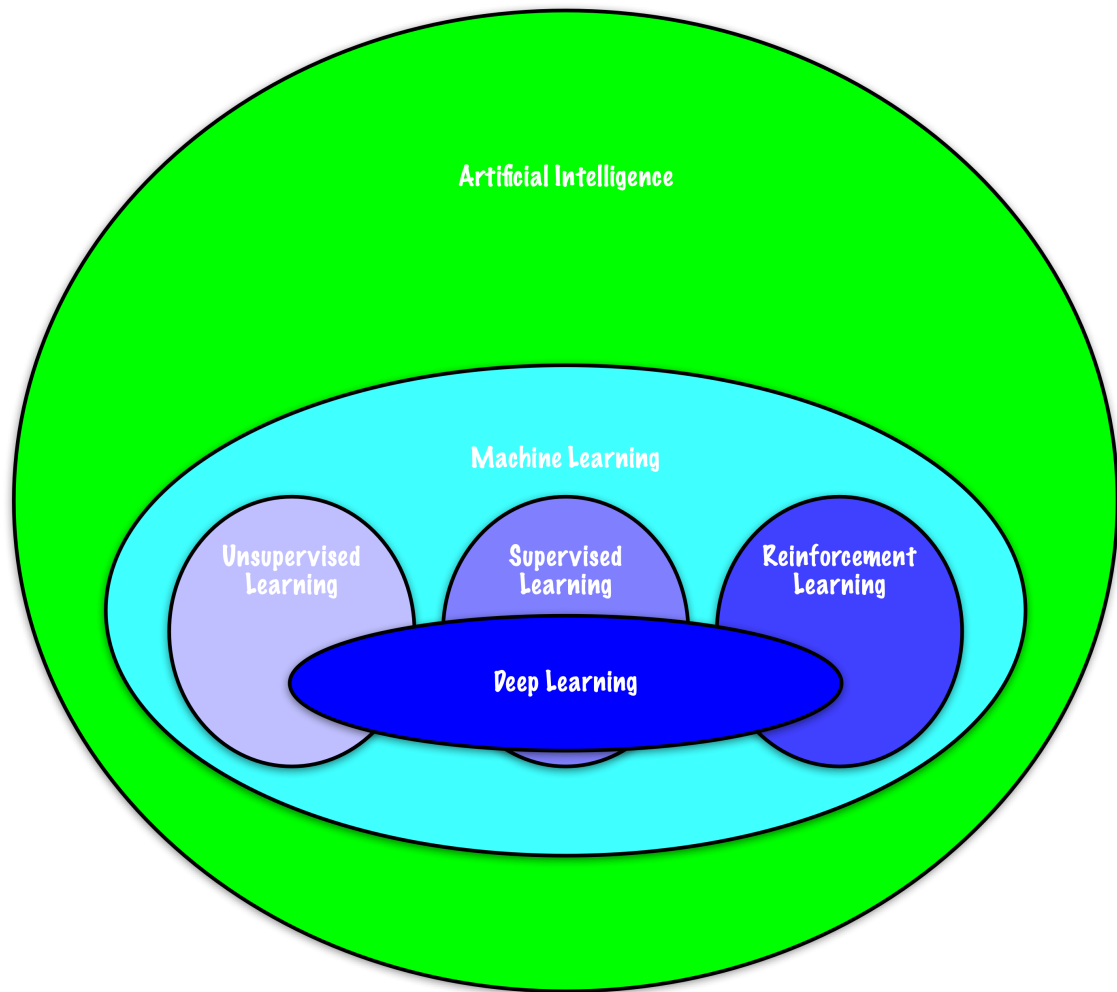
D'haeseleer (2006)

Deep Learning in Genomics Primer

- James Zou, Mikael Huss, Abubakar Abid, Pejman Mohammadi, Ali Torkamani, and Amalio Telenti, A primer on deep learning in genomics, *Nat Genet* **51**:1, 12–18, 2019.
 - Google Colab Notebook

Zou et al. (2019)

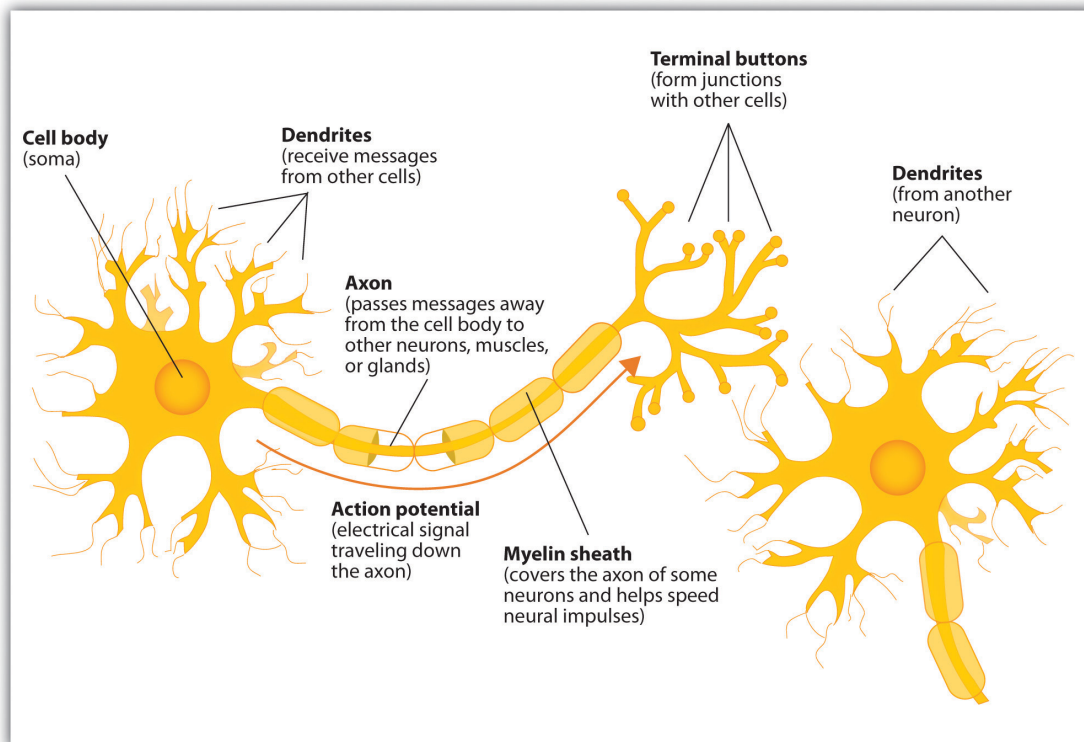
Machine Learning Problems



- **Supervised Learning:** Classification, Regression
- **Unsupervised Learning:** Autoencoders, Self-Supervised
- **Reinforcement Learning:** Now an Integral Component

We will begin our exploration within the framework of supervised learning.

A Neuron



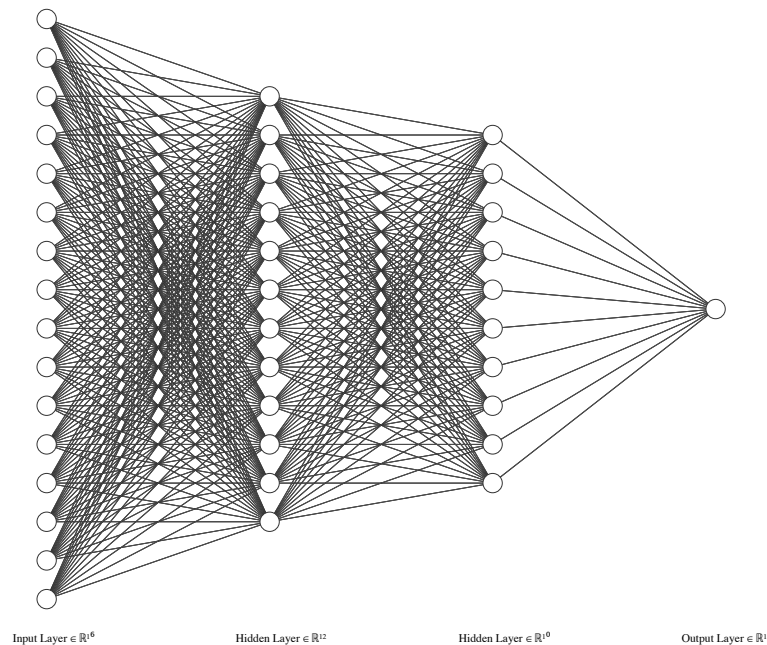
Attribution: Jennifer Walinga, [CC BY-SA 4.0](#)

In the study of artificial intelligence, it is logical to derive inspiration from the most well-understood form of intelligence: the human brain. The brain is composed of a complex network of neurons, which together form biological neural networks. Although each neuron exhibits relatively simple behavior, it is connected to thousands of other neurons, contributing to the intricate functionality of these networks.

A neuron can be conceptualized as a basic computational unit, and the complexity of brain function arises from the interconnectedness of these units.

Yann LeCun and other researchers have frequently noted that artificial neural networks used in machine learning resemble biological neural networks in much the same way that an airplane's wings resemble those of a bird.

Connectionist



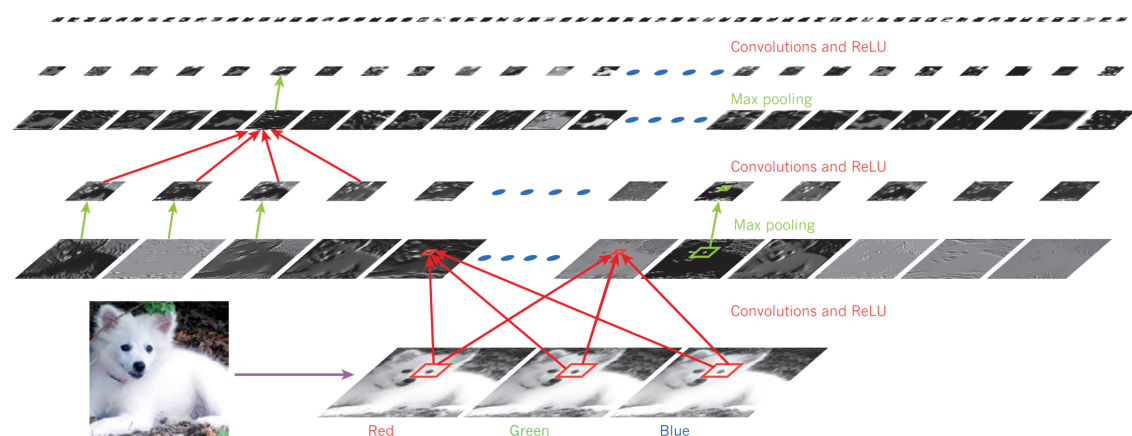
Attribution: LeNail, (2019). NN-SVG: Publication-Ready Neural Network Architecture Schematics. Journal of Open Source Software, 4(33), 747, <https://doi.org/10.21105/joss.00747> (GitHub)

A characteristic of biological neural networks that we adopt is the organization of neurons into layers, particularly evident in the cerebral cortex.

Neural networks (NNs) consist of layers of **interconnected nodes (neurons)**, each **connection** having an associated **weight**.

Neural networks process input data through these weighted connections, and **learning** occurs by **adjusting the weights** based on **errors** in the **training data**.

Hierarchy of Concepts



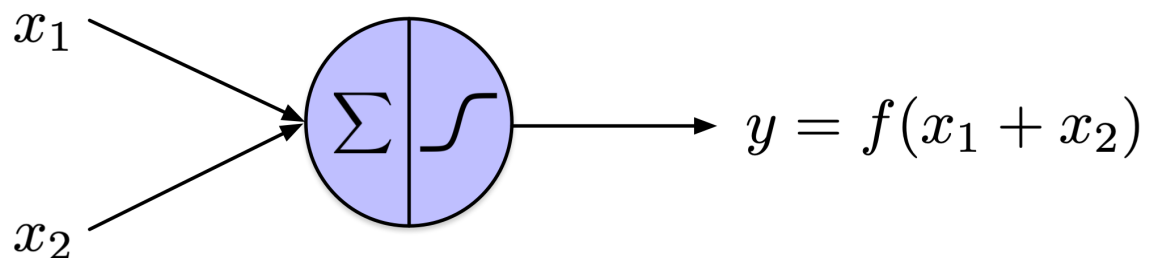
Attribution: LeCun, Bengio, and Hinton (2015)

In the book “Deep Learning” (Goodfellow, Bengio, and Courville 2016), authors Goodfellow, Bengio, and Courville define deep learning as a subset of machine learning that enables computers to “understand the world in terms of a hierarchy of concepts.”

This hierarchical approach is one of deep learning’s most significant contributions. It reduces the need for manual feature engineering and redirects the focus toward the engineering of neural network architectures.

Basics

Computations with Neurodes



where $x_1, x_2 \in \{0,1\}$ and $f(z)$ is an **indicator function**:

$$f(z) = \begin{cases} 0, & z < \theta \\ 1, & z \geq \theta \end{cases}$$

McCulloch and Pitts (1943) termed artificial neurons, **neurodes**, for “**neuron**” + “**node**”.

In mathematics, $f(z)$, as defined above, is known as an **indicator function** or a **characteristic function**.

These neurodes have one or more binary inputs, taking a value of 0 or 1, and one binary output.

They showed that such units could implement Boolean functions such as **AND**, **OR**, and **NOT**.

But also that networks of such units can compute any logical proposition.

Computations with Neurodes

$$y = f(x_1 + x_2) = \begin{cases} 0, & x_1 + x_2 < \theta \\ 1, & x_1 + x_2 \geq \theta \end{cases}$$

- With $\theta = 2$, the neurode implements an **AND** logic gate.
- With $\theta = 1$, the neurode implements an **OR** logic gate.

More **complex logic** can be constructed by multiplying the inputs by **-1**, which is interpreted as **inhibitory**. Namely, this allows building a logical NOT.

With $\theta = 1$, x_1 and x_2 multiplied by (-1) , $y = 0$ when $x_2 = 1$, $y = 1$, if $x_2 = 0$.

$$y = f(x_1 + (-1)x_2) = \begin{cases} 0, & x_1 + x_2 < \theta \\ 1, & x_1 + (-1)x_2 \geq \theta \end{cases}$$

Neurons can be broadly categorized into two primary types: **excitatory** and **inhibitory**.

Computations with Neurodes

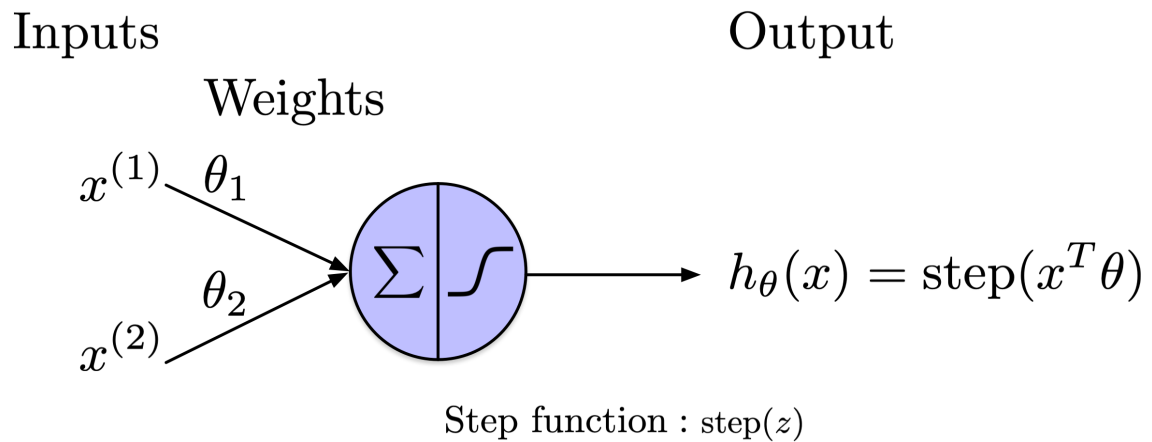
- **Digital computations** can be broken down into a **sequence of logical operations**, enabling neurode networks to **execute any computation**.
- McCulloch and Pitts (1943) did **not** focus on **learning** parameter θ .
- They introduced a machine that **computes any function** but **cannot learn**.

The period roughly from 1930 to 1950 marked a transformative shift in mathematics toward the formalization of computation. Pioneering work by **Gödel**, **Church**, and **Turing** not only established the theoretical limits and capabilities of computation—with Gödel's **incompleteness theorems**, Church's **λ -calculus and thesis**, and Turing's model of **universal machines**—but also set the stage for later developments in computer science.

McCulloch and Pitts' 1943 model of neural networks was inspired by this early mathematical framework linking computation to aspects of intelligence, prefiguring later research in artificial intelligence.

From this work, we take the idea that networks of such units perform computations. Signal propagates from one end of the network to compute a result.

Threshold Logic Unit



Weighted sum : $z = x^T \theta$

Rosenblatt (1958)

In 1957, Frank Rosenblatt developed a conceptually distinct model of a neuron known as the **threshold logic unit**, which he published in 1958.

In this model, both the inputs and the output of the neuron are represented as **real values**. Notably, each input connection has an associated weight.

The left section of the neuron, denoted by the sigma symbol, represents the computation of a weighted sum of its inputs, expressed as $\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_D x_D + b$.

This sum is then processed through a step function, right section of the neuron, to generate the output.

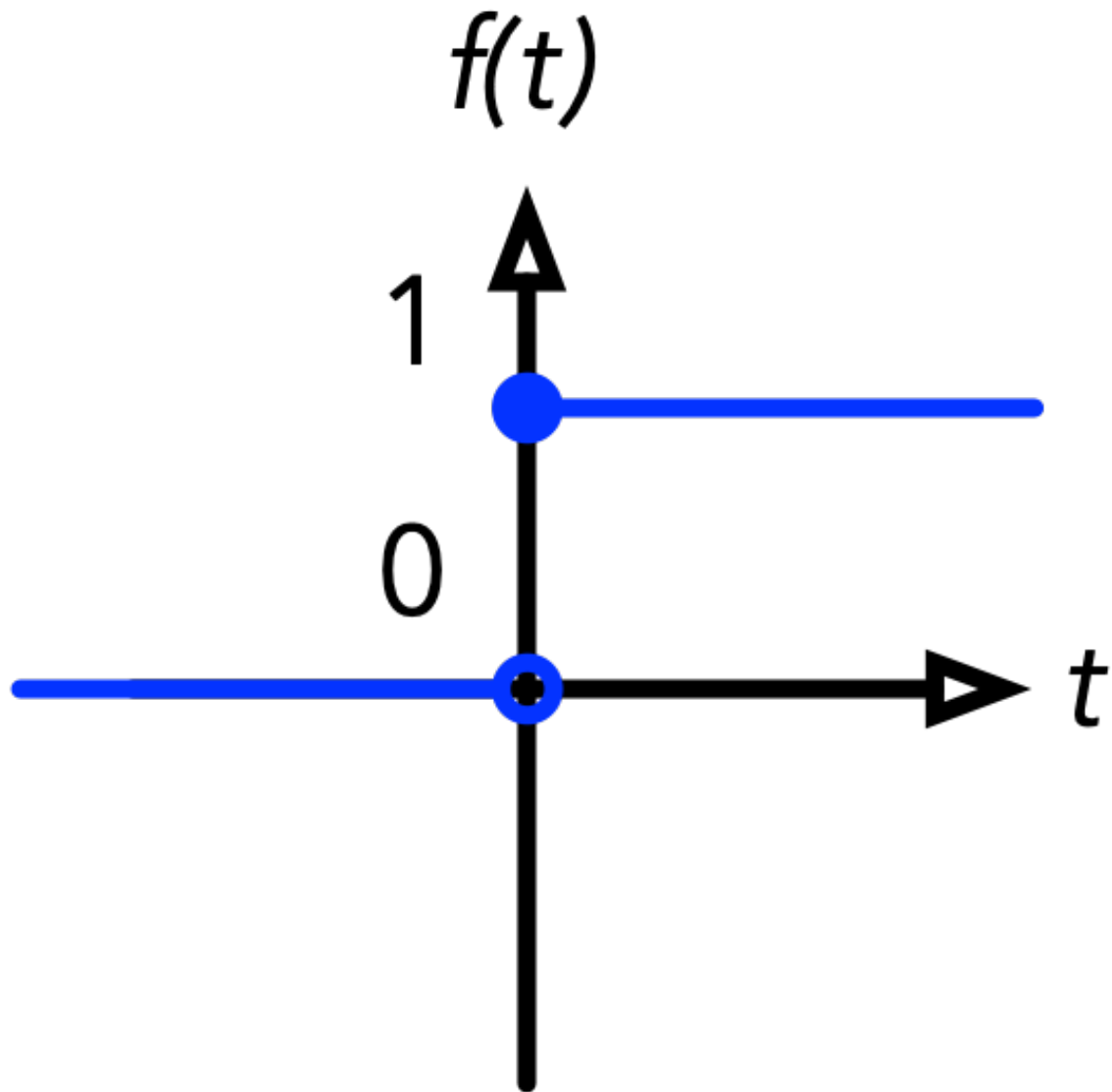
Here, $x^T \theta$ represents the dot product of two vectors: x and θ . Here, x^T denotes the transpose of the vector x , converting it from a row vector to a column vector, allowing the dot product operation to be performed with the vector θ .

The dot product $x^T \theta$ is then a scalar given by:

$$x^T \theta = x^{(1)} \theta_1 + x^{(2)} \theta_2 + \dots + x^{(D)} \theta_D$$

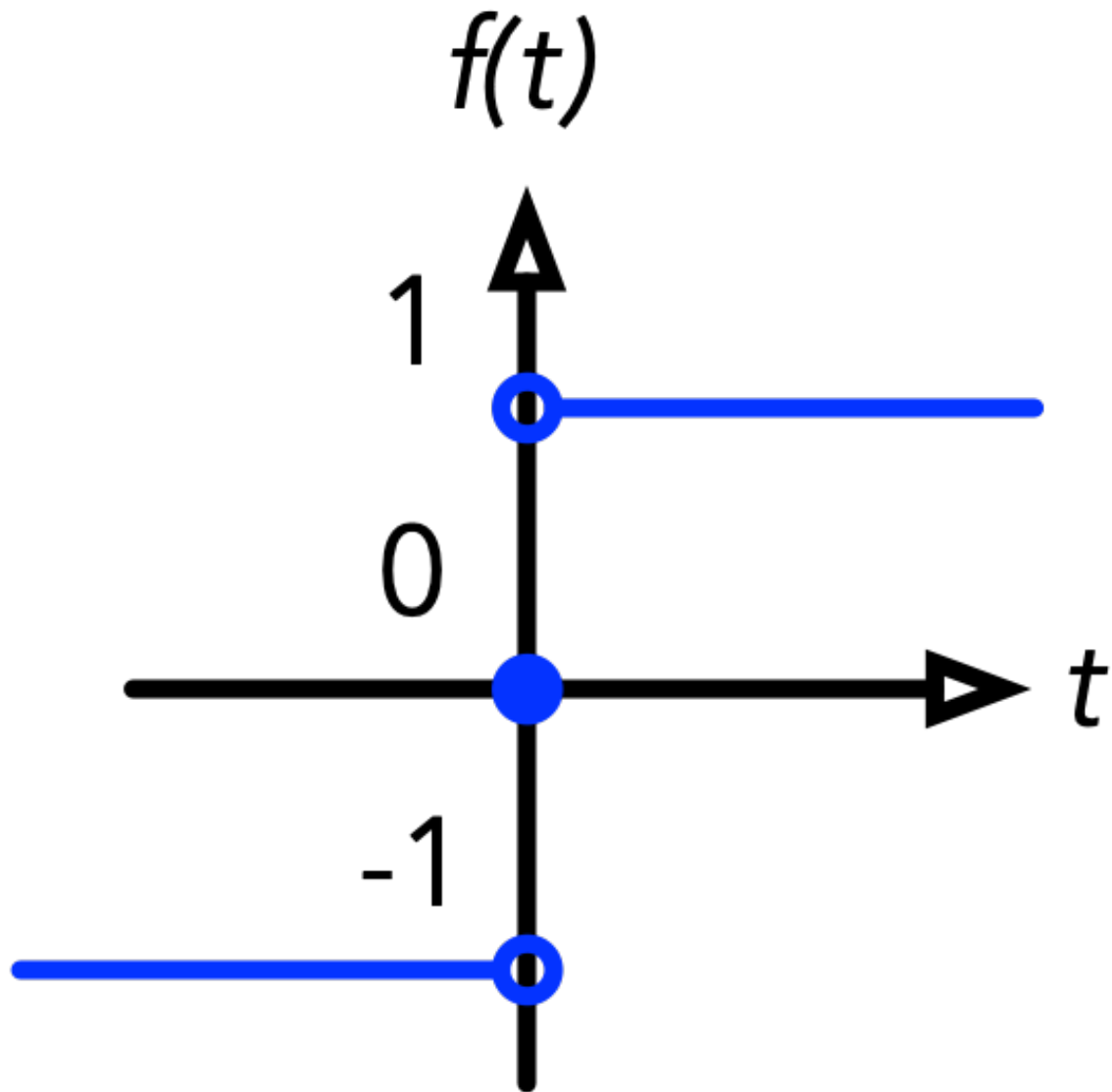
where $x^{(j)}$ and θ_j are the components of the vectors x and θ , respectively.

Simple Step Functions



$\text{Heaviside}(t) =$

- 1, if $t \geq 0$
- 0, if $t < 0$

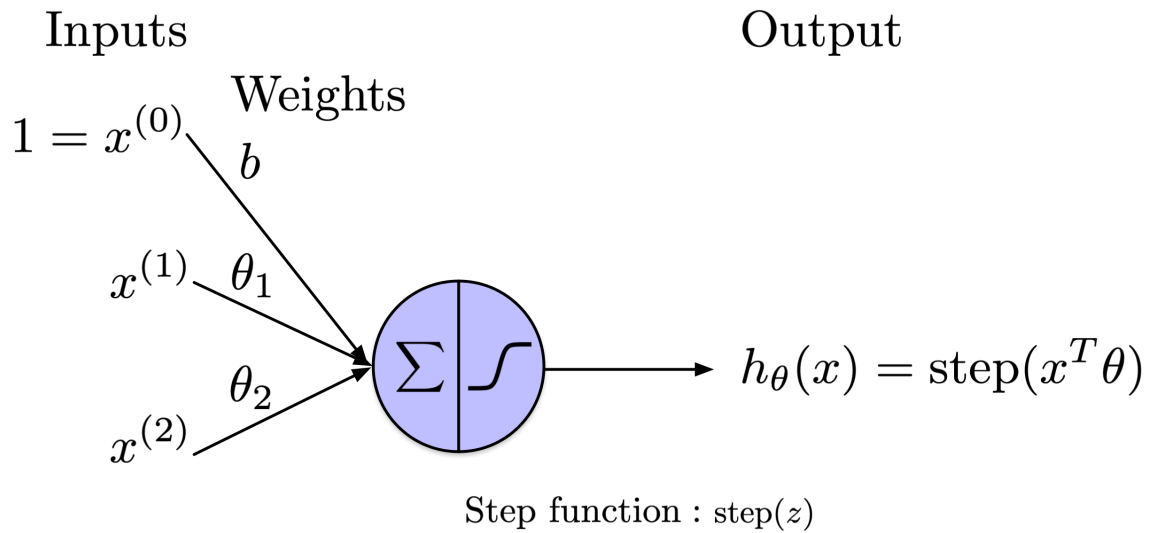


$\text{sign}(t) =$

- 1, if $t > 0$
- 0, if $t = 0$
- -1, if $t < 0$

Common **step functions** include the **heavyside function** (0 if the input is negative and 1 otherwise) or the **sign** function (-1 if the input is negative, 0 if the input is zero, 1 otherwise).

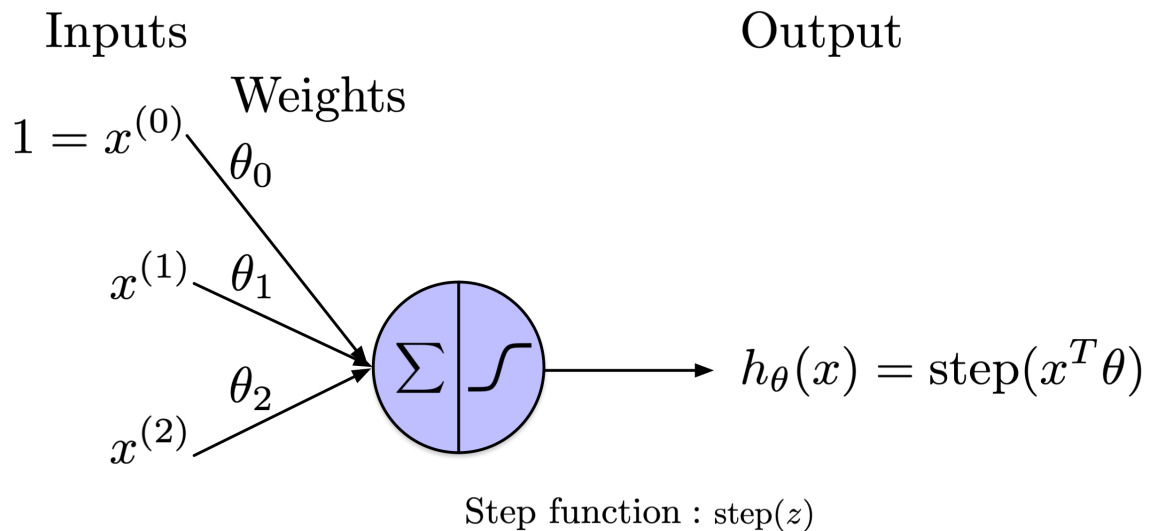
Notation



Weighted sum : $z = x^T \theta$

Add an extra feature with a fixed value of 1 to the input. Associate it with weight θ_0 , where θ_0 is the bias/intercept term.

Notation



Weighted sum : $z = x^T \theta$

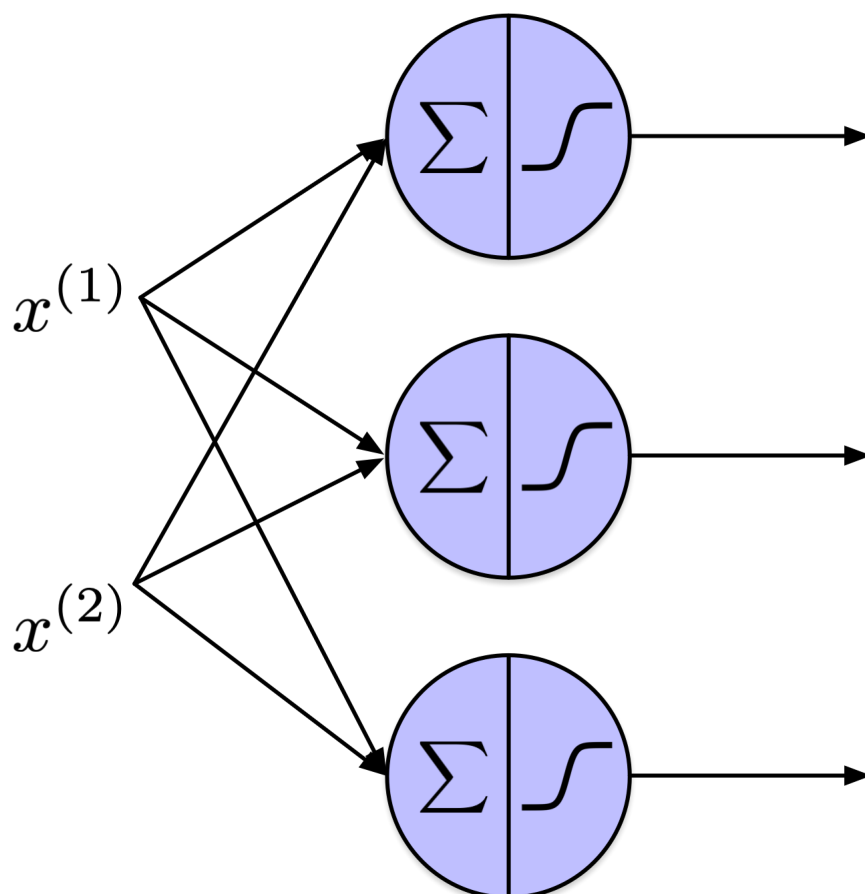
$\theta_0 = b$ is the bias/intercept term.

The threshold logic unit is analogous to logistic regression, with the primary distinction being the substitution of the logistic (sigmoid) function with a step function. Similar to logistic regression, the perceptron is employed for classification tasks.

Perceptron

Inputs

Outputs



Input Layer

Output Layer

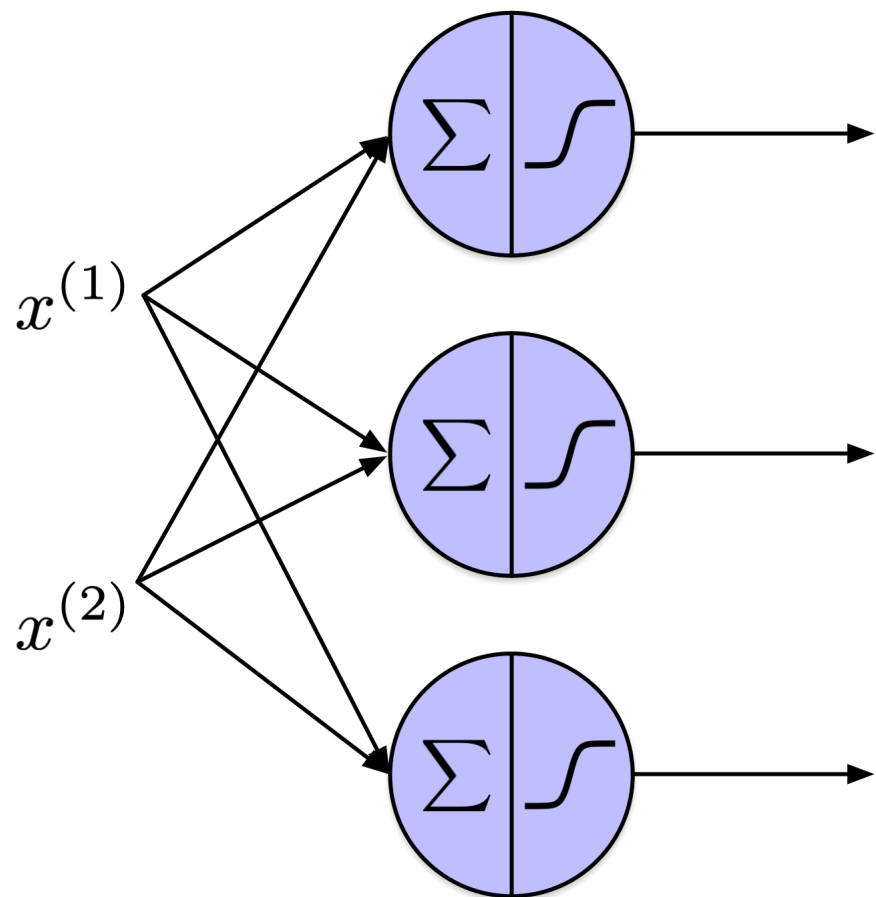
A **perceptron** consists of one or more **threshold logic units** arranged in a **single layer**, with each unit connected to all inputs. This configuration is referred to as **fully connected** or **dense**.

Since the threshold logic units in this single layer also generate the output, it is referred to as the **output layer**.

Perceptron

Inputs

Outputs



Input Layer

Output Layer

As this perceptron generates multiple outputs simultaneously, it performs **multiple binary predictions**, making it as a **multilabel classifier** (can also be used as multiclass classifier).

Classification tasks, can be further divided into **multilabel** and **multiclass** classification.

1. Multiclass Classification:

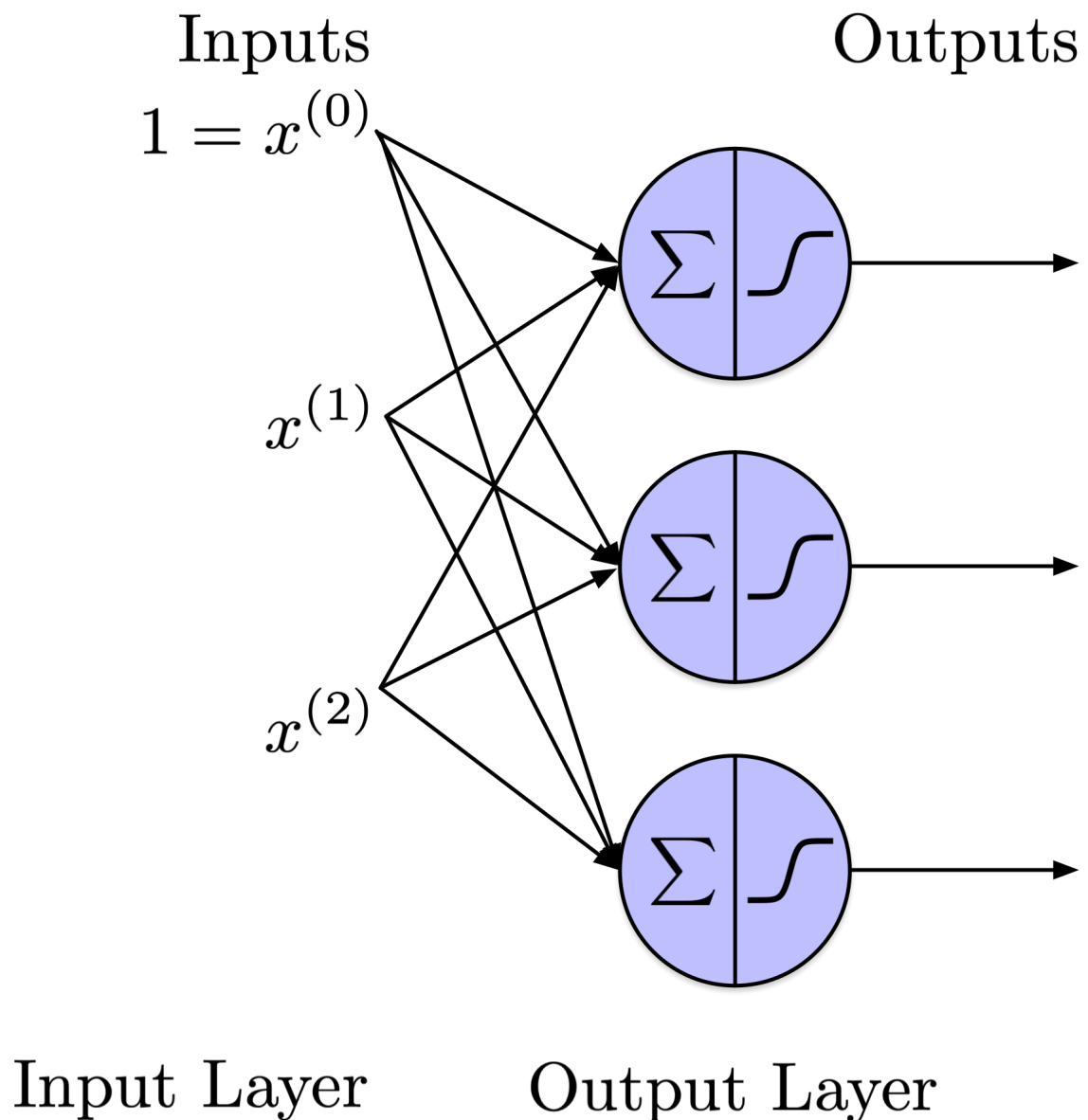
- In multiclass classification, each instance is assigned to one and only one class out of a set of three or more possible classes. The classes are mutually exclusive, meaning that an instance cannot belong to more than one class at the same time.
- **Example:** Classifying an image as either a cat, dog, or bird. Each image can only belong to one of these categories.

2. Multilabel Classification:

- In multilabel classification, each instance can be associated with multiple classes simultaneously. The classes are not mutually exclusive, allowing for the possibility that an instance can belong to several classes at once.
- **Example:** Tagging an image with multiple attributes such as "outdoor," "sunset," and "beach." The image can simultaneously belong to all these labels.

The key difference lies in the relationship between classes: **multiclass classification deals with a single label per instance**, while **multilabel classification handles multiple labels for each instance**.

Notation



As before, introduce an additional feature with a value of 1 to the input. **Assign a bias b to each neuron.** Each incoming connection **implicitly** has an associated weight.

Notation

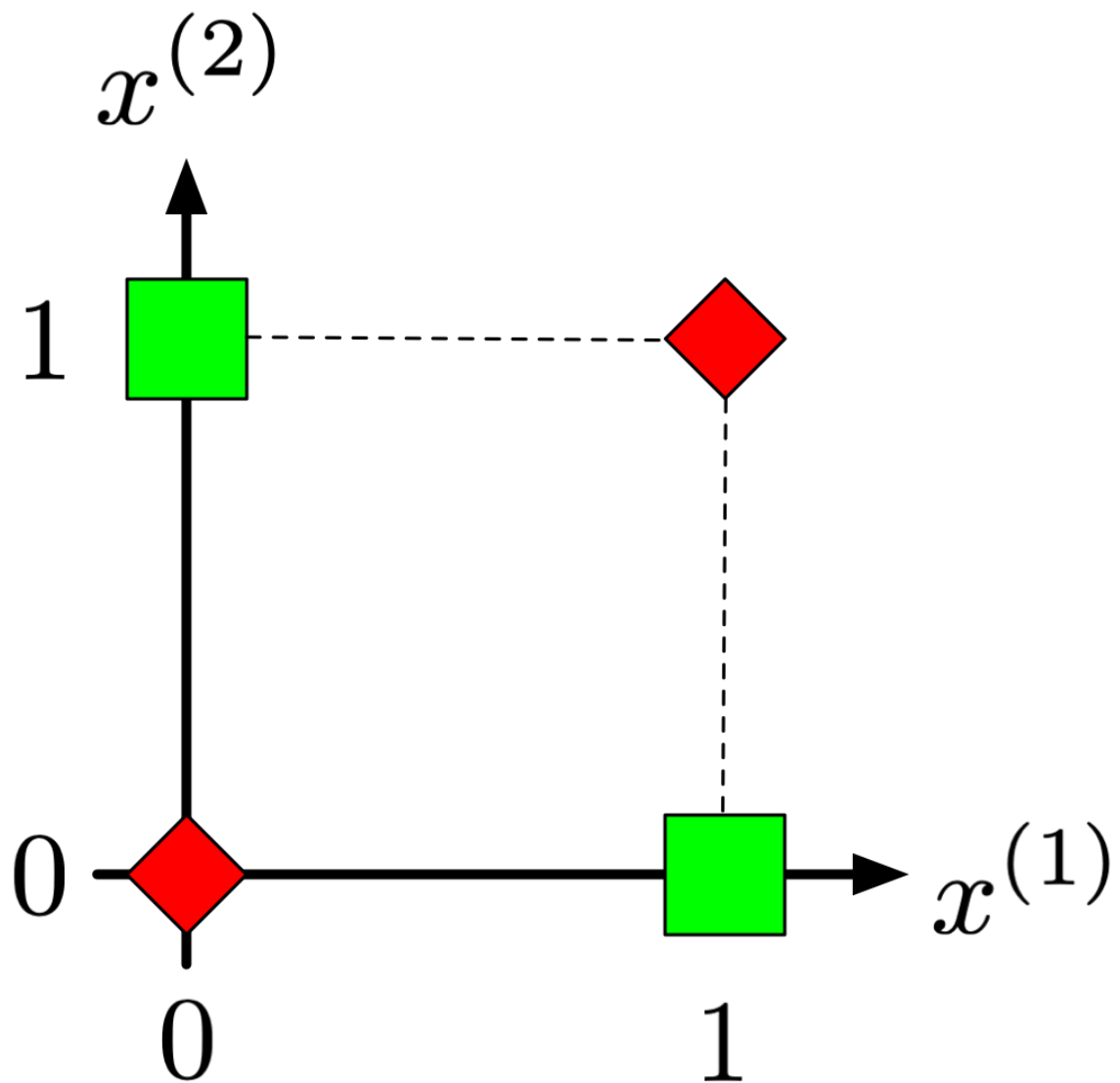
- X is the input **data matrix** where **each row corresponds to an example** and **each column represents one of the D features**.
- W is the **weight matrix**, structured with one **row per input (feature)** and **one column per neuron**.
- **Bias terms** can be represented separately; both approaches appear in the literature. Here, b is a vector with a **length equal to the number of neurons**.

With neural networks, the **parameters** of the model are often referred to as w (vector) or W (matrix), rather than θ .

Discussion

- The algorithm to **train** the perceptron closely resembles stochastic gradient descent.
 - In the **interest of time** and to **avoid confusion**, we will skip this algorithm and focus on **multilayer perception** (MLP) and its training algorithm, **backpropagation**.

Historical Note and Justification

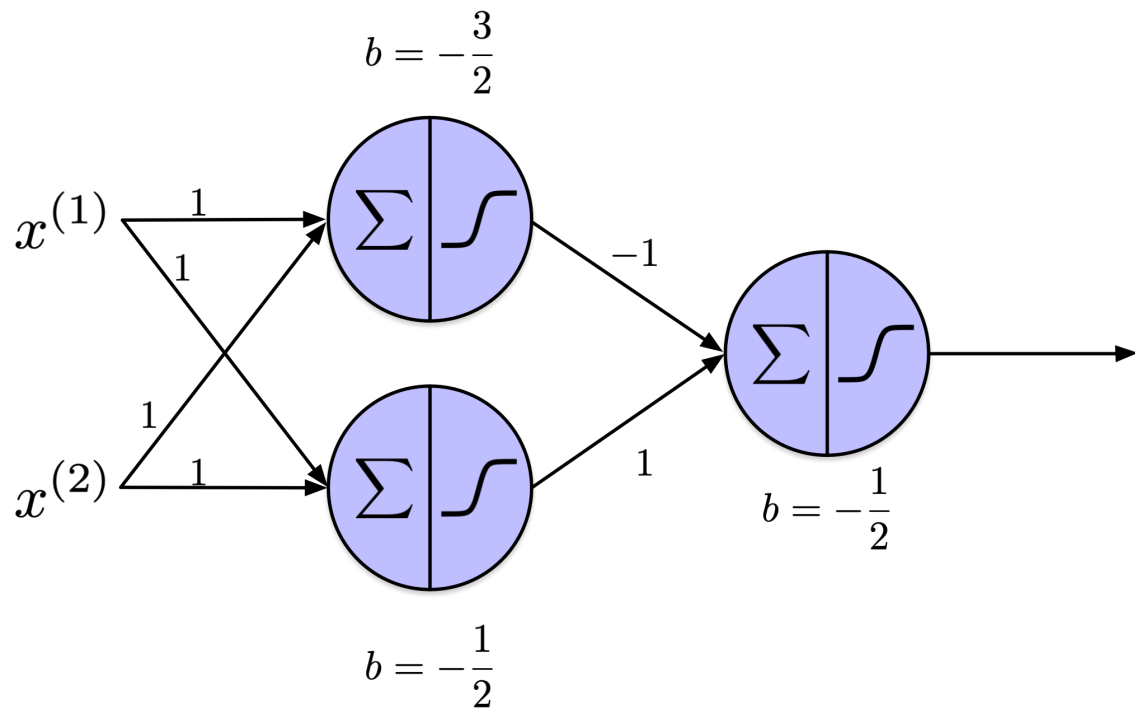


Minsky and Papert (1969) demonstrated the limitations of perceptrons, notably their inability to solve **exclusive OR** (XOR) classification problems: $\{([0,1], \mathrm{true}), ([1,0], \mathrm{true}), ([0,0], \mathrm{false}), ([1,1], \mathrm{false})\}$.

This limitation also applies to other linear classifiers, such as logistic regression.

Consequently, due to these limitations and a lack of practical applications, some researchers abandoned the perceptron.

Multilayer Perceptron



A **multilayer perceptron** (MLP) includes an input layer and one or more layers of threshold logic units. Layers that are neither input nor output are termed **hidden layers**.

XOR Classification Problem

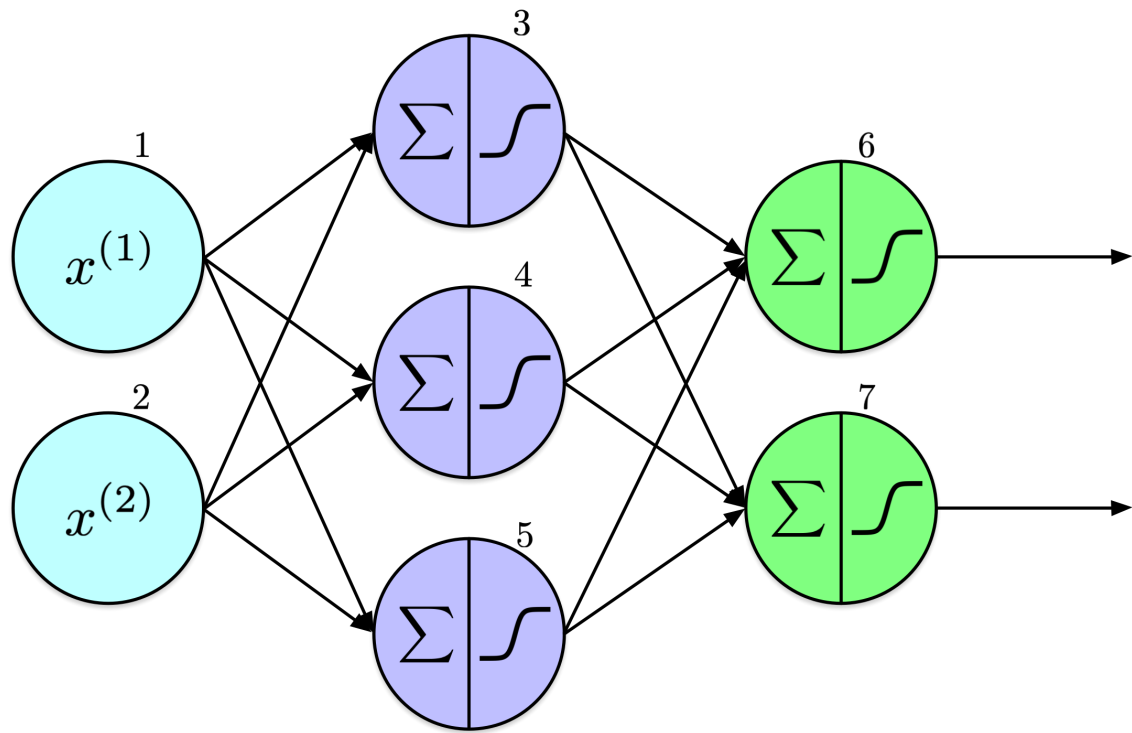
$x^{(1)}$	$x^{(2)}$	y	o_1	o_2	o_3
1	0	1	0	1	1
0	1	1	0	1	1
0	0	0	0	0	0
1	1	0	1	1	0

$x^{(1)}$ and $x^{(2)}$ are two attributes, y is the target, o_1 , o_2 , and $o_3 = h_{\theta}(x)$, are the output of the top left, bottom left, and right threshold units. Clearly $h_{\theta}(x) = y$, $\forall x \in X$. The challenge during Rosenblatt's time was the lack of algorithms to train multi-layer networks.

I developed an Excel spreadsheet to verify that the proposed multilayer perceptron effectively solves the XOR classification problem.

The step function used in the above model is the heavyside function.

Feedforward Neural Network (FNN)



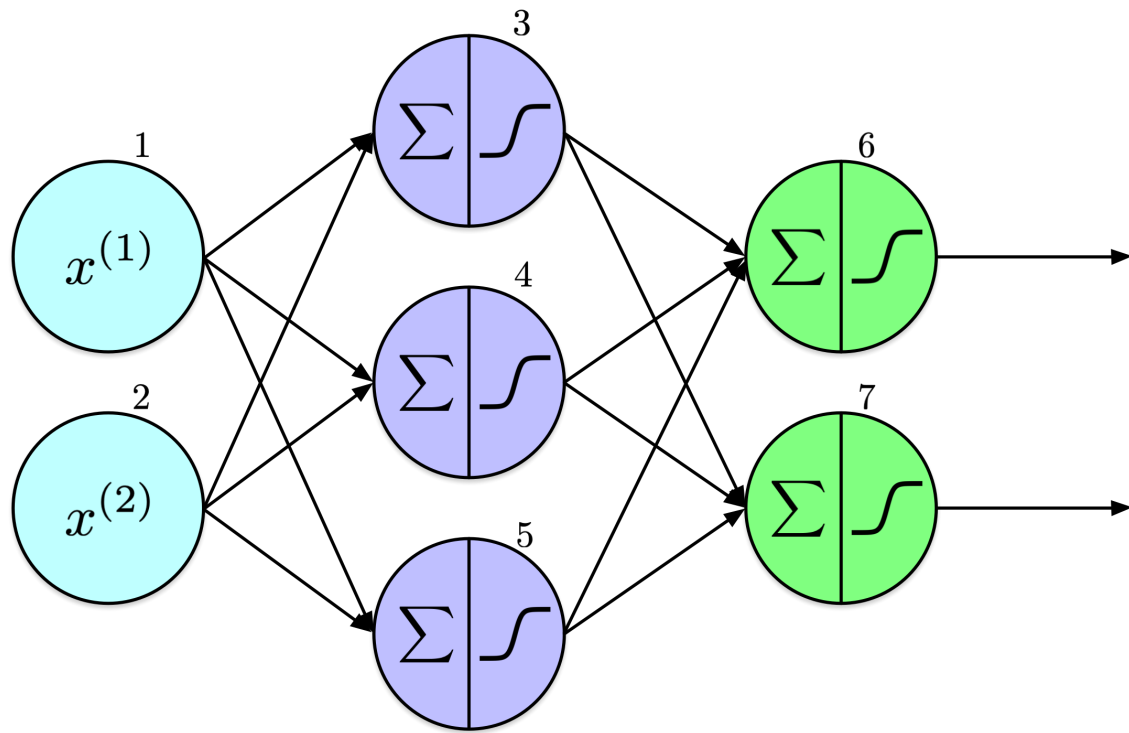
Information in this architecture flows unidirectionally—from left to right, moving from input to output. Consequently, it is termed a **feedforward neural network**.

The network consists of **three layers**: input, hidden, and output. The **input layer** contains two nodes, the **hidden layer** comprises three nodes, and the **output layer** has two nodes. Additional hidden layers and nodes per layer can be added, which will be discussed later.

It is often useful to include explicit input nodes that do not perform calculations, known as **input units** or **input neurons**. These nodes act as placeholders to introduce input features into the network, passing data directly to the next layer without transformation. In the network diagram, these are the light blue nodes on the left, labeled 1 and 2. Typically, **the number of input units corresponds to the number of features**.

For clarity, nodes are labeled to facilitate discussion of the weights between them, such as $w_{1,5}$ between nodes 1 and 5. Similarly, the output of a node is denoted by o_k , where k represents the node's label. For example, for $k=3$, the output would be o_3 .

Forward Pass (Computation)



$$o_3 = \sigma(w_{13} x^{(1)} + w_{23} x^{(2)} + b_3)$$

$$o_4 = \sigma(w_{14} x^{(1)} + w_{24} x^{(2)} + b_4)$$

$$o_5 = \sigma(w_{15} x^{(1)} + w_{25} x^{(2)} + b_5)$$

$$o_6 = \sigma(w_{36} o_3 + w_{46} o_4 + w_{56} o_5 + b_6)$$

$$o_7 = \sigma(w_{37} o_3 + w_{47} o_4 + w_{57} o_5 + b_7)$$

First, it's important to **understand the information flow**: this network **computes** two outputs from its inputs.

To simplify the figure, I have opted not to display the bias terms, though they remain crucial components. Specifically, b_3 represents the bias term associated with node 3.

If bias terms were not significant, the training process would naturally reduce them to zero. Bias terms are essential as they enable the adjustment of the decision boundary, allowing the model to learn more complex patterns that weights alone cannot capture. By offering additional degrees of freedom, they also contribute to faster convergence during training.

Forward Pass (Computation)

```
In [2]: import numpy as np
```

```

# Sigmoid function

def sigma(x):
    return 1 / (1 + np.exp(-x))

# Input (two attributes) vector, one example of our training set

x1, x2 = (0.5, 0.9)

# Initializing the weights of layers 2 and 3 to random values

w13, w14, w15, w23, w24, w25 = np.random.uniform(low=-1, high=1, size=
w36, w46, w56, w37, w47, w57 = np.random.uniform(low=-1, high=1, size=

# Initializing all 5 bias terms to random values

b3, b4, b5, b6, b7 = np.random.uniform(low=-1, high=1, size=5)

o3 = sigma(w13 * x1 + w23 * x2 + b3)
o4 = sigma(w14 * x1 + w24 * x2 + b4)
o5 = sigma(w15 * x1 + w25 * x2 + b5)
o6 = sigma(w36 * o3 + w46 * o4 + w56 * o5 + b6)
o7 = sigma(w37 * o3 + w47 * o4 + w57 * o5 + b7)

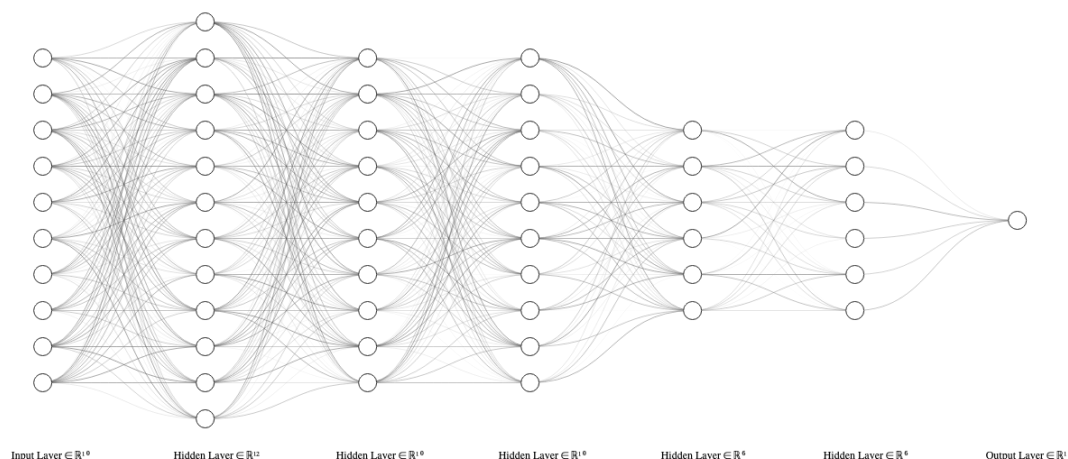
(o6, o7)

```

(0.5027642045632127, 0.39398584263466113)

The example above illustrates the computation process with specific values. Before training a neural network, it is standard practice to initialize the weights and biases with random values. Gradient descent is then employed to iteratively adjust these parameters, aiming to minimize the loss function.

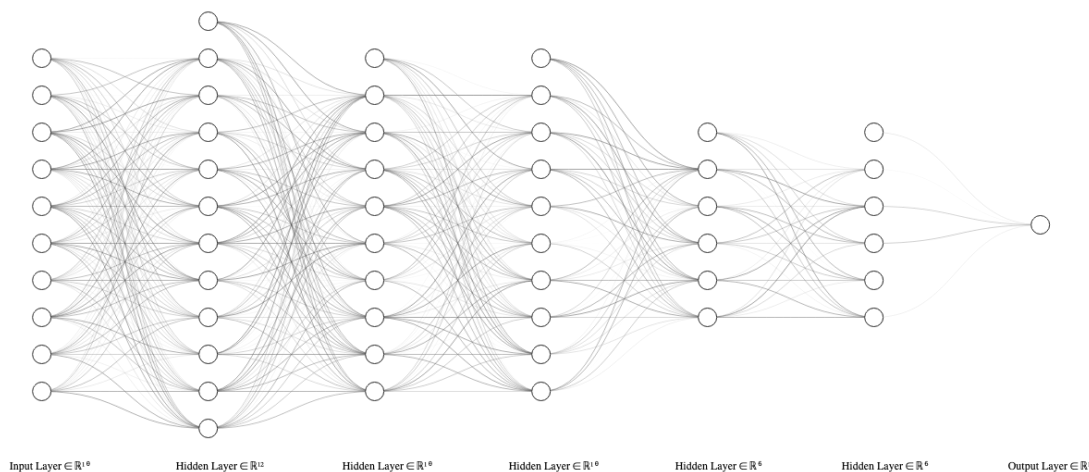
Forward Pass (Computatation)



The information flow remains consistent even in more **complex networks**.
 Networks with many layers are called **deep neural networks** (DNN).

Produced using **NN-SVG**, LeNail (2019).

Forward Pass (Computatation)



Same network with **bias terms shown**.

Produced using **NN-SVG**, LeNail (2019).

Activation Function

- As will be discussed later, the training algorithm, known as **backpropagation**, employs **gradient descent**, necessitating the calculation of the **partial derivatives of the loss function**.
- The **step function** in the multilayer perceptron had to be replaced, as it consists only of flat surfaces. **Gradient descent cannot progress on flat surfaces due to their zero derivative**.

Activation Function

- Nonlinear activation functions are paramount** because, without them, multiple layers in the network would only compute a linear function of the inputs.
- According to the **Universal Approximation Theorem**, sufficiently large deep networks with nonlinear activation functions can **approximate any**

continuous function. See [Universal Approximation Theorem](#).

Sigmoid

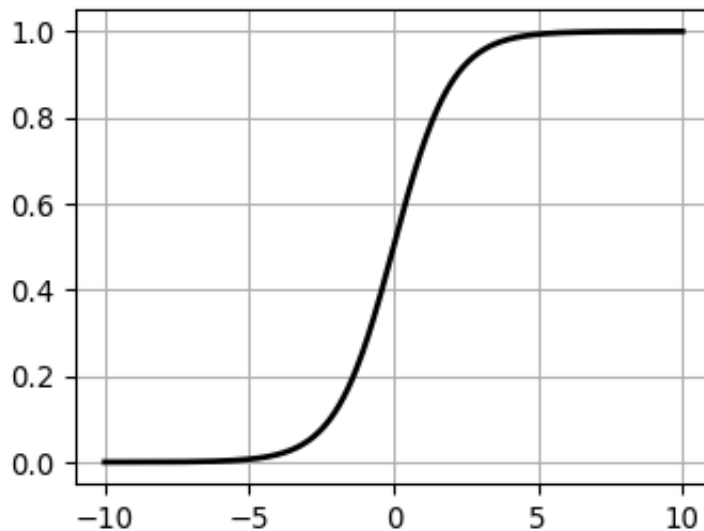
```
In [3]: import numpy as np
import matplotlib.pyplot as plt

# Sigmoid function
def sigmoid(x):
    return 1 / (1 + np.exp(-x))

# Generate x values
x = np.linspace(-10, 10, 400)

# Compute y values for the sigmoid function
y = sigmoid(x)

plt.figure(figsize=(4,3))
plt.plot(x, y, color='black', linewidth=2)
plt.grid(True)
plt.show()
plt.show()
```



$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

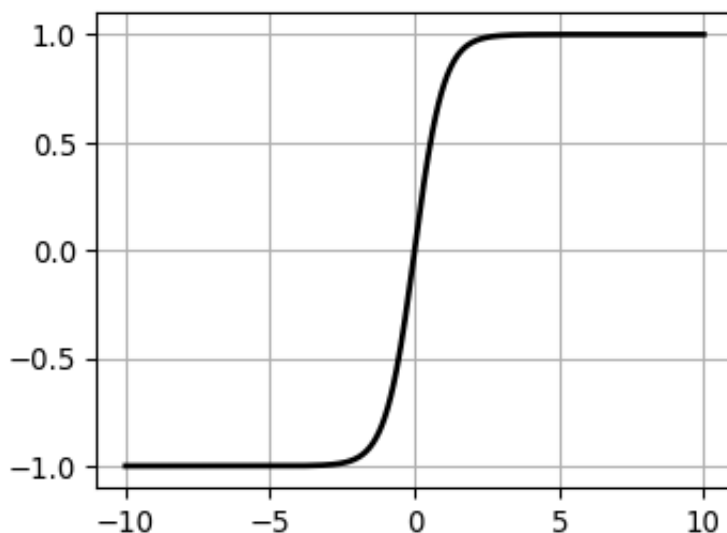
Hyperbolic Tangent Function

```
In [4]: # Compute y values for the hyperbolic tangent function

y = np.tanh(x)

plt.figure(figsize=(4,3))
plt.plot(x, y, color='black', linewidth=2)
```

```
plt.grid(True)
plt.show()
```

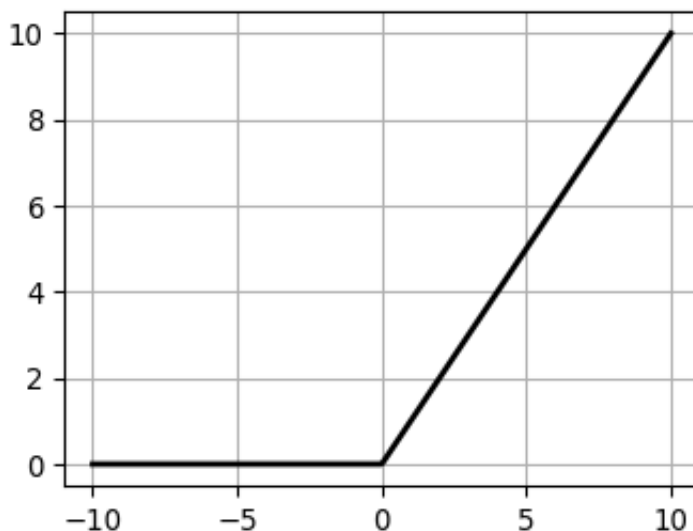


Hyperbolic tangent ($\tanh(t) = 2 \cdot \sigma(2t) - 1$) is an S-shaped curve, similar to the sigmoid function, producing output values ranging from -1 to 1. According to Géron (2022), this range helps each layer's output to be approximately centered around 0 at the start of training, thereby **accelerating convergence**.

Rectified linear unit function (ReLU)

```
In [5]: # Compute y values for the rectified linear unit function (ReLU) funct
y = np.maximum(0, x)

plt.figure(figsize=(4,3))
plt.plot(x, y, color='black', linewidth=2)
plt.grid(True)
plt.show()
```



Although the **ReLU function** ($\mathrm{ReLU}(t) = \max(0, t)$) is not differentiable at $t=0$ and has a derivative of 0 for $t<0$, it performs quite well in practice and is computationally efficient. Consequently, it has become the **default activation function**.

Common Activation Functions

```
In [6]: from scipy.special import expit as sigmoid

def relu(z):
    return np.maximum(0, z)

def derivative(f, z, eps=0.000001):
    return (f(z + eps) - f(z - eps))/(2 * eps)

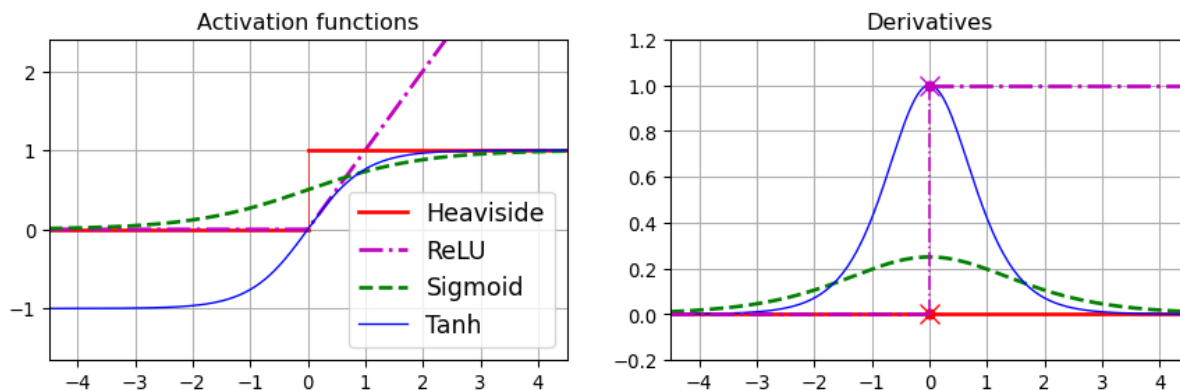
max_z = 4.5
z = np.linspace(-max_z, max_z, 200)

plt.figure(figsize=(11, 3.1))

plt.subplot(121)
plt.plot([-max_z, 0], [0, 0], "r-", linewidth=2, label="Heaviside")
plt.plot(z, relu(z), "m-.", linewidth=2, label="ReLU")
plt.plot([0, 0], [0, 1], "r-", linewidth=0.5)
plt.plot([0, max_z], [1, 1], "r-", linewidth=2)
plt.plot(z, sigmoid(z), "g--", linewidth=2, label="Sigmoid")
plt.plot(z, np.tanh(z), "b-", linewidth=1, label="Tanh")
plt.grid(True)
plt.title("Activation functions")
plt.axis([-max_z, max_z, -1.65, 2.4])
plt.gca().set_yticks([-1, 0, 1, 2])
plt.legend(loc="lower right", fontsize=13)

plt.subplot(122)
plt.plot(z, derivative(np.sign, z), "r-", linewidth=2, label="Heavisid")
plt.plot(0, 0, "ro", markersize=5)
plt.plot(0, 0, "rx", markersize=10)
plt.plot(z, derivative(sigmoid, z), "g--", linewidth=2, label="Sigmoid")
plt.plot(z, derivative(np.tanh, z), "b-", linewidth=1, label="Tanh")
plt.plot([-max_z, 0], [0, 0], "m-.", linewidth=2)
plt.plot([0, max_z], [1, 1], "m-.", linewidth=2)
plt.plot([0, 0], [0, 1], "m-.", linewidth=1.2)
plt.plot(0, 1, "mo", markersize=5)
plt.plot(0, 1, "mx", markersize=10)
plt.grid(True)
plt.title("Derivatives")
plt.axis([-max_z, max_z, -0.2, 1.2])

plt.show()
```



Géron (2022) – [10_neural_nets_with_keras.ipynb](#)

Universal Approximation

Definition

The **Universal Approximation Theorem (UAT)** states that a feedforward neural network with a *single hidden layer* containing a *finite number of neurons* can **approximate any continuous function** on a compact subset of $\mathbb{R}^n$, given appropriate weights and activation functions.

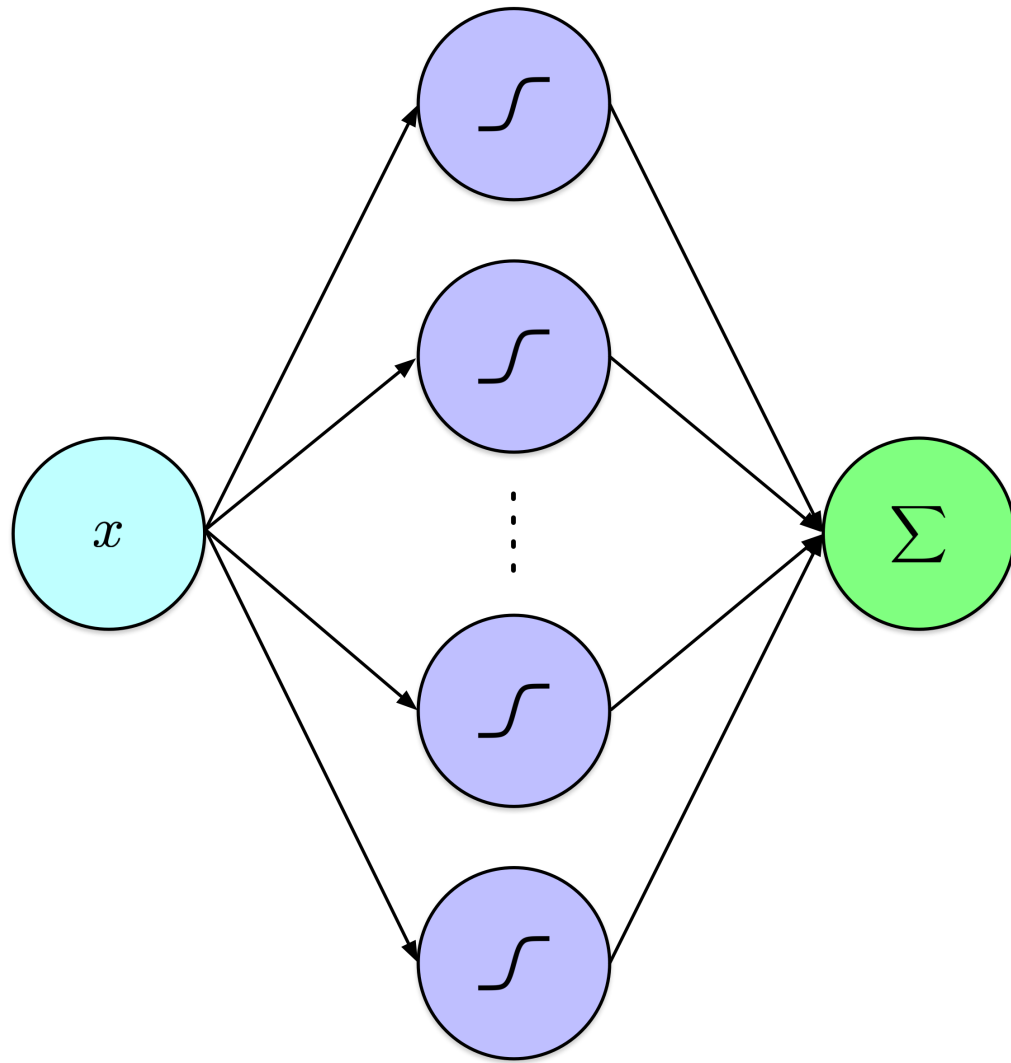
Cybenko (1989); Hornik, Stinchcombe, and White (1989)

In mathematical terms, a subset of $\mathbb{R}^n$ is considered **compact** if it is both **closed** and **bounded**.

- **Closed:** A set is closed if it contains all its boundary points. In other words, it includes its limit points or accumulation points.
- **Bounded:** A set is bounded if there exists a real number (M) such that the distance between any two points in the set is less than M .

In the context of the universal approximation theorem, compactness ensures that the function being approximated is defined on a finite and well-behaved region, which is crucial for the theoretical guarantees provided by the theorem.

Single Hidden Layer



Input Layer Hidden Layer Output Layer

$$y = \sum_{i=1}^N \alpha_i \sigma(w_{1,i} x + b_i)$$

Notation adapted to follow that of Cybenko (1989).

Effect of Varying w

```
In [7]: def logistic(x, w, b):
        """Compute the logistic function with parameters w and b."""
        return 1 / (1 + np.exp(-(w * x + b)))

        # Define a range for x values.
        x = np.linspace(-10, 10, 400)

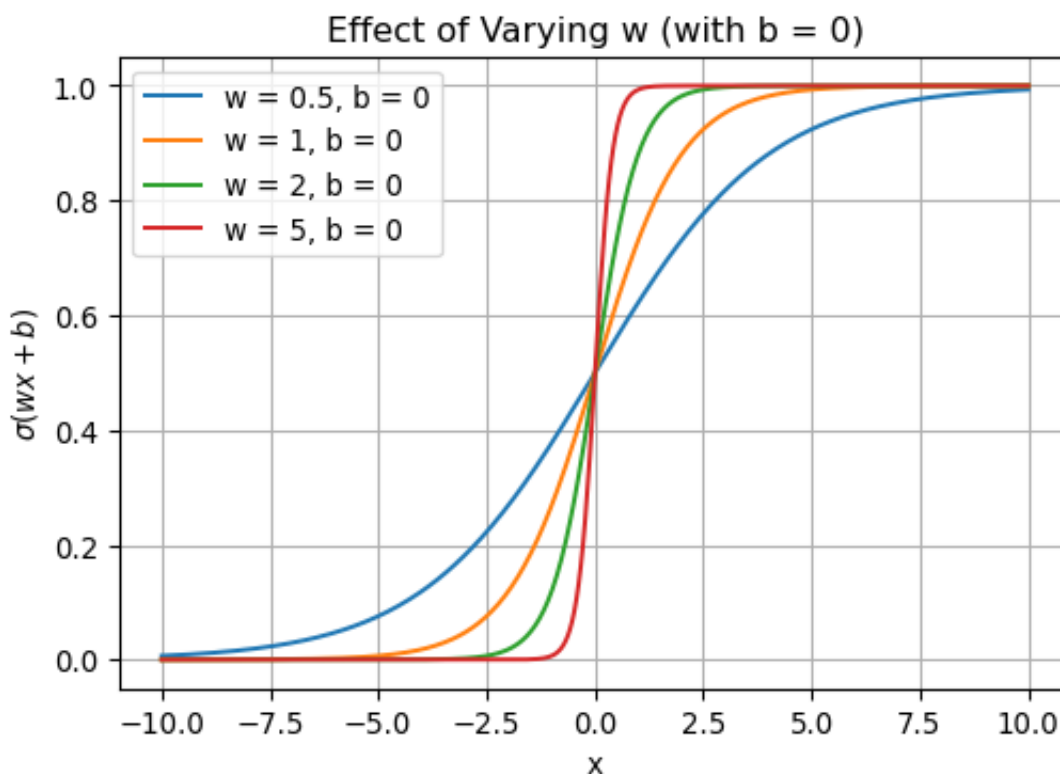
        # Plot 1: Varying w (steepness) with b fixed at 0.
        plt.figure(figsize=(6,4))
        w_values = [0.5, 1, 2, 5] # different steepness values
        b = 0 # fixed bias
```

```

for w in w_values:
    plt.plot(x, logistic(x, w, b), label=f'w = {w}, b = {b}')
plt.title('Effect of Varying w (with b = 0)')
plt.xlabel('x')
plt.ylabel(r'$\sigma(wx+b)$')
plt.legend()
plt.grid(True)

plt.show()

```



Sigmoid activation function: $\sigma(wx+b)$.

Effect of Varying b

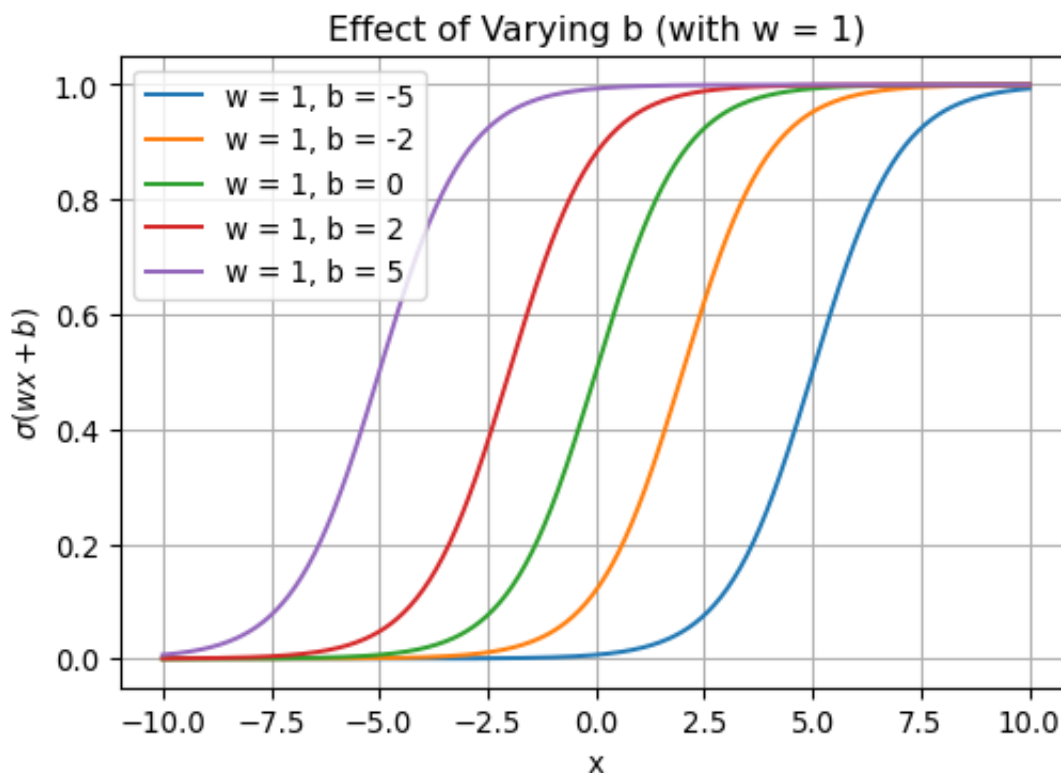
```

In [8]: # Plot 2: Varying b (horizontal shift) with w fixed at 1.
plt.figure(figsize=(6,4))
w = 1 # fixed steepness
b_values = [-5, -2, 0, 2, 5] # different bias values

for b in b_values:
    plt.plot(x, logistic(x, w, b), label=f'w = {w}, b = {b}')
plt.title('Effect of Varying b (with w = 1)')
plt.xlabel('x')
plt.ylabel(r'$\sigma(wx+b)$')
plt.legend()
plt.grid(True)

```

```
plt.show()
```



Sigmoid activation function: $\sigma(wx+b)$.

Effect of Varying w

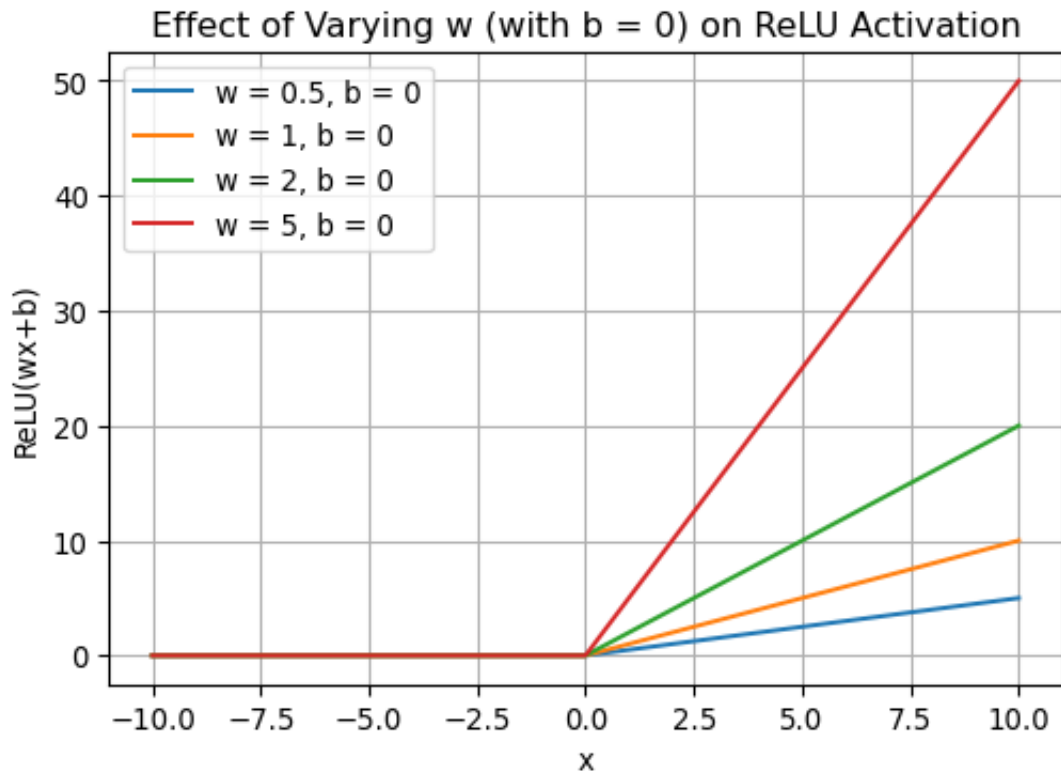
```
In [9]: def relu(x, w, b):
        """Compute the ReLU activation with parameters w and b."""
        return np.maximum(0, w * x + b)

        # Define a range for x values.
        x = np.linspace(-10, 10, 400)

        # Plot 1: Varying w (scaling) with b fixed at 0.
        plt.figure(figsize=(6,4))
        w_values = [0.5, 1, 2, 5] # different scaling values
        b = 0 # fixed bias

        for w in w_values:
            plt.plot(x, relu(x, w, b), label=f'w = {w}, b = {b}')
        plt.title('Effect of Varying w (with b = 0) on ReLU Activation')
        plt.xlabel('x')
        plt.ylabel('ReLU(wx+b)')
        plt.legend()
        plt.grid(True)

        plt.show()
```



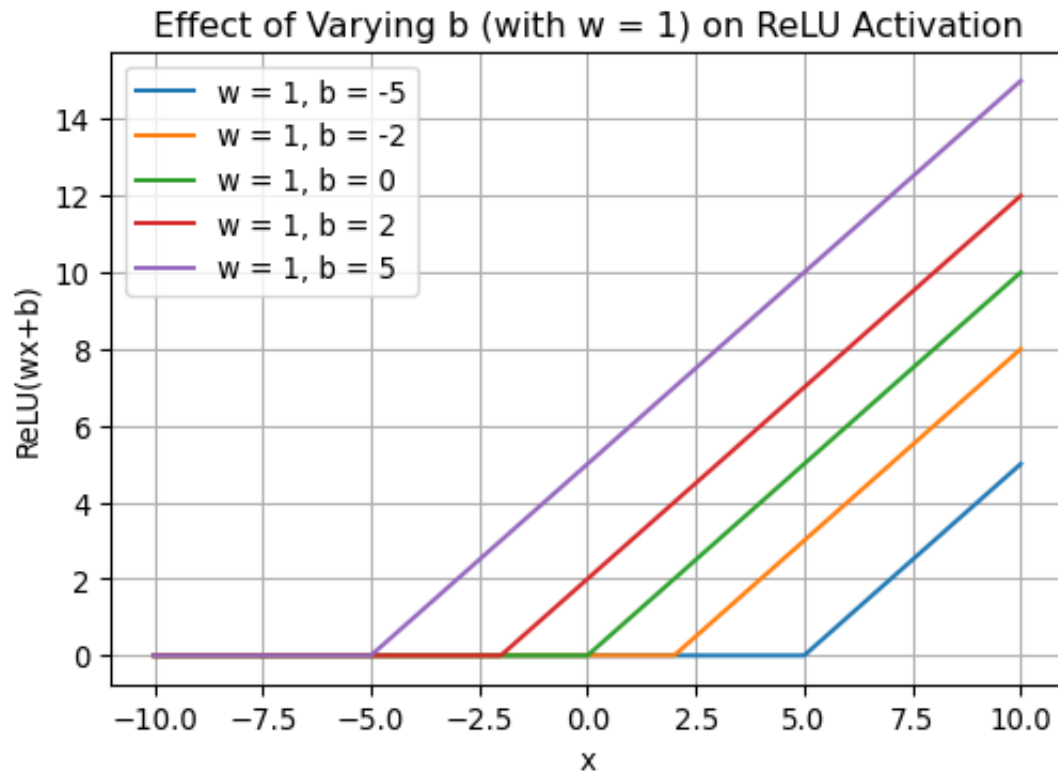
ReLU activation function: `np.maximum(0, w * x + b)`.

Effect of Varying b

```
In [10]: # Plot 2: Varying b (horizontal shift) with w fixed at 1.
plt.figure(figsize=(6,4))
w = 1 # fixed scaling
b_values = [-5, -2, 0, 2, 5] # different bias values

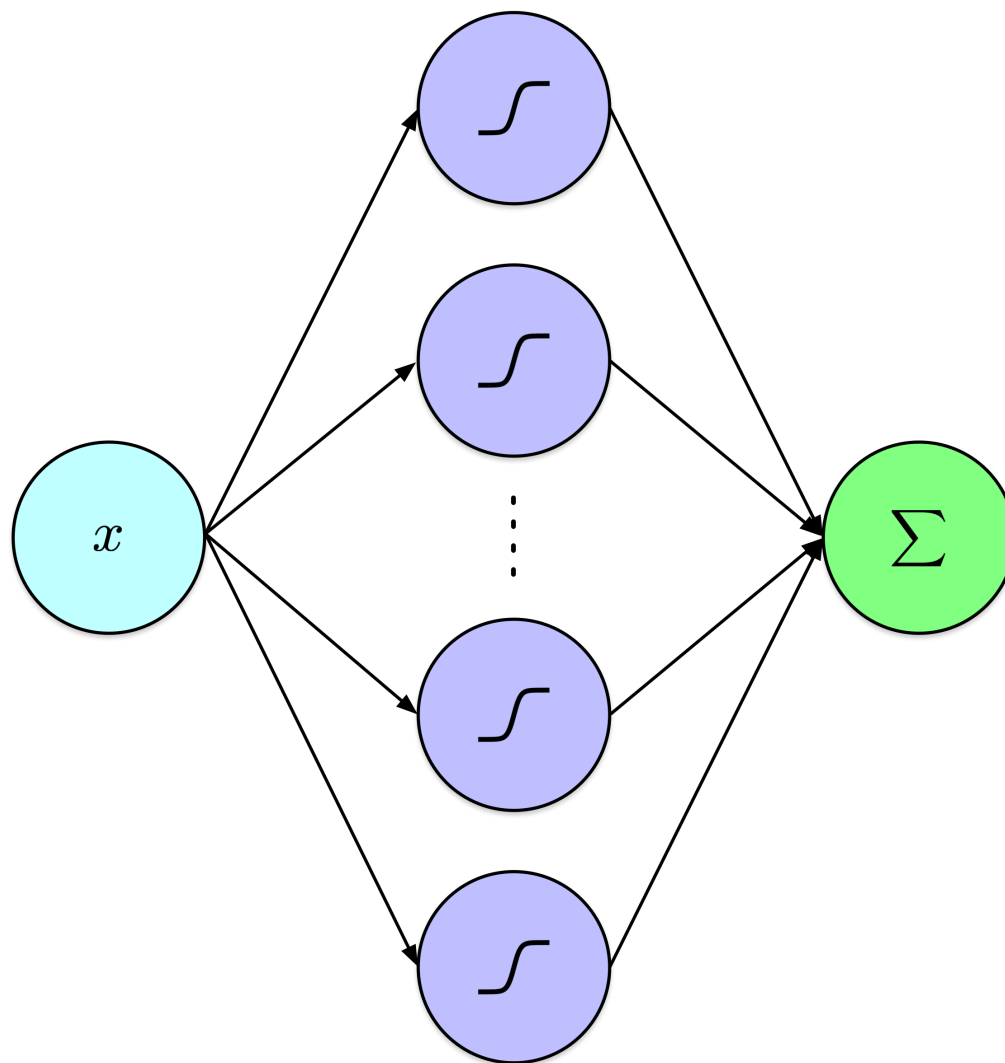
for b in b_values:
    plt.plot(x, relu(x, w, b), label=f'w = {w}, b = {b}')
plt.title('Effect of Varying b (with w = 1) on ReLU Activation')
plt.xlabel('x')
plt.ylabel('ReLU(wx+b)')
plt.legend()
plt.grid(True)

plt.show()
```



ReLU activation function: `np.maximum(0, w * x + b)` .

Single Hidden Layer



Input Layer Hidden Layer Output Layer

$$y = \sum_{i=1}^N \alpha_i \sigma(w_{1,i} x + b_i)$$

See also: Chapter 4: [A visual proof that neural nets can compute any function](#), [Neural Networks and Deep Learning](#) by Michael Nielsen.

Demonstration with Code

```
In [11]: # Defining the function to be approximated

def f(x):
    return 2 * x**3 + 4 * x**2 - 5 * x + 1

# Generating a dataset, x in [-4,2), f(x) as above

X = 6 * np.random.rand(1000, 1) - 4
```



```
y = f(X).flatten()
```

Increasing the Number of Neurons

```
In [12]: from sklearn.neural_network import MLPRegressor
from sklearn.model_selection import train_test_split

X_train, X_valid, y_train, y_valid = train_test_split(X, y, test_size=
models = []

sizes = [1, 2, 5, 10, 100]

for i, n in enumerate(sizes):

    models.append(MLPRegressor(hidden_layer_sizes=[n], max_iter=5000, ra

    models[i].fit(X_train, y_train)
```

`MLPRegressor` is a multi-layer perceptron regressor from sklearn. Its default activation function is `relu`.

Increasing the Number of Neurons

```
In [13]: # Create a colormap
colors = plt.colormaps['cool'].resampled(len(sizes))

X_valid = np.sort(X_valid,axis=0)

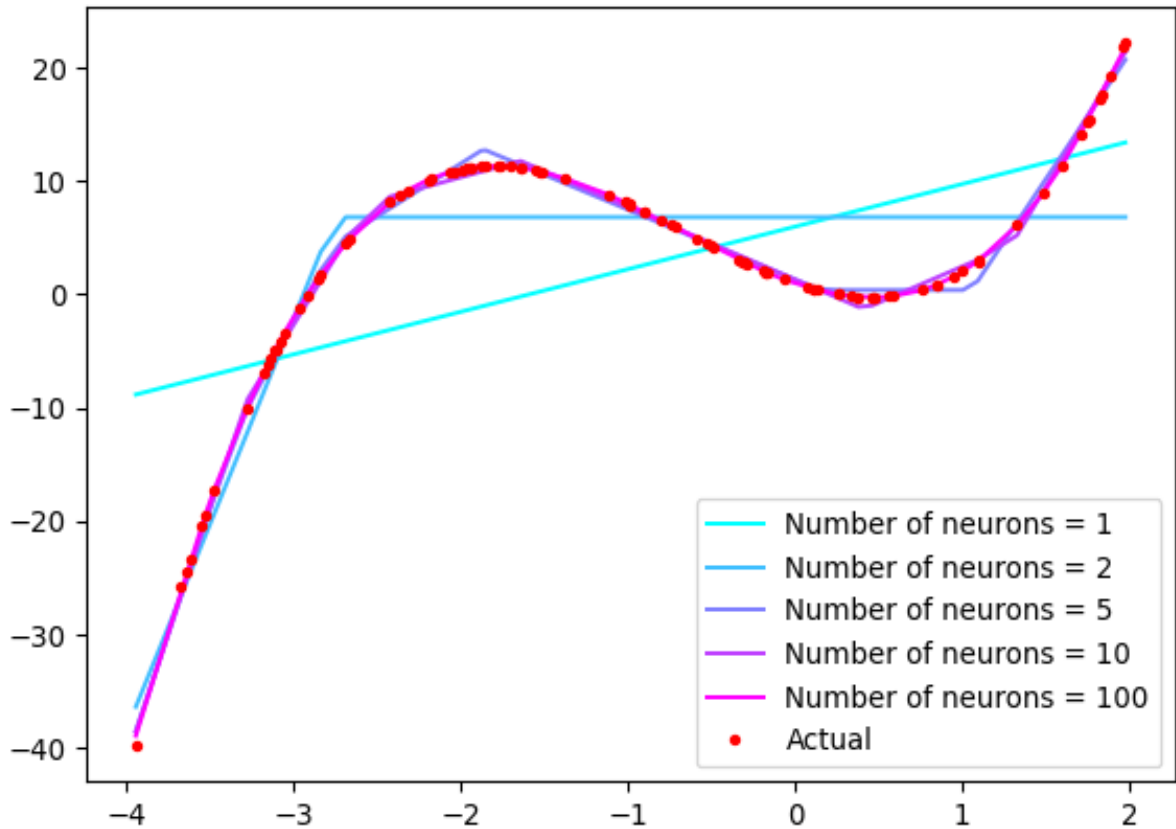
for i, n in enumerate(sizes):

    y_pred = models[i].predict(X_valid)

    plt.plot(X_valid, y_pred, "-", color=colors(i), label="Number of neu

y_true = f(X_valid)
plt.plot(X_valid, y_true, "r.", label='Actual')

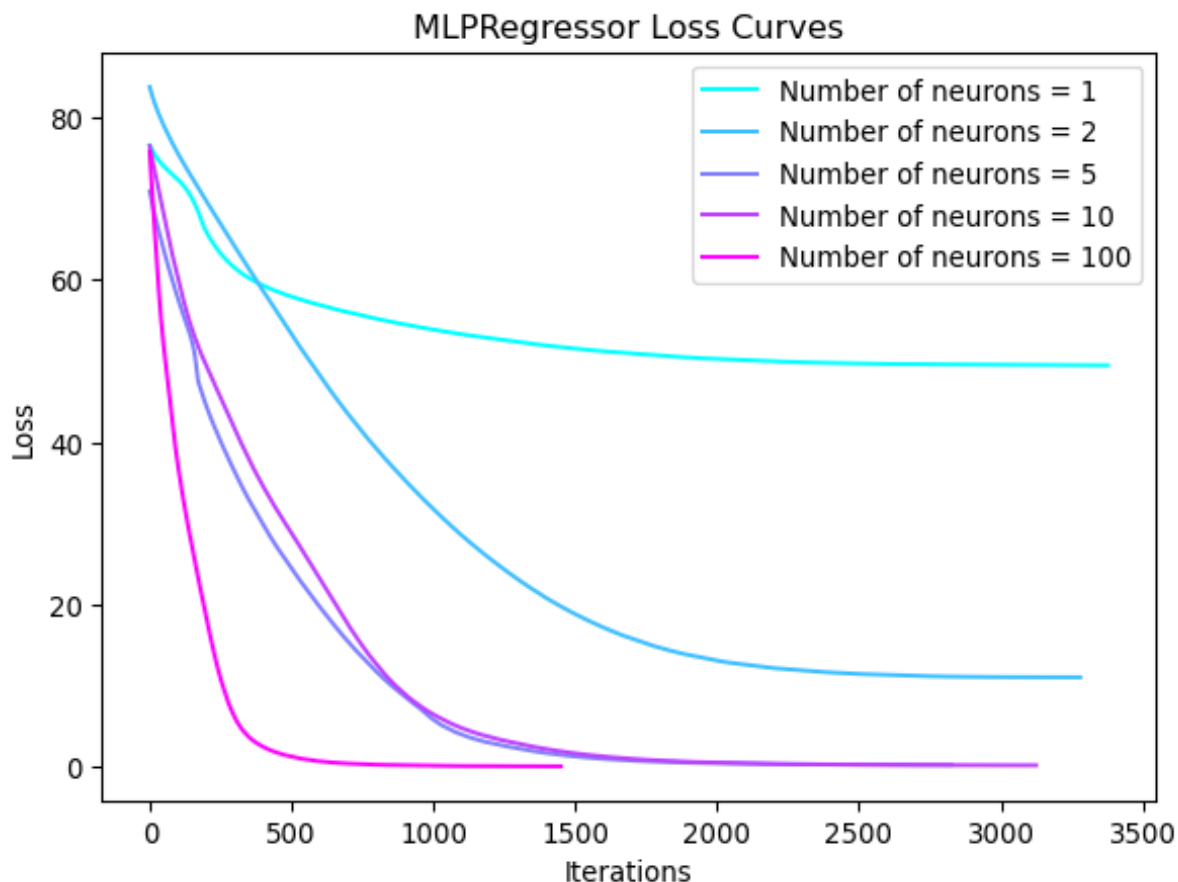
plt.legend()
plt.show()
```



In the example above, I retained only 10% of the data as the test set because the function being approximated is straightforward and noise-free. This decision was made to ensure that the true curve does not overshadow the other results.

Increasing the Number of Neurons

```
In [14]: for i, n in enumerate(sizes):  
    plt.plot(models[i].loss_curve_, "-", color=colors(i), label="Number  
  
plt.title('MLPRegressor Loss Curves')  
plt.xlabel('Iterations')  
plt.ylabel('Loss')  
  
plt.legend()  
plt.show()
```



As expected, increasing neuron count reduces loss.

Universal Approximation

<https://youtu.be/CqOfi41LfDw>

This video effectively conveys the underlying intuition of the universal approximation theorem. (18m 53s)

The video effectively elucidates key concepts (terminology) in neural networks, including nodes, layers, weights, and activation functions. It demonstrates the process of summing activation outputs from a preceding layer, akin to the aggregation of curves. Additionally, the video illustrates how scaling an output by a weight not only alters the amplitude of a curve but also inverts its orientation when the weight is negative. Moreover, it clearly depicts the function of bias terms in vertically shifting the curve, contingent on the sign of the bias.

Let's Code

Frameworks

[PyTorch](#) and [TensorFlow](#) are the leading platforms for deep learning.

- PyTorch has gained considerable traction in the **research community**. Initially developed by **Meta AI**, it is now part of the Linux Foundation.
- TensorFlow, created by **Google**, is widely adopted in **industry** for deploying models in production environments.

Keras

[Keras](#) is a high-level API designed to build, train, evaluate, and execute models across various backends, including PyTorch, TensorFlow, and [JAX](#), Google's high-performance platform.

[Keras](#) is powerful enough for most projects.

As highlighted in previous Quotes of the Day, François Chollet, a Google engineer, is the originator and one of the primary developers of the Keras project.

Fashion-MNIST dataset

"[Fashion-MNIST](#) is a dataset of [Zalando](#)'s article images—consisting of a training set of **60,000 examples** and a test set of **10,000 examples**. Each example is a **28x28 grayscale image**, associated with a label from **10 classes**."

Attribution: Géron (2022) – [10_neural_nets_with_keras.ipynb](#)

Loading

```
In [15]: import tensorflow as tf

fashion_mnist = tf.keras.datasets.fashion_mnist.load_data()

(X_train_full, y_train_full), (X_test, y_test) = fashion_mnist

X_train, y_train = X_train_full[:5000], y_train_full[:5000]
X_valid, y_valid = X_train_full[5000:], y_train_full[5000:]
```

Setting aside 5000 examples as a validation set.

Exploration

```
In [16]: X_train.shape  
(55000, 28, 28)  
...
```

```
In [17]: X_train.dtype  
dtype('uint8')  
...
```

Transforming the pixel intensities from integers in the range 0 to 255 to floats in the range 0 to 1.

```
In [18]: X_train, X_valid, X_test = X_train / 255., X_valid / 255., X_test / 255.
```

What are these Images Anyway?

```
In [19]: plt.figure(figsize=(2, 2))  
plt.imshow(X_train[0], cmap="binary")  
plt.axis('off')  
plt.show()
```



...

```
In [20]: y_train  
array([9, 0, 0, ..., 9, 0, 2], dtype=uint8)  
...
```

Since the labels are integers, 0 to 9. Class names will become handy.

```
In [21]: class_names = ["T-shirt/top", "Trouser", "Pullover", "Dress", "Coat",  
                        "Sandal", "Shirt", "Sneaker", "Bag", "Ankle boot"]
```

First 40 Images

```
In [22]: n_rows = 4
n_cols = 10
plt.figure(figsize=(n_cols * 1.2, n_rows * 1.2))
for row in range(n_rows):
    for col in range(n_cols):
        index = n_cols * row + col
        plt.subplot(n_rows, n_cols, index + 1)
        plt.imshow(X_train[index], cmap="binary", interpolation="nearest")
        plt.axis('off')
        plt.title(class_names[y_train[index]])
plt.subplots_adjust(wspace=0.2, hspace=0.5)
plt.show()
```



Creating a Model

```
In [23]: tf.random.set_seed(42)

model = tf.keras.Sequential()

model.add(tf.keras.layers.InputLayer(shape=[28, 28]))
model.add(tf.keras.layers.Flatten())
model.add(tf.keras.layers.Dense(300, activation="relu"))
model.add(tf.keras.layers.Dense(100, activation="relu"))
model.add(tf.keras.layers.Dense(10, activation="softmax"))
```

```
model.summary()
```

```
In [24]: model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	
flatten (Flatten)	(None, 784)	
dense (Dense)	(None, 300)	
dense_1 (Dense)	(None, 100)	
dense_2 (Dense)	(None, 10)	

Total params: 266,610 (1.02 MB)

Trainable params: 266,610 (1.02 MB)

Non-trainable params: 0 (0.00 B)

As observed, `dense_3` has 235,500 parameters, while $784 \times 300 = 235,200$.

Could you explain the origin of the additional parameters?

Similarly, `dense_3` has 30,100 parameters, while $300 \times 100 = 30,000$.

Can you explain why?

Creating a Model (Alternative)

```
In [25]: # extra code – clear the session to reset the name counters
tf.keras.backend.clear_session()
tf.random.set_seed(42)
```

```
In [26]: model = tf.keras.Sequential([
    tf.keras.layers.Flatten(input_shape=[28, 28]),
    tf.keras.layers.Dense(300, activation="relu"),
    tf.keras.layers.Dense(100, activation="relu"),
    tf.keras.layers.Dense(10, activation="softmax")
])
```

/Users/turcotte/opt/micromamba/envs/ml4bio/lib/python3.10/site-packages/keras/src/layers/reshaping/flatten.py:37: UserWarning:

Do not pass an `input\_shape`/`input\_dim` argument to a layer. When using Sequential models, prefer using an `Input(shape)` object as the first layer in the model instead.

```
model.summary()
```

```
In [27]: model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	
flatten (Flatten)	(None, 784)	
dense (Dense)	(None, 300)	
dense_1 (Dense)	(None, 100)	
dense_2 (Dense)	(None, 10)	

Total params: 266,610 (1.02 MB)

Trainable params: 266,610 (1.02 MB)

Non-trainable params: 0 (0.00 B)

Compiling the Model

```
In [28]: model.compile(loss="sparse_categorical_crossentropy",
                      optimizer="sgd",
                      metrics=["accuracy"])
```

`sparse_categorical_crossentropy` is the appropriate function for a multiclass classification problem (more later).

The method `compile` allows to set the loss function, as well as other parameters. Keras then prepares the model for training.

Training the Model

```
In [29]: history = model.fit(X_train, y_train, epochs=30,
                           validation_data=(X_valid, y_valid))
```

Epoch 1/30

```
1/1719 _____ 3:28 122ms/step - accuracy: 0.0625 - loss: 2.6486
63/1719 _____ 1s 812us/step - accuracy: 0.2777 - loss: 2.1578
129/1719 _____ 1s 787us/step - accuracy: 0.3854 - loss: 1.9433
195/1719 _____ 1s 779us/step - accuracy: 0.4432 - loss: 1.7913
262/1719 _____ 1s 772us/step - accuracy: 0.4824 - loss: 1.6747
329/1719 _____ 1s 768us/step - accuracy: 0.5108 - loss: 1.5833
397/1719 _____ 1s 763us/step - accuracy: 0.5271 - loss: 1.5408
```



```
cy: 0.5330 - loss: 1.5084
463/1719 0s 764u
s/step - accuracy: 0.5507 - loss: 1.4479
529/1719 0s 765us/step - accuracy: 0.5656 - loss: 1.3965
596/1719 0s 763us/step - accuracy: 0.5786 - loss: 1.3514
662/1719 0s 763us/step - accuracy: 0.5899 - loss: 1.3126
728/1719 0s 763us/step - accuracy: 0.5999 - loss: 1.2781
794/1719 0s 763u
s/step - accuracy: 0.6089 - loss: 1.2473
859/1719 0s 764us/step - accuracy: 0.6169 - loss: 1.2198
925/1719 0s 764us/step - accuracy: 0.6244 - loss: 1.1943
991/1719 0s 764us/step - accuracy: 0.6312 - loss: 1.1709
1056/1719 0s 765us/step - accuracy: 0.6374 - loss: 1.1497
1123/1719 0s 764u
s/step - accuracy: 0.6433 - loss: 1.1296
1189/1719 0s 764us/step - accuracy: 0.6487 - loss: 1.1111
1254/1719 0s 765us/step - accuracy: 0.6537 - loss: 1.0944
1319/1719 0s 765us/step - accuracy: 0.6583 - loss: 1.0786
1386/1719 0s 765us/step - accuracy: 0.6628 - loss: 1.0634
1453/1719 0s 764u
s/step - accuracy: 0.6670 - loss: 1.0491
1518/1719 0s 765us/step - accuracy: 0.6708 - loss: 1.0360
1584/1719 0s 765us/step - accuracy: 0.6745 - loss: 1.0234
1651/1719 0s 765us/step - accuracy: 0.6781 - loss: 1.0114
1714/1719 0s 766us/step - accuracy: 0.6812 - loss: 1.0006
1719/1719 2s 860u
s/step - accuracy: 0.6815 - loss: 0.9996 - val_accuracy: 0.8278 - val_
loss: 0.5071
Epoch 2/30
1/1719 18s 11ms/step - accuracy: 0.8750 - loss:
0.4716
65/1719 1s 785us/step - accuracy:
```

Page 44 of 83

Page 45 of 83

Page 46 of 83

```

1225/1719 _____
- 0s 741us/step - accuracy: 0.8535 - loss: 0.4245
1291/1719 _____
0s 743us/step - accuracy: 0.8536 - loss: 0.4242
1356/1719 _____ 0s 744us/step - accuracy: 0.8537 - loss:
0.4239
1420/1719 _____ 0s 746us/step - accuracy:
0.8538 - loss: 0.4235
1485/1719 _____ 0s 747us/ste
p - accuracy: 0.8539 - loss: 0.4232
1549/1719 _____
- 0s 749us/step - accuracy: 0.8540 - loss: 0.4228
1614/1719 _____
0s 750us/step - accuracy: 0.8541 - loss: 0.4225
1679/1719 _____ 0s 751us/step - accuracy: 0.8542 - loss:
0.4222
1719/1719 _____ 1s 810us/step - accuracy:
0.8543 - loss: 0.4220 - val_accuracy: 0.8488 - val_loss: 0.4176
Epoch 5/30
1/1719 _____ 18s 11ms/step - accuracy: 0.8438 - loss:
0.4341
65/1719 _____ 1s 784us/step - accuracy:
0.8733 - loss: 0.3710
131/1719 _____ 1s 774us/ste
p - accuracy: 0.8668 - loss: 0.3883
197/1719 _____
- 1s 770us/step - accuracy: 0.8629 - loss: 0.3969
264/1719 _____
1s 767us/step - accuracy: 0.8615 - loss: 0.3996
330/1719 _____ 1s 767us/step - accuracy: 0.8606 - loss:
0.4012
397/1719 _____ 1s 765us/step - accuracy:
0.8600 - loss: 0.4024
463/1719 _____ 0s 765us/ste
p - accuracy: 0.8597 - loss: 0.4030
529/1719 _____
- 0s 764us/step - accuracy: 0.8596 - loss: 0.4034
595/1719 _____
0s 764us/step - accuracy: 0.8595 - loss: 0.4036
662/1719 _____ 0s 763us/step - accuracy: 0.8596 - loss:
0.4035
727/1719 _____ 0s 763us/step - accuracy:
0.8598 - loss: 0.4035
793/1719 _____ 0s 763us/ste
p - accuracy: 0.8599 - loss: 0.4035
859/1719 _____
- 0s 763us/step - accuracy: 0.8600 - loss: 0.4034
926/1719 _____
```

Page 48 of 83

Epoch 7/30

Page 49 of 83

Page 50 of 83

136/1719 _____ 1s 748us/ste
p - accuracy: 0.8820 - loss: 0.3428
205/1719 _____
- 1s 743us/step - accuracy: 0.8792 - loss: 0.3506
273/1719 _____
_____ 1s 742us/step - accuracy: 0.8779 - loss: 0.3533
341/1719 _____ 1s 742us/step - accuracy: 0.8769 - loss:
0.3550
409/1719 _____ 0s 741us/step - accuracy:
0.8763 - loss: 0.3564
478/1719 _____ 0s 739us/ste
p - accuracy: 0.8760 - loss: 0.3572
547/1719 _____
- 0s 738us/step - accuracy: 0.8757 - loss: 0.3577
616/1719 _____
_____ 0s 737us/step - accuracy: 0.8756 - loss: 0.3578
685/1719 _____ 0s 737us/step - accuracy: 0.8756 - loss:
0.3578
752/1719 _____ 0s 737us/step - accuracy:
0.8756 - loss: 0.3579
821/1719 _____ 0s 737us/ste
p - accuracy: 0.8756 - loss: 0.3580
890/1719 _____
- 0s 737us/step - accuracy: 0.8755 - loss: 0.3581
957/1719 _____
_____ 0s 738us/step - accuracy: 0.8755 - loss: 0.3581
1026/1719 _____ 0s 738us/step - accuracy: 0.8755 - loss:
0.3580
1095/1719 _____ 0s 737us/step - accuracy:
0.8756 - loss: 0.3578
1164/1719 _____ 0s 737us/ste
p - accuracy: 0.8756 - loss: 0.3575
1232/1719 _____
- 0s 737us/step - accuracy: 0.8757 - loss: 0.3574
1299/1719 _____
_____ 0s 738us/step - accuracy: 0.8757 - loss: 0.3572
1367/1719 _____ 0s 738us/step - accuracy: 0.8757 - loss:
0.3570
1436/1719 _____ 0s 737us/step - accuracy:
0.8757 - loss: 0.3568
1505/1719 _____ 0s 737us/ste
p - accuracy: 0.8758 - loss: 0.3566
1574/1719 _____
- 0s 736us/step - accuracy: 0.8758 - loss: 0.3564
1643/1719 _____
_____ 0s 736us/step - accuracy: 0.8758 - loss: 0.3562
1712/1719 _____ 0s 736us/step - accuracy: 0.8758 - loss:

Page 52 of 83

```

1467/1719 _____ 0s 756us/ste
p - accuracy: 0.8789 - loss: 0.3457
1532/1719 _____
- 0s 756us/step - accuracy: 0.8789 - loss: 0.3455
1597/1719 _____
_____ 0s 757us/step - accuracy: 0.8789 - loss: 0.3454
1665/1719 _____ 0s 756us/step - accuracy: 0.8790 - loss:
0.3452
1719/1719 _____ 1s 812us/step - accuracy:
0.8790 - loss: 0.3451 - val_accuracy: 0.8620 - val_loss: 0.3739
Epoch 10/30
  1/1719 _____ 18s 11ms/step - accuracy: 0.8750 - loss:
0.3685
  66/1719 _____ 1s 777us/step - accuracy:
0.8917 - loss: 0.3041
 132/1719 _____ 1s 769us/ste
p - accuracy: 0.8868 - loss: 0.3192
 199/1719 _____
- 1s 765us/step - accuracy: 0.8839 - loss: 0.3277
 265/1719 _____
_____ 1s 765us/step - accuracy: 0.8829 - loss: 0.3307
 330/1719 _____ 1s 766us/step - accuracy: 0.8820 - loss:
0.3326
 397/1719 _____ 1s 764us/step - accuracy:
0.8815 - loss: 0.3342
 463/1719 _____ 0s 764us/ste
p - accuracy: 0.8813 - loss: 0.3352
 529/1719 _____
- 0s 763us/step - accuracy: 0.8811 - loss: 0.3358
 595/1719 _____
_____ 0s 763us/step - accuracy: 0.8810 - loss: 0.3362
 661/1719 _____ 0s 763us/step - accuracy: 0.8810 - loss:
0.3363
 727/1719 _____ 0s 764us/step - accuracy:
0.8811 - loss: 0.3364
 793/1719 _____ 0s 763us/ste
p - accuracy: 0.8811 - loss: 0.3365
 861/1719 _____
- 0s 762us/step - accuracy: 0.8810 - loss: 0.3367
 928/1719 _____
_____ 0s 761us/step - accuracy: 0.8810 - loss: 0.3368
 995/1719 _____ 0s 760us/step - accuracy: 0.8810 - loss:
0.3368
1062/1719 _____ 0s 759us/step - accuracy:
0.8811 - loss: 0.3367
1129/1719 _____ 0s 758us/ste
p - accuracy: 0.8812 - loss: 0.3365
1196/1719 _____
```

```
- 0s 758us/step - accuracy: 0.8812 - loss: 0.3363
1261/1719
0s 759us/step - accuracy: 0.8812 - loss: 0.3362
1328/1719
0s 758us/step - accuracy: 0.8813 - loss: 0.3361
1396/1719
0s 757us/step - accuracy: 0.8813 - loss: 0.3359
1462/1719
0s 757us/step - accuracy: 0.8814 - loss: 0.3357
1529/1719
- 0s 757us/step - accuracy: 0.8814 - loss: 0.3356
1597/1719
0s 757us/step - accuracy: 0.8814 - loss: 0.3354
1661/1719
0s 758us/step - accuracy: 0.8815 - loss: 0.3353
1719/1719
1s 819us/step - accuracy: 0.8815 - loss: 0.3352 - val_accuracy: 0.8640 - val_loss: 0.3690
Epoch 11/30
1/1719
19s 11ms/step - accuracy: 0.9062 - loss: 0.3526
65/1719
1s 791us/step - accuracy: 0.8956 - loss: 0.2945
130/1719
1s 781us/step - accuracy: 0.8895 - loss: 0.3093
194/1719
- 1s 782us/step - accuracy: 0.8864 - loss: 0.3179
262/1719
1s 772us/step - accuracy: 0.8854 - loss: 0.3211
330/1719
1s 765us/step - accuracy: 0.8846 - loss: 0.3232
396/1719
1s 764us/step - accuracy: 0.8842 - loss: 0.3248
462/1719
0s 764us/step - accuracy: 0.8839 - loss: 0.3258
527/1719
- 0s 765us/step - accuracy: 0.8838 - loss: 0.3265
593/1719
0s 765us/step - accuracy: 0.8837 - loss: 0.3269
658/1719
0s 765us/step - accuracy: 0.8838 - loss: 0.3270
724/1719
0s 766us/step - accuracy: 0.8838 - loss: 0.3271
789/1719
0s 766us/step - accuracy: 0.8838 - loss: 0.3273
855/1719
- 0s 766us/step - accuracy: 0.8838 - loss: 0.3274
923/1719
0s 765us/step - accuracy: 0.8838 - loss: 0.3276
```

```

992/1719 _____ 0s 763us/step - accuracy: 0.8838 - loss:
0.3276
1059/1719 _____ 0s 762us/step - accuracy:
0.8839 - loss: 0.3275
1126/1719 _____ 0s 761us/ste
p - accuracy: 0.8840 - loss: 0.3273
1193/1719 _____
- 0s 761us/step - accuracy: 0.8840 - loss: 0.3272
1260/1719 _____
_____ 0s 760us/step - accuracy: 0.8840 - loss: 0.3271
1325/1719 _____ 0s 761us/step - accuracy: 0.8841 - loss:
0.3270
1393/1719 _____ 0s 760us/step - accuracy:
0.8841 - loss: 0.3268
1461/1719 _____ 0s 759us/ste
p - accuracy: 0.8842 - loss: 0.3266
1527/1719 _____
- 0s 759us/step - accuracy: 0.8842 - loss: 0.3265
1594/1719 _____
_____ 0s 759us/step - accuracy: 0.8842 - loss: 0.3263
1661/1719 _____ 0s 759us/step - accuracy: 0.8843 - loss:
0.3262
1719/1719 _____ 1s 814us/step - accuracy:
0.8843 - loss: 0.3261 - val_accuracy: 0.8662 - val_loss: 0.3648
Epoch 12/30
  1/1719 _____ 19s 11ms/step - accuracy: 0.9062 - loss:
0.3407
  66/1719 _____ 1s 781us/step - accuracy:
0.9027 - loss: 0.2855
 131/1719 _____ 1s 777us/ste
p - accuracy: 0.8965 - loss: 0.3000
 196/1719 _____
- 1s 775us/step - accuracy: 0.8930 - loss: 0.3085
 264/1719 _____
_____ 1s 767us/step - accuracy: 0.8914 - loss: 0.3118
 331/1719 _____ 1s 765us/step - accuracy: 0.8902 - loss:
0.3139
 398/1719 _____ 1s 763us/step - accuracy:
0.8894 - loss: 0.3156
 463/1719 _____ 0s 764us/ste
p - accuracy: 0.8889 - loss: 0.3167
 531/1719 _____
- 0s 761us/step - accuracy: 0.8884 - loss: 0.3174
 598/1719 _____
_____ 0s 760us/step - accuracy: 0.8882 - loss: 0.3179
 659/1719 _____ 0s 766us/step - accuracy: 0.8881 - loss:
0.3180
```

Page 56 of 83

```
p - accuracy: 0.8903 - loss: 0.3084
529/1719
- 0s 763us/step - accuracy: 0.8900 - loss: 0.3092
595/1719
0s 763us/step - accuracy: 0.8897 - loss: 0.3097
660/1719
0s 764us/step - accuracy: 0.8897 - loss: 0.3098
726/1719
0s 765us/step - accuracy: 0.8897 - loss: 0.3100
792/1719
0s 765us/step - accuracy: 0.8896 - loss: 0.3102
859/1719
- 0s 764us/step - accuracy: 0.8895 - loss: 0.3104
925/1719
0s 764us/step - accuracy: 0.8894 - loss: 0.3106
987/1719
0s 767us/step - accuracy: 0.8894 - loss: 0.3106
1051/1719
0s 768us/step - accuracy: 0.8895 - loss: 0.3106
1115/1719
0s 769us/step - accuracy: 0.8895 - loss: 0.3105
1182/1719
- 0s 768us/step - accuracy: 0.8895 - loss: 0.3104
1248/1719
0s 768us/step - accuracy: 0.8895 - loss: 0.3103
1307/1719
0s 772us/step - accuracy: 0.8896 - loss: 0.3102
1351/1719
0s 784us/step - accuracy: 0.8896 - loss: 0.3102
1414/1719
0s 785us/step - accuracy: 0.8896 - loss: 0.3100
1478/1719
- 0s 785us/step - accuracy: 0.8897 - loss: 0.3099
1543/1719
0s 785us/step - accuracy: 0.8897 - loss: 0.3098
1604/1719
0s 786us/step - accuracy: 0.8897 - loss: 0.3097
1668/1719
0s 786us/step - accuracy: 0.8897 - loss: 0.3096
1719/1719
1s 844us/step - accuracy: 0.8897 - loss: 0.3096 - val_accuracy: 0.8678 - val_loss: 0.3591
Epoch 14/30
1/1719
19s 12ms/step - accuracy: 0.9375 - loss: 0.3176
66/1719
1s 773us/step - accuracy: 0.9079 - loss: 0.2706
131/1719
1s 775us/step
```

```
p - accuracy: 0.9004 - loss: 0.2845
195/1719
- 1s 777us/step - accuracy: 0.8970 - loss: 0.2928
262/1719
1s 771us/step - accuracy: 0.8956 - loss: 0.2960
327/1719
1s 773us/step - accuracy: 0.8945 - loss: 0.2981
390/1719
1s 778us/step - accuracy: 0.8939 - loss: 0.2998
456/1719
0s 776us/step - accuracy: 0.8935 - loss: 0.3010
522/1719
- 0s 775us/step - accuracy: 0.8931 - loss: 0.3018
589/1719
0s 773us/step - accuracy: 0.8928 - loss: 0.3023
656/1719
0s 771us/step - accuracy: 0.8928 - loss: 0.3024
722/1719
0s 770us/step - accuracy: 0.8927 - loss: 0.3026
789/1719
0s 768us/step - accuracy: 0.8925 - loss: 0.3028
855/1719
- 0s 768us/step - accuracy: 0.8924 - loss: 0.3030
922/1719
0s 766us/step - accuracy: 0.8923 - loss: 0.3033
987/1719
0s 767us/step - accuracy: 0.8923 - loss: 0.3033
1052/1719
0s 767us/step - accuracy: 0.8922 - loss: 0.3033
1119/1719
0s 767us/step - accuracy: 0.8922 - loss: 0.3032
1185/1719
- 0s 767us/step - accuracy: 0.8922 - loss: 0.3031
1252/1719
0s 766us/step - accuracy: 0.8922 - loss: 0.3031
1315/1719
0s 768us/step - accuracy: 0.8922 - loss: 0.3030
1380/1719
0s 768us/step - accuracy: 0.8922 - loss: 0.3029
1445/1719
0s 768us/step - accuracy: 0.8923 - loss: 0.3028
1511/1719
- 0s 767us/step - accuracy: 0.8923 - loss: 0.3027
1578/1719
0s 767us/step - accuracy: 0.8923 - loss: 0.3026
1645/1719
0s 766us/step - accuracy: 0.8923 - loss: 0.3025
```


Page 59 of 83

```

1346/1719 _____ 0s 787us/step - accuracy:
0.8940 - loss: 0.2962
1412/1719 _____ 0s 786us/ste
p - accuracy: 0.8940 - loss: 0.2961
1479/1719 _____
- 0s 785us/step - accuracy: 0.8941 - loss: 0.2960
1545/1719 _____
_____ 0s 784us/step - accuracy: 0.8941 - loss: 0.2959
1612/1719 _____ 0s 783us/step - accuracy: 0.8941 - loss:
0.2958
1679/1719 _____ 0s 781us/step - accuracy:
0.8941 - loss: 0.2957
1719/1719 _____ 1s 838us/ste
p - accuracy: 0.8941 - loss: 0.2957 - val_accuracy: 0.8684 - val_loss:
0.3557

```

Epoch 16/30

```

1/1719 _____ 18s 11ms/step - accuracy: 0.9375 - loss:
0.3022
65/1719 _____ 1s 788us/step - accuracy:
0.9128 - loss: 0.2575
131/1719 _____ 1s 773us/ste
p - accuracy: 0.9059 - loss: 0.2711
198/1719 _____
- 1s 766us/step - accuracy: 0.9019 - loss: 0.2794
265/1719 _____
_____ 1s 764us/step - accuracy: 0.9003 - loss: 0.2826
333/1719 _____ 1s 760us/step - accuracy: 0.8990 - loss:
0.2847
399/1719 _____ 1s 761us/step - accuracy:
0.8982 - loss: 0.2865
464/1719 _____ 0s 763us/ste
p - accuracy: 0.8977 - loss: 0.2876
529/1719 _____
- 0s 764us/step - accuracy: 0.8974 - loss: 0.2884
594/1719 _____
_____ 0s 766us/step - accuracy: 0.8971 - loss: 0.2888
659/1719 _____ 0s 767us/step - accuracy: 0.8970 - loss:
0.2890
721/1719 _____ 0s 770us/step - accuracy:
0.8969 - loss: 0.2892
782/1719 _____ 0s 775us/ste
p - accuracy: 0.8968 - loss: 0.2894
843/1719 _____
- 0s 779us/step - accuracy: 0.8967 - loss: 0.2896
905/1719 _____
_____ 0s 781us/step - accuracy: 0.8966 - loss: 0.2899
969/1719 _____ 0s 782us/step - accuracy: 0.8965 - loss:
0.2900

```

Epoch 17/30

Page 61 of 83

Epoch 18/30

Page 62 of 83

Page 63 of 83

Epoch 19/30

[illegible]

```

1389/1719 _____
- 0s 834us/step - accuracy: 0.9029 - loss: 0.2723
1448/1719 _____
0s 835us/step - accuracy: 0.9029 - loss: 0.2722
1506/1719 _____ 0s 836us/step - accuracy: 0.9029 - loss:
0.2722
1566/1719 _____ 0s 836us/step - accuracy:
0.9029 - loss: 0.2721
1626/1719 _____ 0s 836us/ste
p - accuracy: 0.9029 - loss: 0.2721
1687/1719 _____
- 0s 836us/step - accuracy: 0.9029 - loss: 0.2721
1719/1719 _____
2s 898us/step - accuracy: 0.9029 - loss: 0.2720 - val_a
ccuracy: 0.8730 - val_loss: 0.3520
Epoch 20/30
1/1719 _____ 20s 12ms/step - accuracy: 0.9375 - loss:
0.2773
62/1719 _____ 1s 822us/step - accuracy:
0.9210 - loss: 0.2348
125/1719 _____ 1s 812us/ste
p - accuracy: 0.9148 - loss: 0.2467
190/1719 _____
- 1s 802us/step - accuracy: 0.9110 - loss: 0.2554
251/1719 _____
1s 807us/step - accuracy: 0.9092 - loss: 0.2587
312/1719 _____ 1s 811us/step - accuracy: 0.9079 - loss:
0.2608
374/1719 _____ 1s 810us/step - accuracy:
0.9070 - loss: 0.2625
436/1719 _____ 1s 810us/ste
p - accuracy: 0.9064 - loss: 0.2640
498/1719 _____
- 0s 811us/step - accuracy: 0.9060 - loss: 0.2649
559/1719 _____
0s 813us/step - accuracy: 0.9057 - loss: 0.2656
621/1719 _____ 0s 812us/step - accuracy: 0.9055 - loss:
0.2659
682/1719 _____ 0s 814us/step - accuracy:
0.9054 - loss: 0.2661
742/1719 _____ 0s 815us/ste
p - accuracy: 0.9053 - loss: 0.2663
803/1719 _____
- 0s 816us/step - accuracy: 0.9051 - loss: 0.2665
864/1719 _____
0s 817us/step - accuracy: 0.9050 - loss: 0.2668
925/1719 _____ 0s 818us/step - accuracy: 0.9049 - loss:
0.2670
```

[illegible]

Epoch 22/30

Page 67 of 83

Page 68 of 83

1/1719 _____ 18s 11ms/step - accuracy: 0.9688 - loss: 0.2519
68/1719 _____ 1s 756us/step - accuracy: 0.9276 - loss: 0.2201
136/1719 _____ 1s 750us/step - accuracy: 0.9202 - loss: 0.2327
203/1719 _____
- 1s 752us/step - accuracy: 0.9163 - loss: 0.2403
270/1719 _____
_____ 1s 752us/step - accuracy: 0.9145 - loss: 0.2437
335/1719 _____ 1s 758us/step - accuracy: 0.9132 - loss: 0.2458
400/1719 _____ 1s 760us/step - accuracy: 0.9124 - loss: 0.2477
467/1719 _____ 0s 758us/step - accuracy: 0.9118 - loss: 0.2491
532/1719 _____
- 0s 761us/step - accuracy: 0.9115 - loss: 0.2500
598/1719 _____
_____ 0s 761us/step - accuracy: 0.9113 - loss: 0.2505
662/1719 _____ 0s 763us/step - accuracy: 0.9112 - loss: 0.2508
729/1719 _____ 0s 762us/step - accuracy: 0.9111 - loss: 0.2511
795/1719 _____ 0s 762us/step - accuracy: 0.9110 - loss: 0.2514
861/1719 _____
- 0s 762us/step - accuracy: 0.9109 - loss: 0.2517
927/1719 _____
_____ 0s 762us/step - accuracy: 0.9107 - loss: 0.2519
994/1719 _____ 0s 761us/step - accuracy: 0.9107 - loss: 0.2521
1061/1719 _____ 0s 761us/step - accuracy: 0.9106 - loss: 0.2521
1129/1719 _____ 0s 760us/step - accuracy: 0.9106 - loss: 0.2521
1196/1719 _____
- 0s 759us/step - accuracy: 0.9106 - loss: 0.2521
1261/1719 _____
_____ 0s 760us/step - accuracy: 0.9105 - loss: 0.2521
1327/1719 _____ 0s 760us/step - accuracy: 0.9105 - loss: 0.2521
1394/1719 _____ 0s 760us/step - accuracy: 0.9105 - loss: 0.2520
1462/1719 _____ 0s 759us/step - accuracy: 0.9105 - loss: 0.2519
1528/1719 _____
- 0s 759us/step - accuracy: 0.9105 - loss: 0.2519

```

1595/1719 _____
_____ 0s 759us/step - accuracy: 0.9105 - loss: 0.2519
1664/1719 _____ 0s 757us/step - accuracy: 0.9105 - loss:
0.2518
1719/1719 _____ 1s 813us/step - accuracy:
0.9105 - loss: 0.2518 - val_accuracy: 0.8750 - val_loss: 0.3488
Epoch 24/30
 1/1719 _____ 18s 11ms/step - accuracy: 0.9688 - loss:
0.2443
 66/1719 _____ 1s 777us/step - accuracy:
0.9280 - loss: 0.2153
131/1719 _____ 1s 777us/ste
p - accuracy: 0.9208 - loss: 0.2273
195/1719 _____
- 1s 779us/step - accuracy: 0.9174 - loss: 0.2351
260/1719 _____
_____ 1s 777us/step - accuracy: 0.9156 - loss: 0.2386
326/1719 _____ 1s 774us/step - accuracy: 0.9143 - loss:
0.2408
392/1719 _____ 1s 773us/step - accuracy:
0.9135 - loss: 0.2428
457/1719 _____ 0s 774us/ste
p - accuracy: 0.9129 - loss: 0.2443
523/1719 _____
- 0s 772us/step - accuracy: 0.9126 - loss: 0.2452
588/1719 _____
_____ 0s 773us/step - accuracy: 0.9124 - loss: 0.2458
653/1719 _____ 0s 773us/step - accuracy: 0.9123 - loss:
0.2461
720/1719 _____ 0s 771us/step - accuracy:
0.9122 - loss: 0.2463
786/1719 _____ 0s 771us/ste
p - accuracy: 0.9121 - loss: 0.2466
852/1719 _____
- 0s 770us/step - accuracy: 0.9120 - loss: 0.2469
919/1719 _____
_____ 0s 769us/step - accuracy: 0.9119 - loss: 0.2472
985/1719 _____ 0s 769us/step - accuracy: 0.9118 - loss:
0.2474
1050/1719 _____ 0s 769us/step - accuracy:
0.9118 - loss: 0.2474
1116/1719 _____ 0s 768us/ste
p - accuracy: 0.9118 - loss: 0.2474
1182/1719 _____
- 0s 768us/step - accuracy: 0.9118 - loss: 0.2474
1249/1719 _____
_____ 0s 767us/step - accuracy: 0.9117 - loss: 0.2474
1
```

313/1719 _____ 0s 768us/step - accuracy: 0.9117 - loss: 0.2474
1380/1719 _____ 0s 767us/step - accuracy: 0.9118 - loss: 0.2474
1447/1719 _____ 0s 766us/step - accuracy: 0.9118 - loss: 0.2473
1513/1719 _____ 0s 766us/step - accuracy: 0.9118 - loss: 0.2473
1578/1719 _____ 0s 766us/step - accuracy: 0.9118 - loss: 0.2472
1643/1719 _____ 0s 767us/step - accuracy: 0.9118 - loss: 0.2472
1709/1719 _____ 0s 766us/step - accuracy: 0.9118 - loss: 0.2472
1719/1719 _____ 1s 823us/step - accuracy: 0.9118 - loss: 0.2472 - val_accuracy: 0.8760 - val_loss: 0.3457
Epoch 25/30

1/1719 _____ 19s 11ms/step - accuracy: 0.9688 - loss: 0.2290
70/1719 _____ 1s 733us/step - accuracy: 0.9309 - loss: 0.2111
138/1719 _____ 1s 739us/step - accuracy: 0.9237 - loss: 0.2237
205/1719 _____ 1s 744us/step - accuracy: 0.9201 - loss: 0.2310
271/1719 _____ 1s 747us/step - accuracy: 0.9183 - loss: 0.2343
336/1719 _____ 1s 752us/step - accuracy: 0.9171 - loss: 0.2363
402/1719 _____ 0s 754us/step - accuracy: 0.9161 - loss: 0.2382
469/1719 _____ 0s 753us/step - accuracy: 0.9156 - loss: 0.2396
535/1719 _____ 0s 754us/step - accuracy: 0.9152 - loss: 0.2405
601/1719 _____ 0s 755us/step - accuracy: 0.9149 - loss: 0.2410
668/1719 _____ 0s 755us/step - accuracy: 0.9147 - loss: 0.2413
733/1719 _____ 0s 757us/step - accuracy: 0.9146 - loss: 0.2416
798/1719 _____ 0s 759us/step - accuracy: 0.9144 - loss: 0.2419
864/1719 _____ 0s 759us/step - accuracy: 0.9143 - loss: 0.2422
930/1719 _____ 0s 760us/step - accuracy: 0.9141 - loss: 0.2425

Page 72 of 83

Epoch 27/30

Page 73 of 83

Page 74 of 83


```
0.9367 - loss: 0.19780000000000000 129/1719 _____ 1s 785us/ste
p - accuracy: 0.9296 - loss: 0.20930000000000000 194/1719 _____
- 1s 784us/step - accuracy: 0.9258 - loss: 0.21730000000000000 259/1719 _____
_____ 1s 783us/step - accuracy: 0.9239 - loss: 0.22080000000000000
324/1719 _____ 1s 782us/step - accuracy: 0.9226 - loss:
0.22290000000000000 387/1719 _____ 1s 784us/step - accuracy:
0.9218 - loss: 0.22480000000000000 451/1719 _____ 0s 784us/ste
p - accuracy: 0.9211 - loss: 0.22630000000000000 514/1719 _____
- 0s 786us/step - accuracy: 0.9207 - loss: 0.22730000000000000 578/1719 _____
_____ 0s 786us/step - accuracy: 0.9204 - loss: 0.22790000000000000
643/1719 _____ 0s 785us/step - accuracy: 0.9202 - loss:
0.22820000000000000 707/1719 _____ 0s 785us/step - accuracy:
0.9201 - loss: 0.22850000000000000 771/1719 _____ 0s 785us/ste
p - accuracy: 0.9199 - loss: 0.22870000000000000 836/1719 _____
- 0s 785us/step - accuracy: 0.9197 - loss: 0.22900000000000000 900/1719 _____
_____ 0s 785us/step - accuracy: 0.9196 - loss: 0.22940000000000000
963/1719 _____ 0s 786us/step - accuracy: 0.9194 - loss:
0.22960000000000000 1029/1719 _____ 0s 784us/step - accuracy:
0.9193 - loss: 0.22970000000000000 1094/1719 _____ 0s 784us/ste
p - accuracy: 0.9193 - loss: 0.22970000000000000 1159/1719 _____
- 0s 783us/step - accuracy: 0.9192 - loss: 0.22970000000000000 1224/1719 _____
_____ 0s 783us/step - accuracy: 0.9191 - loss: 0.22970000000000000
1288/1719 _____ 0s 783us/step - accuracy: 0.9191 - loss:
0.22980000000000000 1352/1719 _____ 0s 783us/step - accuracy:
0.9191 - loss: 0.22980000000000000 1416/1719 _____ 0s 783us/ste
p - accuracy: 0.9191 - loss: 0.22970000000000000 1480/1719 _____
- 0s 784us/step - accuracy: 0.9190 - loss: 0.22970000000000000 1545/1719 _____
_____ 0s 783us/step - accuracy: 0.9190 - loss: 0.22970000000000000
1609/1719 _____ 0s 783us/step - accuracy: 0.9190 - loss: 0.22970000000000000
```

Page 76 of 83

297/1719 _____ 0s 776us/step - accuracy: 0.9209 - loss: 0.2258
1361/1719 _____ 0s 777us/step - accuracy: 0.9209 - loss: 0.2258
1428/1719 _____ 0s 776us/step - accuracy: 0.9209 - loss: 0.2257
1494/1719 _____
- 0s 776us/step - accuracy: 0.9209 - loss: 0.2257
1560/1719 _____
_____ 0s 775us/step - accuracy: 0.9208 - loss: 0.2257
1626/1719 _____ 0s 775us/step - accuracy: 0.9208 - loss: 0.2257
1692/1719 _____ 0s 775us/step - accuracy: 0.9208 - loss: 0.2257
1719/1719 _____ 1s 831us/step - accuracy: 0.9208 - loss: 0.2257 - val_accuracy: 0.8794 - val_loss: 0.3451

Epoch 30/30

1/1719 _____ 19s 11ms/step - accuracy: 0.9688 - loss: 0.1983
65/1719 _____ 1s 789us/step - accuracy: 0.9391 - loss: 0.1896
129/1719 _____ 1s 786us/step - accuracy: 0.9324 - loss: 0.2011
193/1719 _____
- 1s 785us/step - accuracy: 0.9290 - loss: 0.2088
259/1719 _____
_____ 1s 781us/step - accuracy: 0.9272 - loss: 0.2124
324/1719 _____ 1s 779us/step - accuracy: 0.9260 - loss: 0.2145
390/1719 _____ 1s 776us/step - accuracy: 0.9252 - loss: 0.2165
454/1719 _____ 0s 777us/step - accuracy: 0.9246 - loss: 0.2180
519/1719 _____
- 0s 778us/step - accuracy: 0.9242 - loss: 0.2190
584/1719 _____
_____ 0s 778us/step - accuracy: 0.9239 - loss: 0.2196
650/1719 _____ 0s 776us/step - accuracy: 0.9237 - loss: 0.2200
716/1719 _____ 0s 775us/step - accuracy: 0.9236 - loss: 0.2202
781/1719 _____ 0s 775us/step - accuracy: 0.9234 - loss: 0.2206
848/1719 _____
- 0s 773us/step - accuracy: 0.9232 - loss: 0.2209
914/1719 _____
_____ 0s 773us/step - accuracy: 0.9231 - loss: 0.2212

```
982/1719 _____ 0s 770us/step - accuracy: 0.9229 - loss:
0.2214_____
1049/1719 _____ 0s 770us/step - accuracy:
0.9228 - loss: 0.2215_____
1116/1719 _____ 0s 769us/st
e p - accuracy: 0.9227 - loss: 0.2215_____
1182/1719 _____
- 0s 769us/step - accuracy: 0.9226 - loss: 0.2215_____
1249/1719 _____
_____ 0s 767us/step - accuracy: 0.9226 - loss: 0.2216_____
1316/1719 _____ 0s 766us/step- accuracy: 0.9225 - loss:
0.2216_____
1381/1719 _____ 0s 767us/step - accuracy:
0.9225 - loss: 0.2216_____
1446/1719 _____ 0s 767us/st
e p - accuracy: 0.9224 - loss: 0.2216_____
1512/1719 _____
- 0s 767us/step - accuracy: 0.9224 - loss: 0.2215_____
1578/1719 _____
_____ 0s 767us/step - accuracy: 0.9224 - loss: 0.2215_____
1644/1719 _____ 0s 767us/step - accuracy: 0.9223 - loss:
0.2215_____
1711/1719 _____ 0s 766us/step - accuracy:
0.9223 - loss: 0.2215_____
1719/1719 _____ 1s 823us/st
e p - accuracy: 0.9223 - loss: 0.2215 - val_accuracy: 0.8802 - val_loss:
0.3446
```

The model is provided with both a training set and a validation set. At each step, the model will report its performance on both sets. This will also allow to visualize the accuracy and loss curves on both sets (more later).

When calling the `fit` method in Keras (or similar frameworks), each step corresponds to the evaluation of a mini-batch. A mini-batch is a subset of the training data, and during each step, the model updates its weights based on the error calculated from this mini-batch.

An epoch is defined as one complete pass through the entire training dataset. During an epoch, the model processes multiple mini-batches until it has seen all the training data once. This process is repeated for a specified number of epochs to optimize the model's performance.

Visualization

```
In [30]: import pandas as pd
```



```
1/1 _____ 0s 20ms/step
1/1 _____ 0s 28ms/step
array([[0. , 0. , 0. , 0. , 0. , 0.19, 0. , 0.01, 0. , 0.81],
       [0. , 0. , 1. , 0. , 0. , 0. , 0. , 0. , 0. , 0. ],
       [0. , 1. , 0. , 0. , 0. , 0. , 0. , 0. , 0. , 0. ]],
      dtype=float32)
```

...

```
In [33]: y_pred = y_proba.argmax(axis=-1)
         y_pred
```

```
array([9, 2, 1])
```

...

```
In [34]: y_new = y_test[:3]
         y_new
```

```
array([9, 2, 1], dtype=uint8)
```

As can be seen, the predictions are unambiguous, with only one class per prediction exhibiting a high value.

Predicted vs Observed

```
In [35]: plt.figure(figsize=(7.2, 2.4))
         for index, image in enumerate(X_new):
             plt.subplot(1, 3, index + 1)
             plt.imshow(image, cmap="binary", interpolation="nearest")
             plt.axis('off')
             plt.title(class_names[y_test[index]])
         plt.subplots_adjust(wspace=0.2, hspace=0.5)
         plt.show()
```

Ankle boot



Pullover



Trouser



```
In [36]: np.array(class_names)[y_pred]
```

```
array(['Ankle boot', 'Pullover', 'Trouser'], dtype='<U11')
```

Test Set Performance

```
In [37]: from sklearn.metrics import classification_report

y_proba = model.predict(X_test)
y_pred = y_proba.argmax(axis=-1)
```

Test Set Performance

```
In [38]: print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.86	0.80	0.83	1000
1	0.99	0.97	0.98	1000
2	0.76	0.84	0.80	1000
3	0.79	0.94	0.86	1000
4	0.82	0.78	0.80	1000
5	0.90	0.98	0.94	1000
6	0.76	0.63	0.69	1000
7	0.94	0.92	0.93	1000
8	0.96	0.96	0.96	1000
9	0.98	0.92	0.95	1000
accuracy			0.87	10000
macro avg	0.88	0.87	0.87	10000
weighted avg	0.88	0.87	0.87	10000

Prologue

Summary

- **Neural Networks Foundations:**

We introduced bio-inspired computation with neurodes and threshold logic units, outlining the perceptron model and its limitations (e.g., the XOR problem).

- **From Perceptrons to Deep Networks:**

We explained the evolution to multilayer perceptrons (MLPs) and feedforward architectures, emphasizing the critical role of nonlinear activation functions (sigmoid, tanh, ReLU) in enabling gradient-based learning and complex function approximation.

- **Universal Approximation:**

We discussed how even single hidden layer networks can approximate any

continuous function on a compact set, highlighting the theoretical underpinning of deep learning.

- **Practical Frameworks and Applications:**

Finally, we reviews leading deep learning frameworks (PyTorch, TensorFlow, Keras) and demonstrates practical model-building using the Fashion-MNIST dataset, covering model training, evaluation, and prediction.

3Blue1Brown on Deep Learning

- But what is a Neural Network?
 - youtu.be/aircAruvnKk
 - 19 minutes
- Gradient descent, how neural networks learn?
 - youtu.be/IHZwWFHWa-w
 - 21 minutes
- What is backpropagation really doing?
 - youtu.be/Ilg3gGewQ5U
 - 14 minutes
- Backpropagation calculus
 - youtu.be/Ilg3gGewQ5U
 - 10 minutes

Next lecture

- Training Deep Learning Models

References

Cybenko, George V. 1989. "Approximation by Superpositions of a Sigmoidal Function." *Mathematics of Control, Signals and Systems* 2: 303–14.

<https://api.semanticscholar.org/CorpusID:3958369>.

D'haeseleer, Patrik. 2006. "How Does DNA Sequence Motif Discovery Work?" *Nature Biotechnology* 24 (8): 959–61. <https://doi.org/10.1038/nbt0806-959>.

Géron, Aurélien. 2022. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 3rd ed. O'Reilly Media, Inc.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Adaptive Computation and Machine Learning. MIT Press.

<https://dblp.org/rec/books/daglib/0040158>.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White. 1989. "Multilayer Feedforward Networks Are Universal Approximators." *Neural Networks* 2 (5): 359–66. [https://doi.org/https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/https://doi.org/10.1016/0893-6080(89)90020-8).

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. <https://doi.org/10.1038/nature14539>.

LeNail, Alexander. 2019. "NN-SVG: Publication-Ready Neural Network Architecture Schematics." *Journal of Open Source Software* 4 (33): 747. <https://doi.org/10.21105/joss.00747>.

McCulloch, Warren S, and Walter Pitts. 1943. "A logical calculus of the ideas immanent in nervous activity." *The Bulletin of Mathematical Biophysics* 5 (4): 115–33. <https://doi.org/10.1007/bf02478259>.

Minsky, Marvin, and Seymour Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA, USA: MIT Press.

Rosenblatt, F. 1958. "The perceptron: A probabilistic model for information storage and organization in the brain." *Psychological Review* 65 (6): 386–408. <https://doi.org/10.1037/h0042519>.

Wasserman, WW, and A Sandelin. 2004. "Applied bioinformatics for the identification of regulatory elements." *Nature Reviews Genetics* 5 (4): 276–87. <https://doi.org/10.1038/nrg1315>.

Zou, James, Mikael Huss, Abubakar Abid, Pejman Mohammadi, Ali Torkamani, and Amalio Telenti. 2019. "A Primer on Deep Learning in Genomics." *Nature Genetics* 51 (1): 12–18. <https://doi.org/10.1038/s41588-018-0295-5>.

Marcel **Turcotte**

Marcel.Turcotte@uOttawa.ca

School of Electrical Engineering and **Computer Science (EECS)**

University of Ottawa