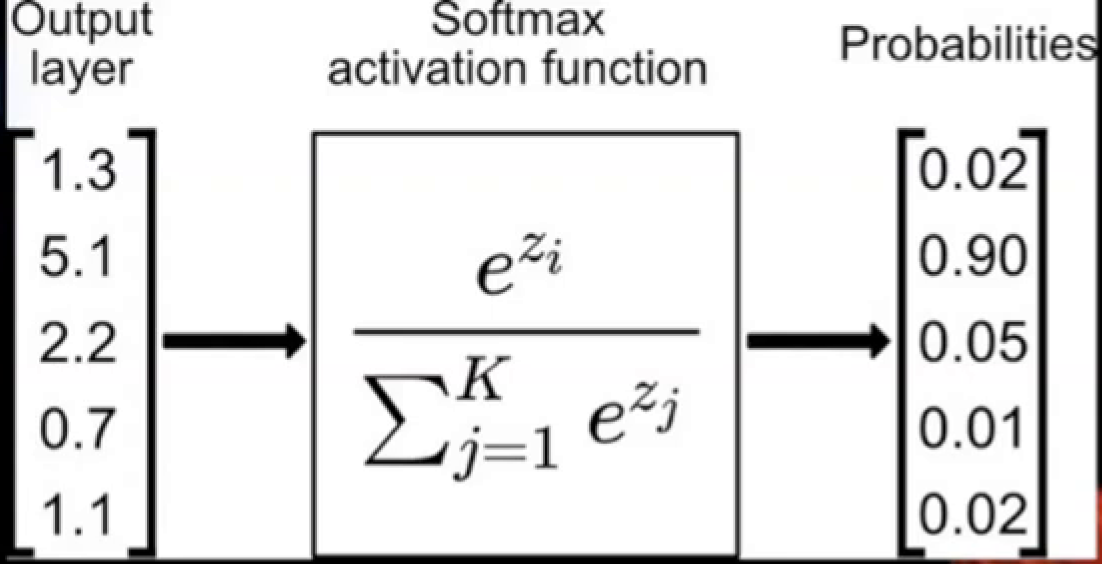
ACTIVATION FUNCTIONS



SOFTMAX ACTIVATION

Softmax Activation

ML For Nerds

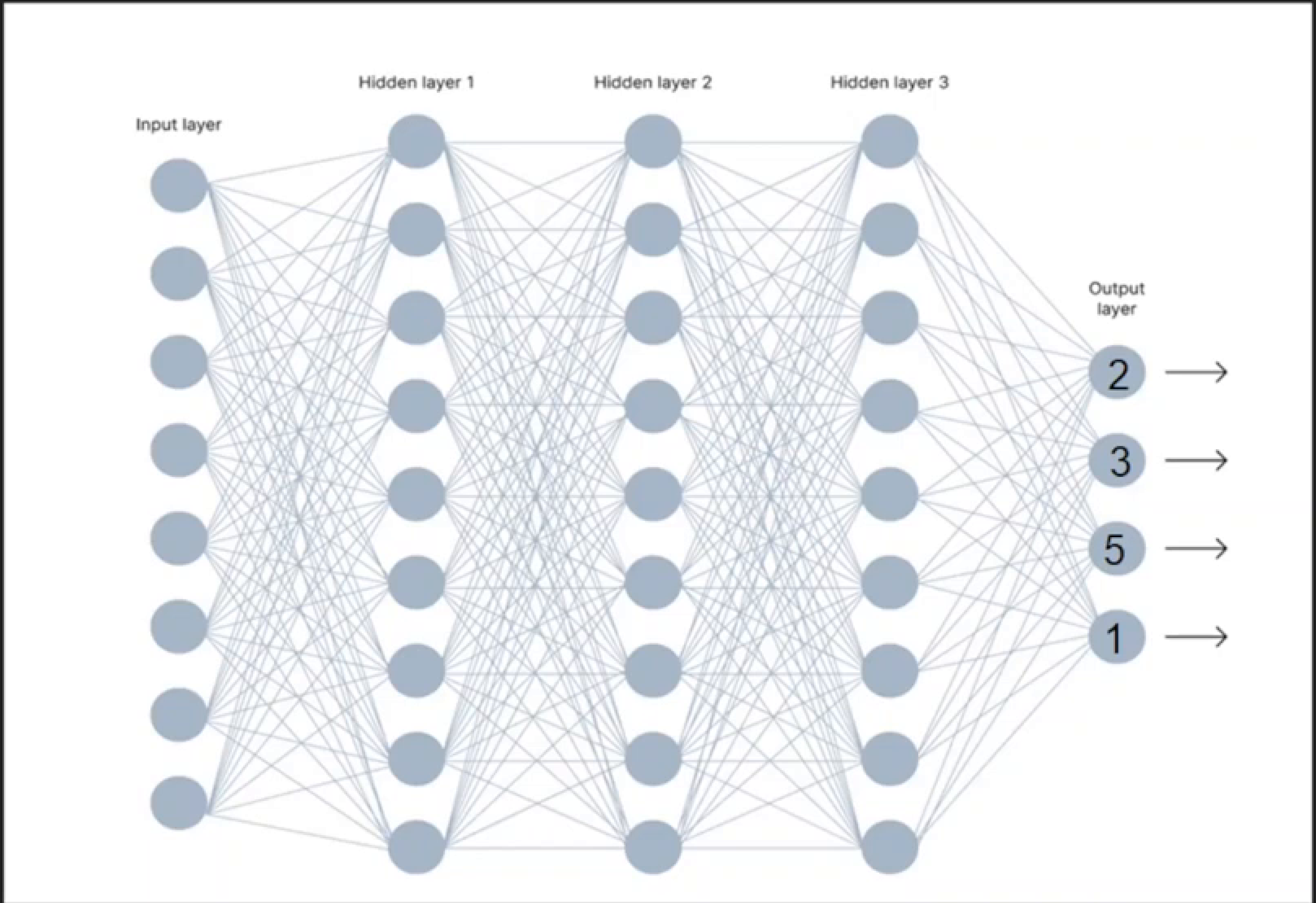
Softmax Activation

ML For Nerds

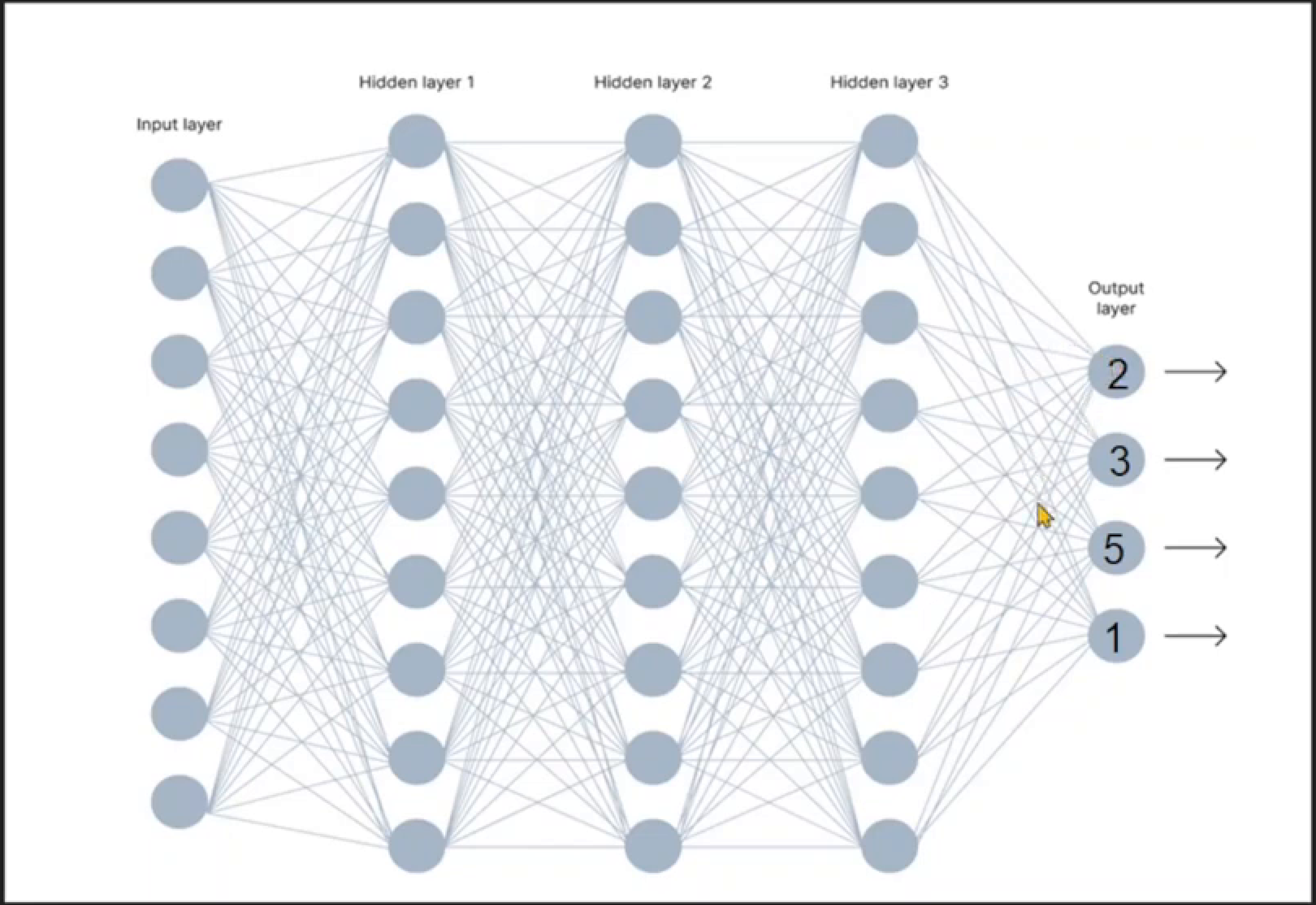
Softmax Activation

ML For Nerds

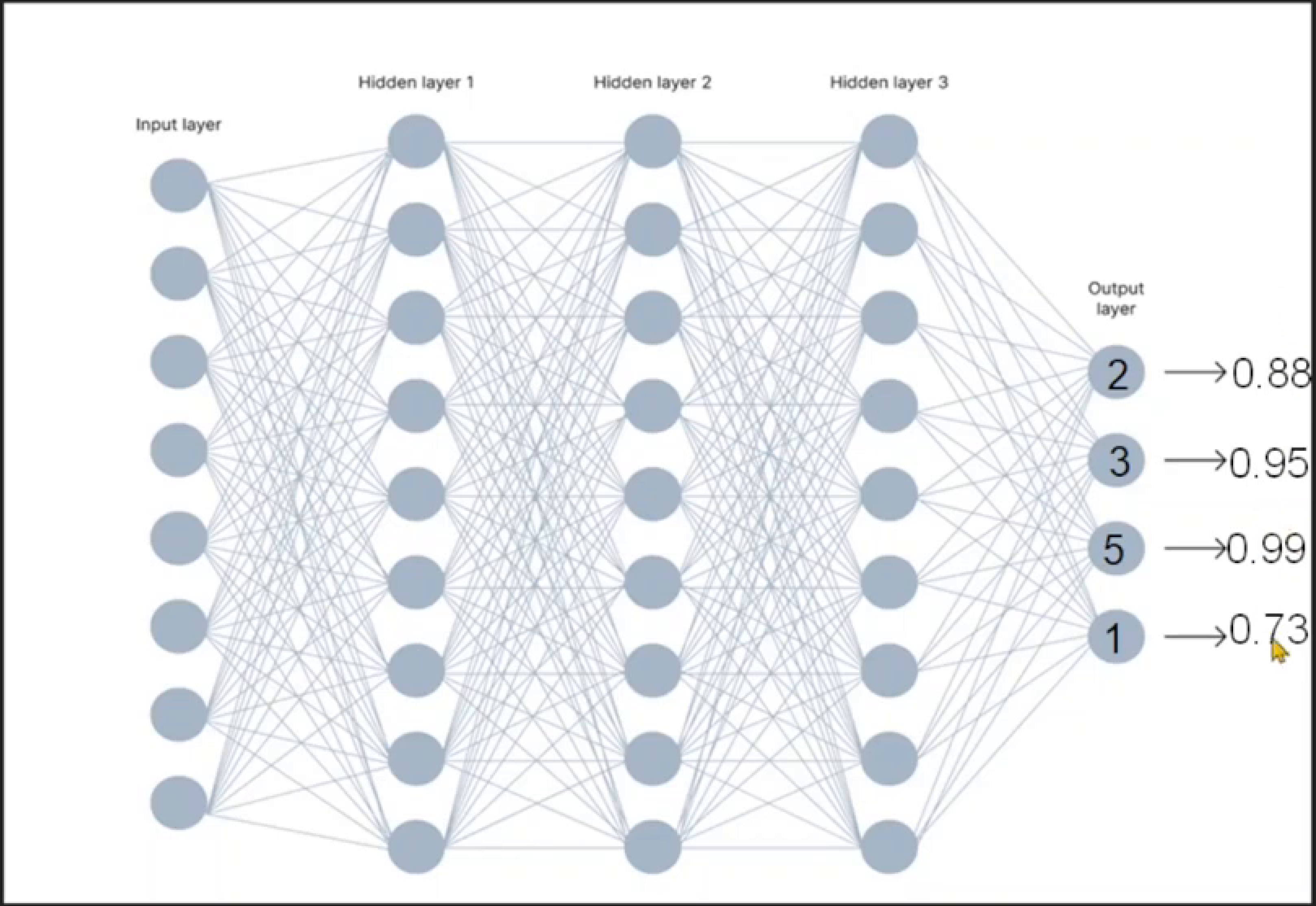
Multi-class classification



Multi-class classification



Multi-class classification

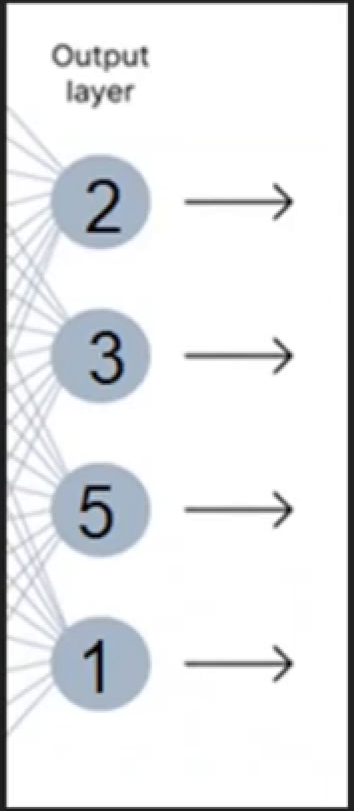


Requirements

- Range - [0,1]
- Sum of outputs = 1

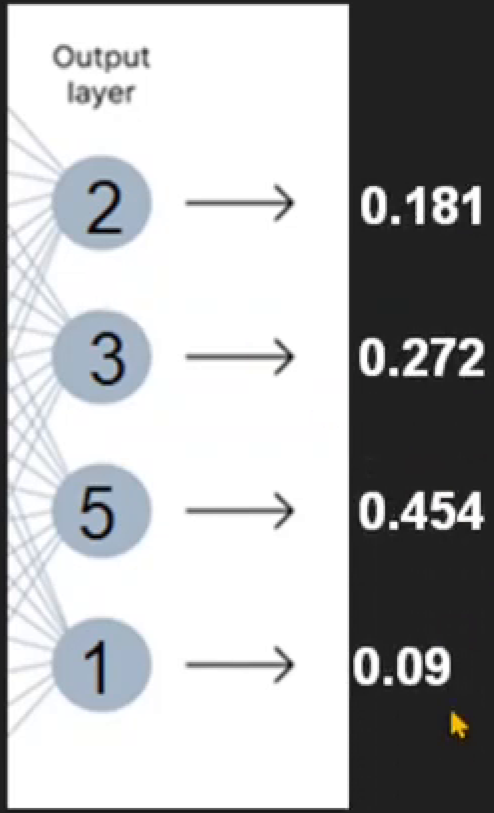
Requirements

- Range - [0,1]
- Sum of outputs = 1



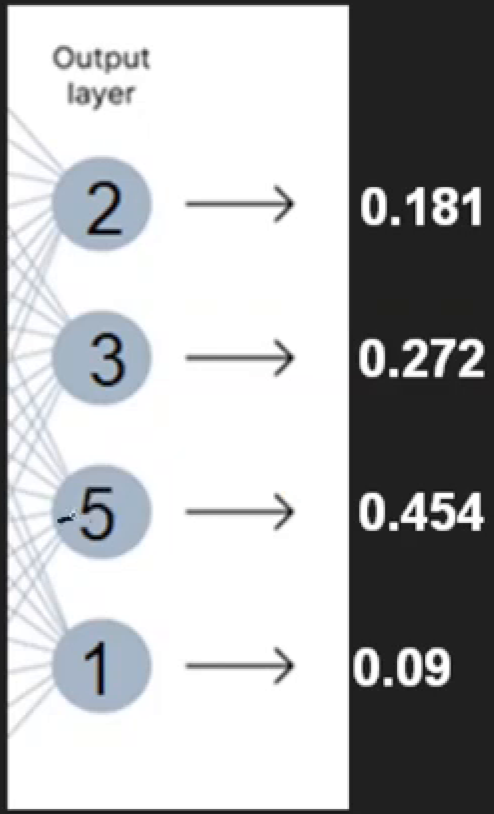
Requirements

- Range - [0,1]
- Sum of outputs = 1



Requirements

- Range - [0,1]
- Sum of outputs = 1

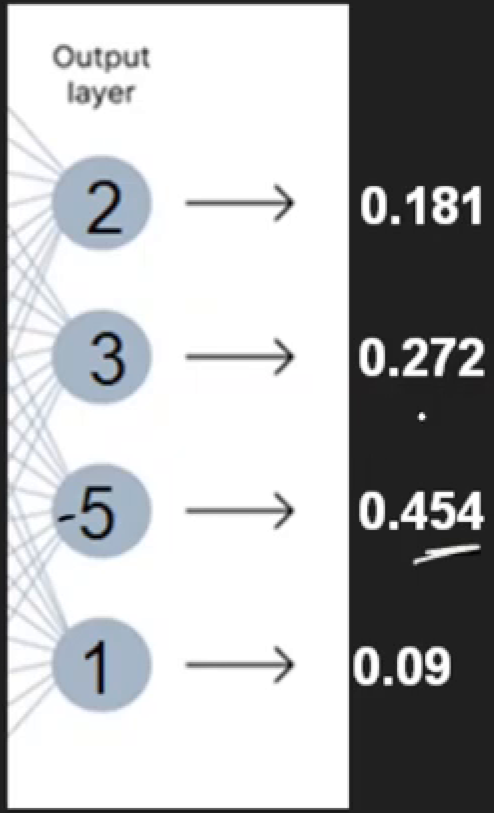


Requirements

- Range - [0,1]
- Sum of outputs = 1

Requirements

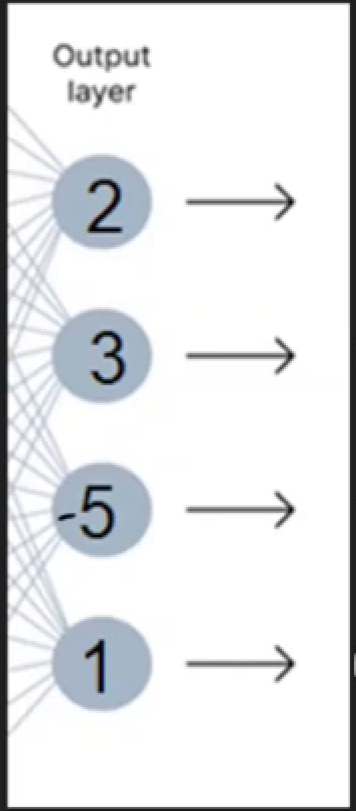
- Range - [0,1]
- Sum of outputs = 1



Requirements

- Range - [0,1]
- Sum of outputs = 1

$-\infty$ to ∞

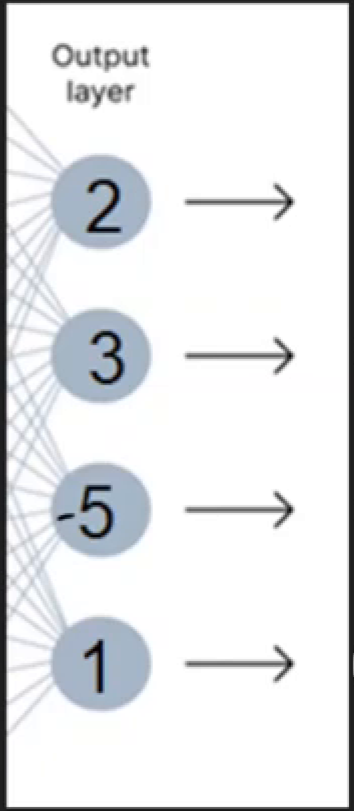


0.181
0.272
0.454
0.09

Requirements

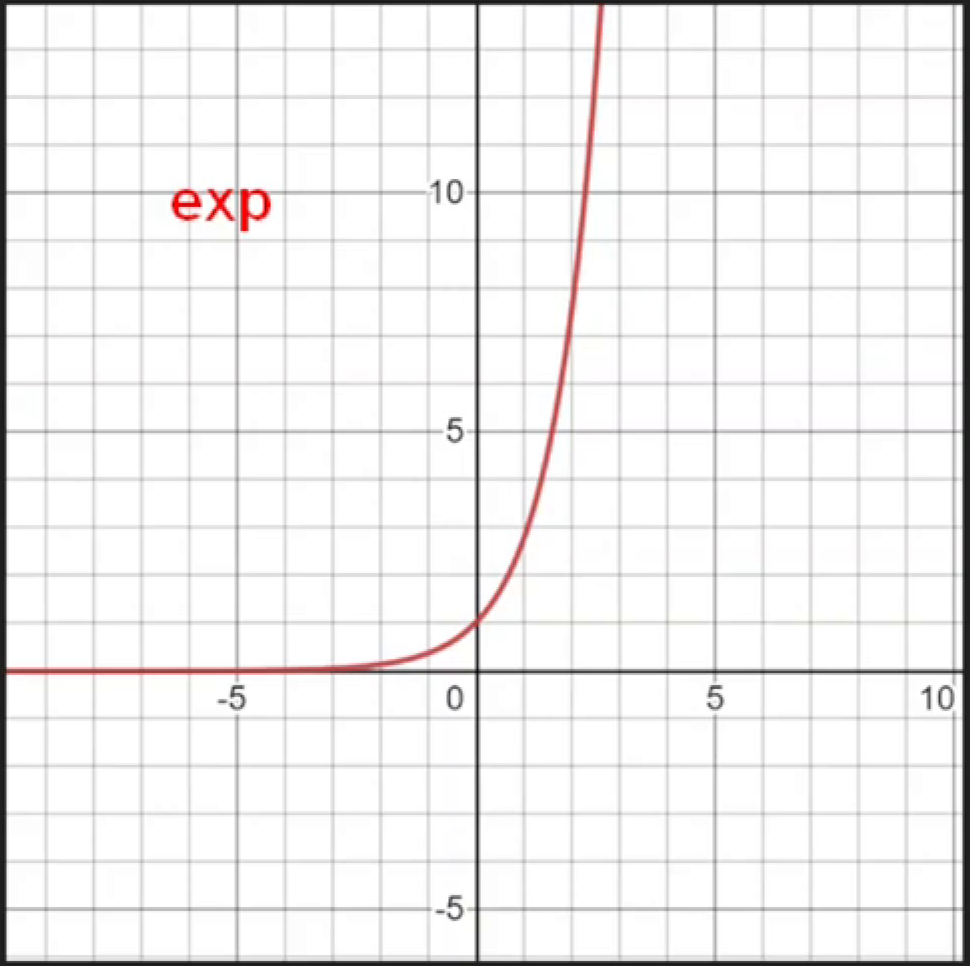
- Range - [0,1]
- Sum of outputs = 1

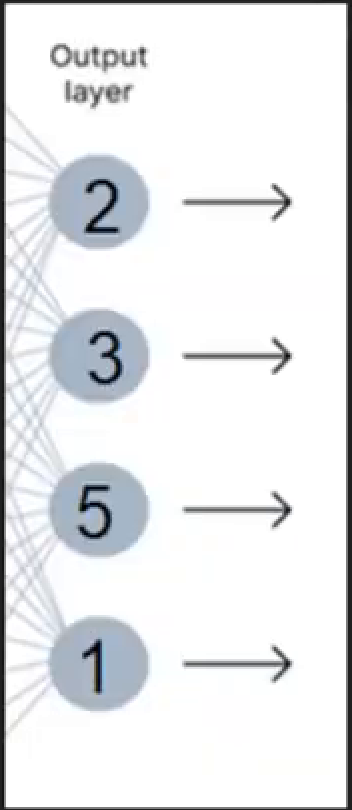
$-\infty$ to ∞
$\downarrow$
0 to ∞



0.181
0.272
0.454
0.09

X	-inf	0	inf
exp(x)	0	1	inf



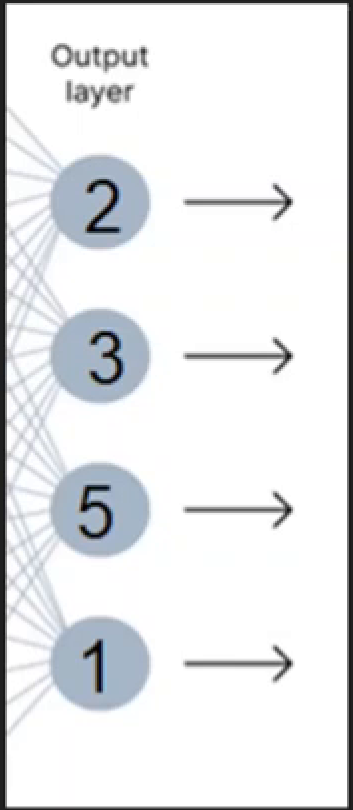


7.38

20.08

148.41

2.718



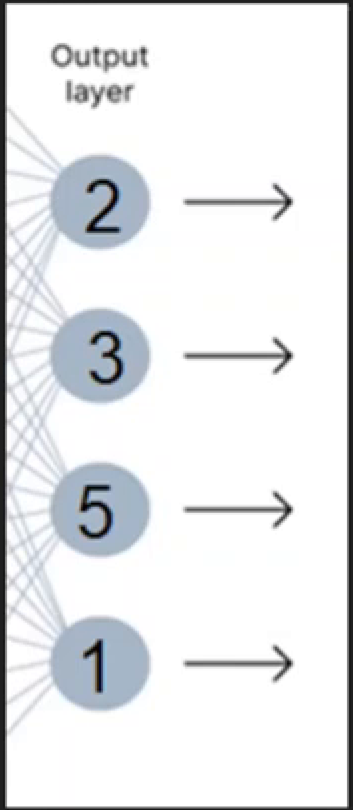
7.38

20.08

148.41

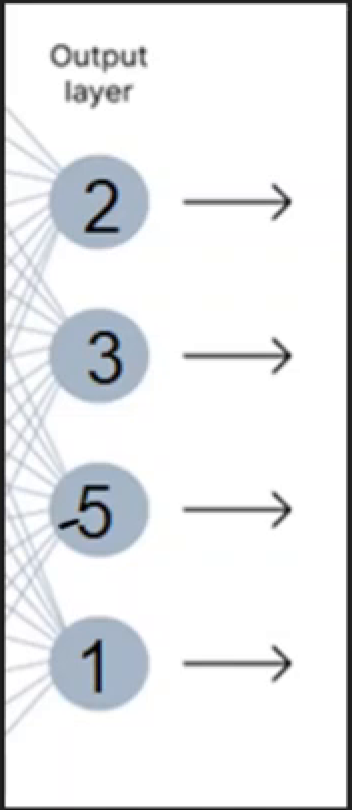
2.718

Sum 178.588



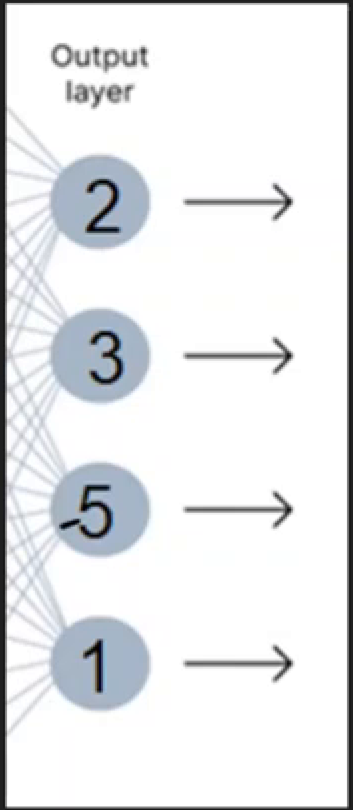
	Sum = 1
7.38	0.041
20.08	0.1124
148.41	0.831
2.718	0.0152
Sum = 178.588	

Sum = 1



7.38	0.041
20.08	0.1124
148.41	0.831
2.718	0.0152

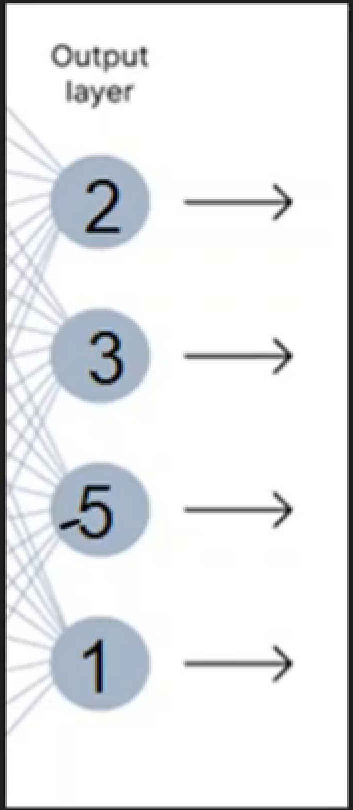
Sum = 178.588



Sum = 1

7.38	0.041
20.08	0.1124
148.41	0.831
2.718	0.0152

Sum = 178.588



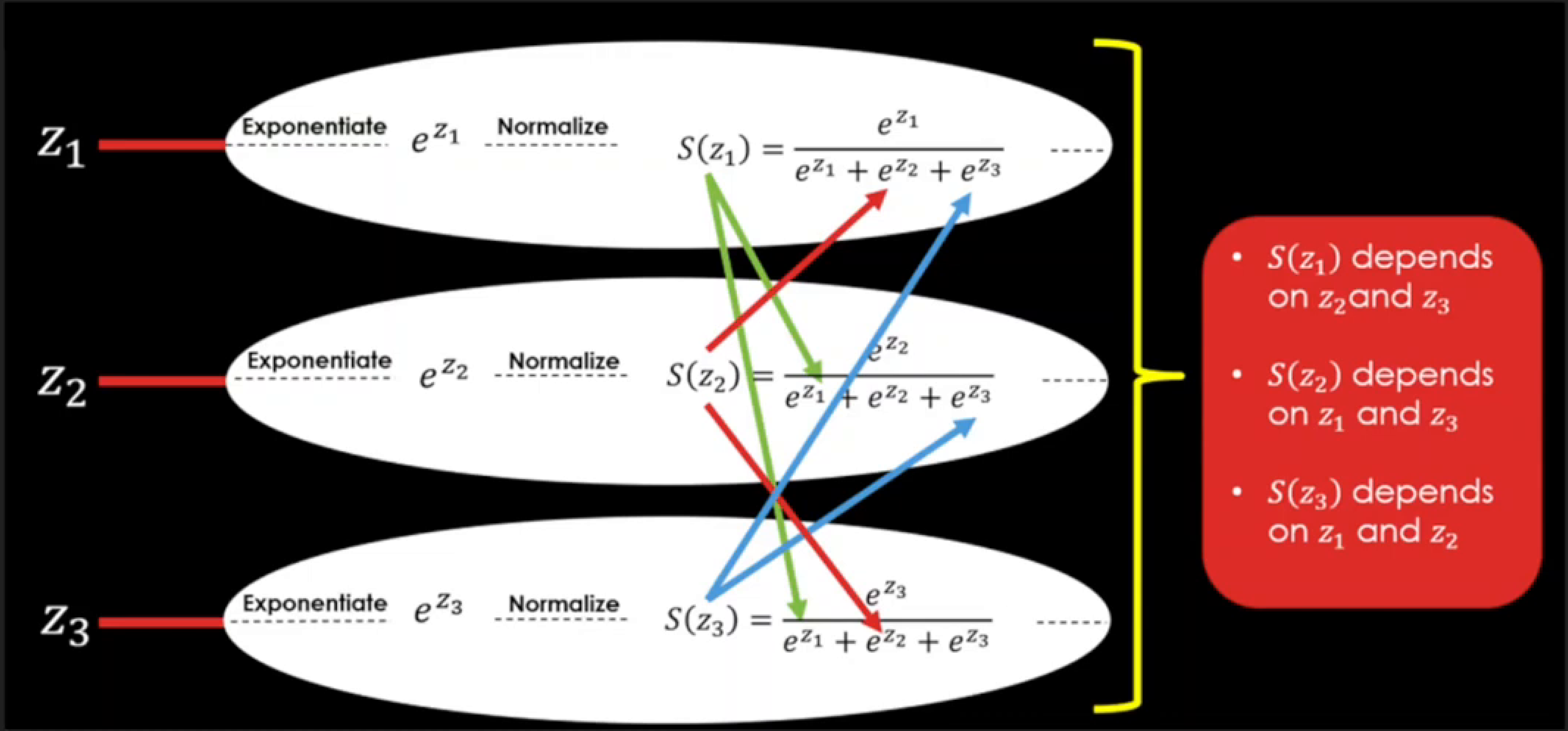
Sum = 1

7.38	0.041
20.08	0.1124
148.41	0.831
2.718	0.0152

Sum = 178.588

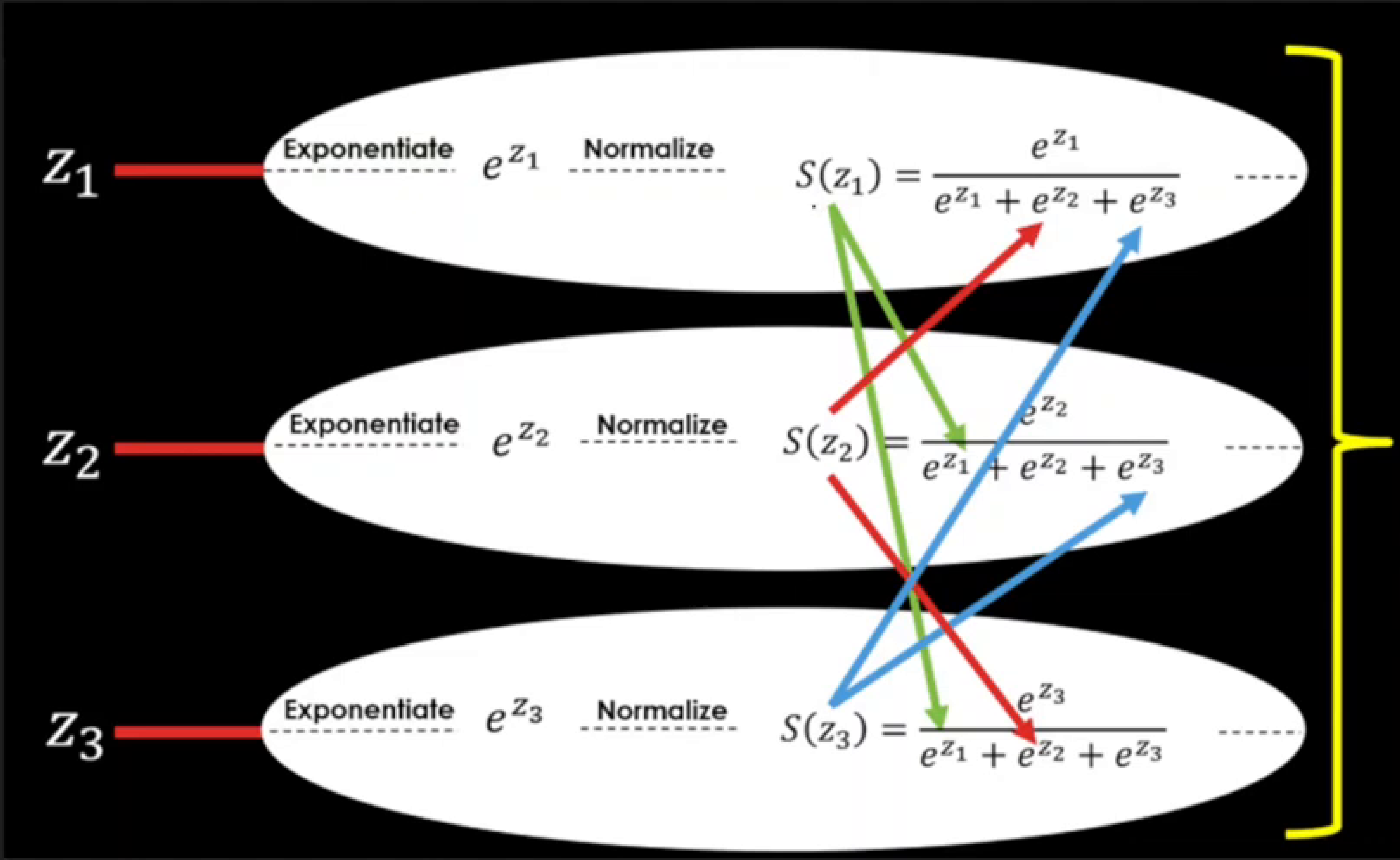
$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

K - number of classes



$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

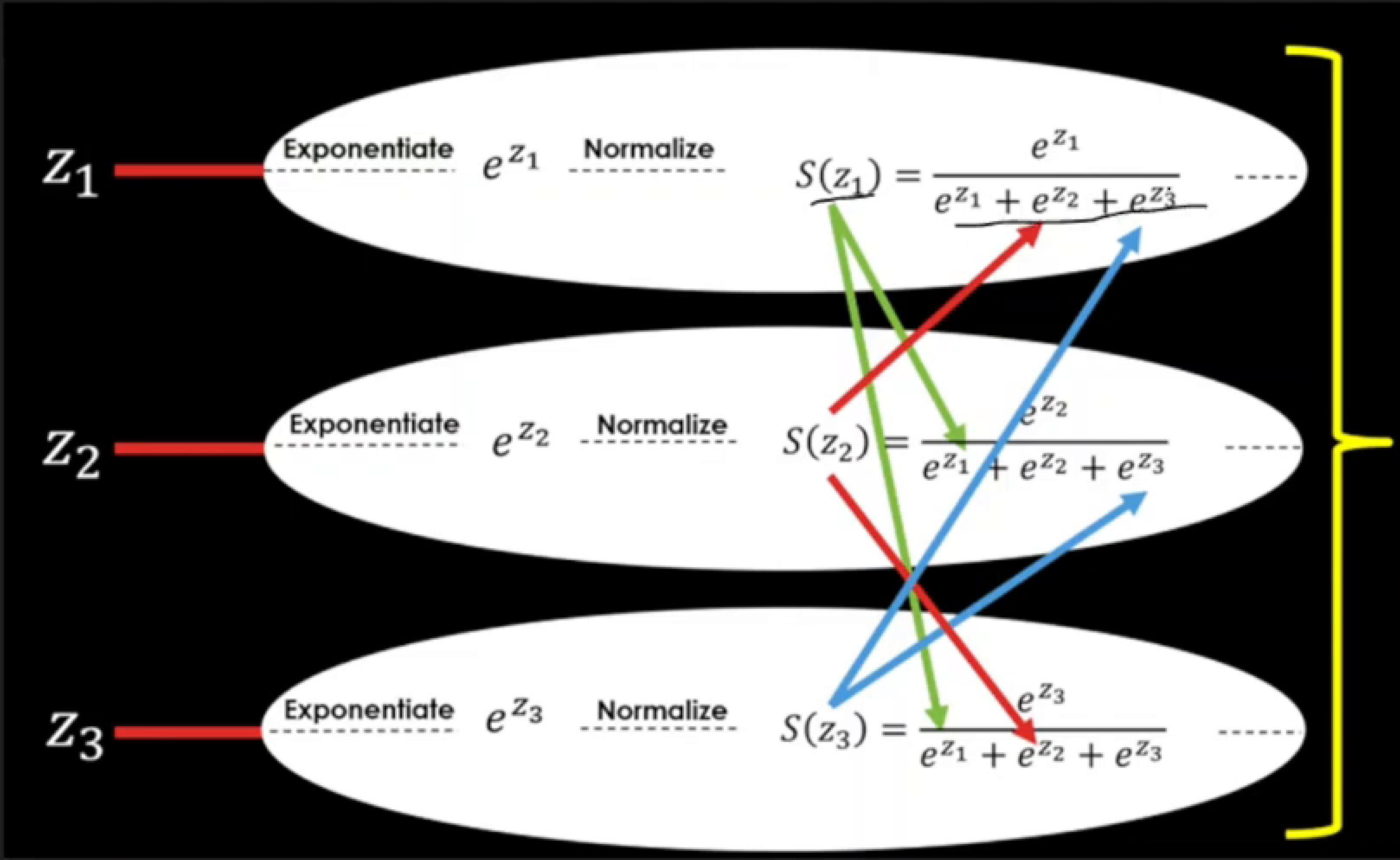
K - number of classes



- $S(z_1)$ depends on z_2 and z_3
- $S(z_2)$ depends on z_1 and z_3
- $S(z_3)$ depends on z_1 and z_2

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

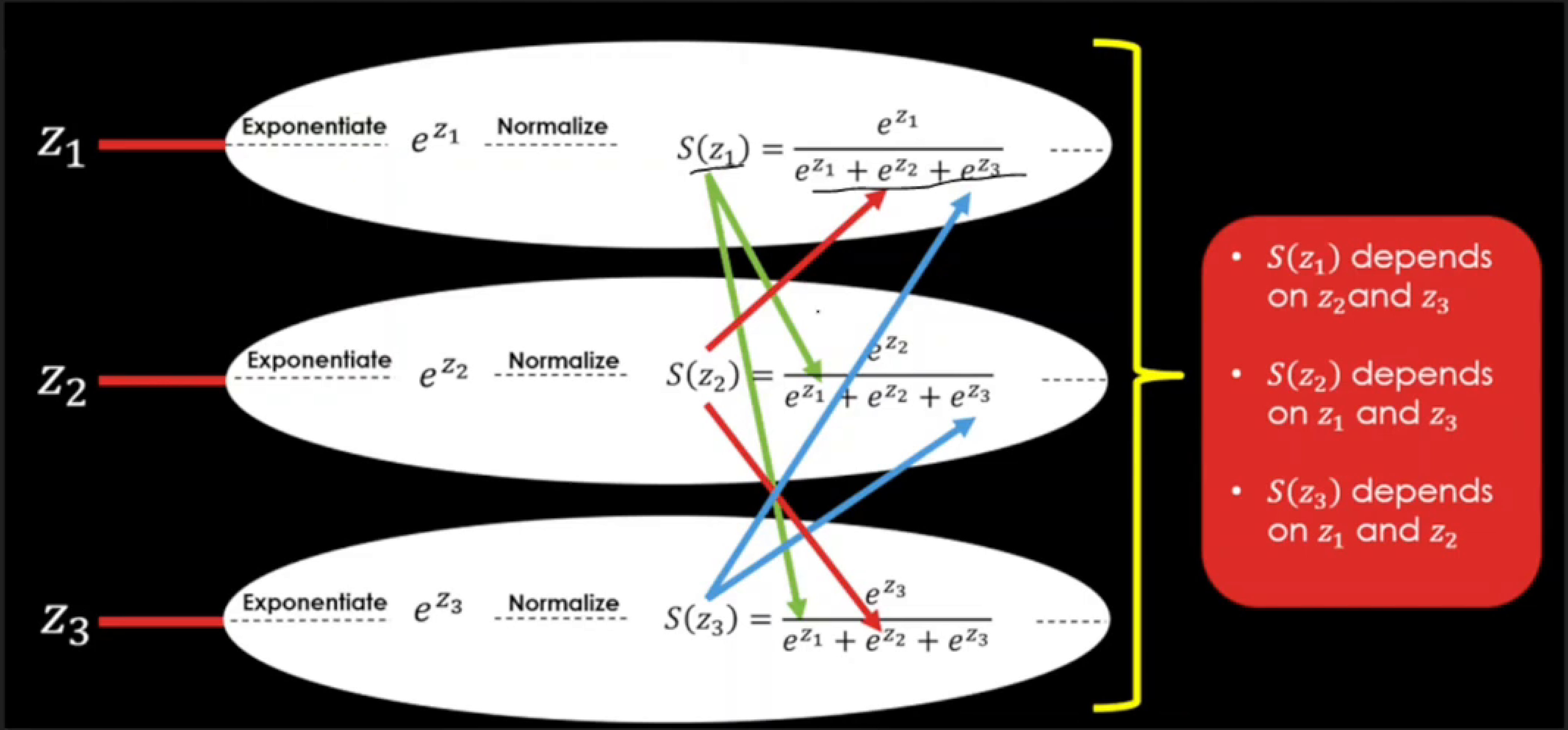
K - number of classes



- $S(z_1)$ depends on z_2 and z_3
- $S(z_2)$ depends on z_1 and z_3
- $S(z_3)$ depends on z_1 and z_2

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

K - number of classes



Derivative

S(z)

$$Z1 \longrightarrow S(z1) = \frac{e^{z1}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z2 \longrightarrow S(z2) = \frac{e^{z2}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z3 \longrightarrow S(z3) = \frac{e^{z3}}{e^{z1} + e^{z2} + e^{z3}}$$

Derivative

$$\frac{S(z)}{z}$$

$$Z1 \longrightarrow S(z1) = \frac{e^{z1}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z2 \longrightarrow S(z2) = \frac{e^{z2}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z3 \longrightarrow S(z3) = \frac{e^{z3}}{e^{z1} + e^{z2} + e^{z3}}$$

Derivative

$$\frac{S(z)}{z}$$

$$Z1 \longrightarrow \underline{S(z1)} = \frac{e^{z1}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z2 \longrightarrow S(z2) = \frac{e^{z2}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z3 \longrightarrow S(z3) = \frac{e^{z3}}{e^{z1} + e^{z2} + e^{z3}}$$

$$\frac{\partial S(z1)}{\partial z1} \quad \frac{\partial S(z1)}{\partial z2} \quad \frac{\partial S(z1)}{\partial z3}$$

$$\frac{\partial S(z2)}{\partial z1} \quad \frac{\partial S(z2)}{\partial z2} \quad \frac{\partial S(z2)}{\partial z3}$$

$$\frac{\partial S(z3)}{\partial z1} \quad \frac{\partial S(z3)}{\partial z2} \quad \frac{\partial S(z3)}{\partial z3}$$

Derivative

$$\frac{S(z)}{z}$$

$$Z1 \longrightarrow S(z1) = \frac{e^{z1}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z2 \longrightarrow S(z2) = \frac{e^{z2}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z3 \longrightarrow S(z3) = \frac{e^{z3}}{e^{z1} + e^{z2} + e^{z3}}$$

$$\frac{\partial S(z1)}{\partial z1} \quad \frac{\partial S(z1)}{\partial z2} \quad \frac{\partial S(z1)}{\partial z3}$$

$$\frac{\partial S(z2)}{\partial z1} \quad \frac{\partial S(z2)}{\partial z2} \quad \frac{\partial S(z2)}{\partial z3}$$

$$\frac{\partial S(z3)}{\partial z1} \quad \frac{\partial S(z3)}{\partial z2} \quad \frac{\partial S(z3)}{\partial z3}$$

Derivative

$$\frac{\partial S(z1)}{\partial z1} = S(z1) \cdot (1 - S(z1))$$

$$\frac{\partial S(z2)}{\partial z1} = -S(z1) \cdot S(z2)$$

$$\frac{\partial S(z3)}{\partial z1} = -S(z1) \cdot S(z3)$$

Derivative

$$\frac{\partial S(z1)}{\partial z1} = S(z1) \cdot (1 - S(z1))$$

$$\frac{\partial S(z2)}{\partial z1} = -S(z1) \cdot S(z2)$$

$$\frac{\partial S(z3)}{\partial z1} = -S(z1) \cdot S(z3)$$

Derivative

$$\frac{\partial S(z1)}{\partial z1} = S(z1) \cdot (1 - S(z1))$$

$$\frac{\partial S(z2)}{\partial z1} = -S(z1) \cdot S(z2)$$

$$\frac{\partial S(z3)}{\partial z1} = -S(z1) \cdot S(z3)$$

Derivative

$$\frac{\partial S(z1)}{\partial z1} = S(z1) \cdot (1 - S(z1))$$

$$\frac{\partial S(z2)}{\partial z1} = -S(z1) \cdot S(z2)$$

$$\frac{\partial S(z3)}{\partial z1} = -S(z1) \cdot S(z3)$$

$$\frac{\partial S(z_i)}{\partial z_j} = \begin{cases} S(z_i) \times (1 - S(z_i)) & if\ i = j \\ -S(z_i) \times S(z_j) & if\ i \neq j \end{cases}$$

Derivative

$$\frac{\partial S(z1)}{\partial z1} = S(z1) \cdot (1 - S(z1))$$

$$\frac{\partial S(z2)}{\partial z1} = -S(z1) \cdot S(z2)$$

$$\frac{\partial S(z3)}{\partial z1} = -S(z1) \cdot S(z3)$$

$$\frac{\partial S(z_i)}{\partial z_j} = \begin{cases} S(z_i) \times (1 - S(z_i)) & if\ i = j \\ -S(z_i) \times S(z_j) & if\ i \neq j \end{cases}$$

Derivative

$$\frac{S(z)}{z}$$

$$Z1 \longrightarrow S(z1) = \frac{e^{z1}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z2 \longrightarrow S(z2) = \frac{e^{z2}}{e^{z1} + e^{z2} + e^{z3}}$$

$$Z3 \longrightarrow S(z3) = \frac{e^{z3}}{e^{z1} + e^{z2} + e^{z3}}$$

$$\begin{array}{ccc}
\frac{\partial S(z1)}{\partial z1} & \frac{\partial S(z1)}{\partial z2} & \frac{\partial S(z1)}{\partial z3} \\[2ex]
\frac{\partial S(z2)}{\partial z1} & \frac{\partial S(z2)}{\partial z2} & \frac{\partial S(z2)}{\partial z3} \\[2ex]
\frac{\partial S(z3)}{\partial z1} & \frac{\partial S(z3)}{\partial z2} & \frac{\partial S(z3)}{\partial z3}
\end{array}$$

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

Drawbacks

- Computationally expensive
- Used for only output layers
- Not 0-centered function – not a drawback

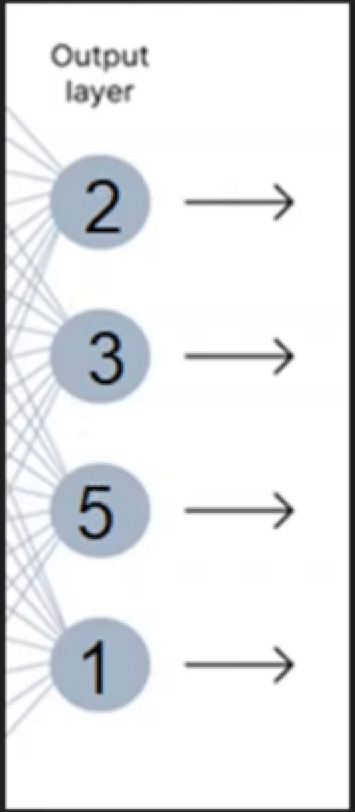
Python Implementation

```
def softmax_function(x):
    z = np.exp(x)
    z_ = z/z.sum()
    return z_
```

Python Implementation

```
def softmax_function(x):
    z = np.exp(x)
    z_ = z/z.sum()
    return z_
```

X = [2, 3, 5, 1]

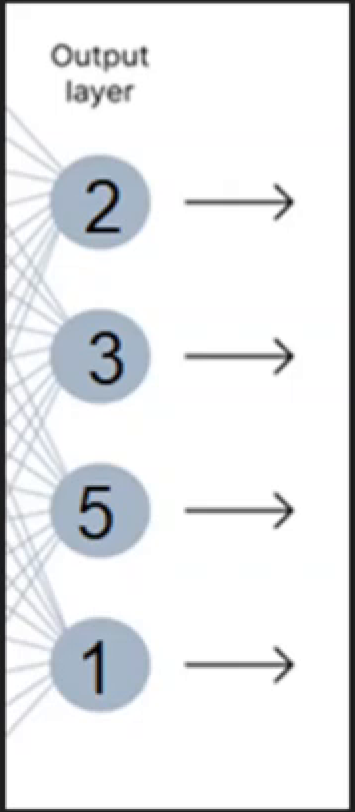


Python Implementation

```
def softmax_function(x):
    z = np.exp(x)
    z_ = z/z.sum()
    return z_
```

X = [2, 3, 5, 1]

Z= [7.38, 20.08, 148.41, 2.718]



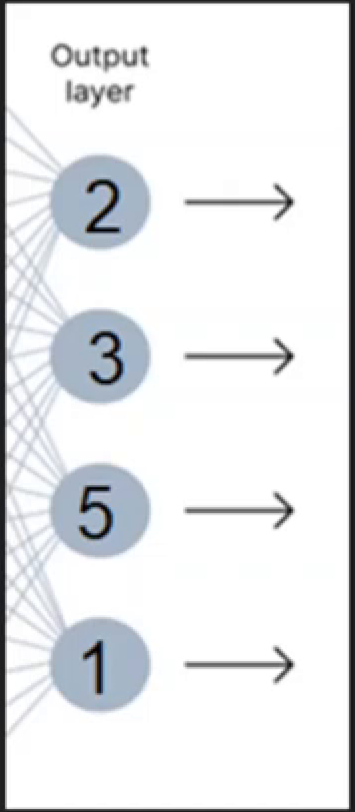
Python Implementation

```
def softmax_function(x):
    z = np.exp(x)
    z_ = z/z.sum()
    return z_
```

X = [2, 3, 5, 1]

Z= [7.38, 20.08, 148.41, 2.718]

Z_= [0.041, 0.1124, 0.831, 0.0152]



Python Implementation

```
def softmax_function(x):
    z = np.exp(x)
    z_ = z/z.sum()
    return z_
```

X = [2, 3, 5, 1]

Z= [7.38, 20.08, 148.41, 2.718]

Z_= [0.041, 0.1124, 0.831, 0.0152]