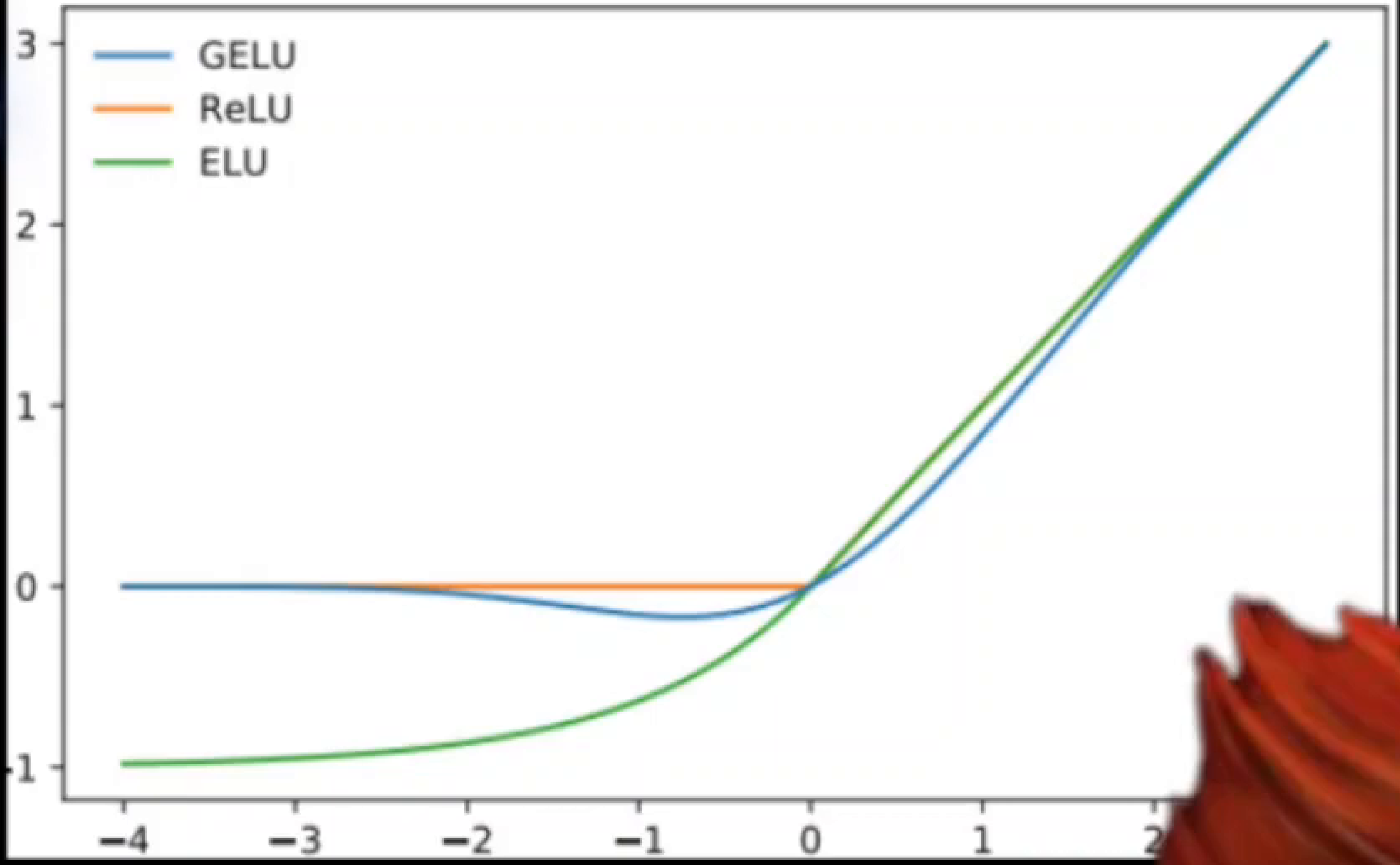
ACTIVATION FUNCTIONS




# GELU ACTIVATION

# Motivation

- Combine ReLU & Dropout
- ReLU - multiplies the input by 1 or 0
- Dropout - Randomly drops neurons

# Motivation

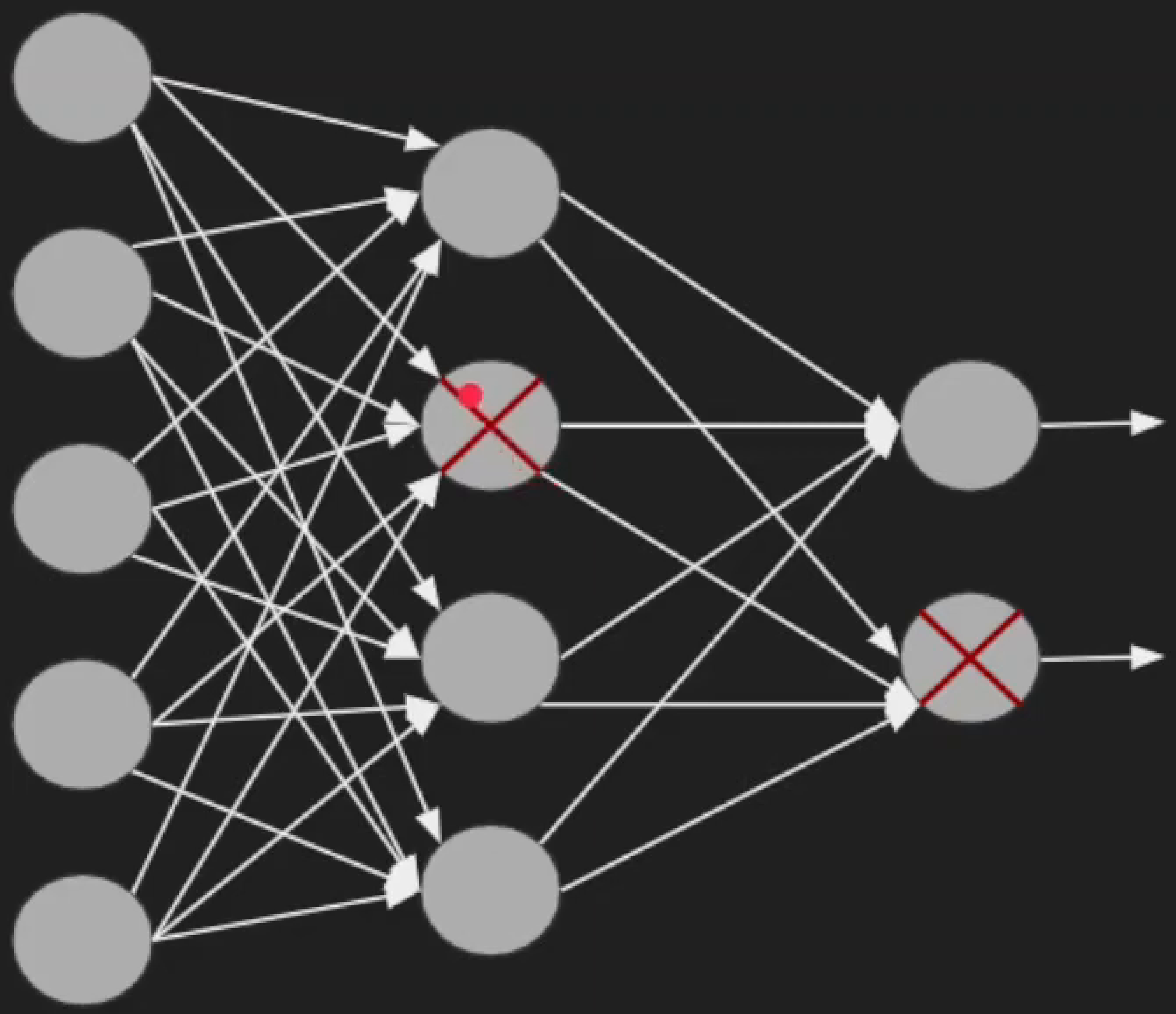
- Combine ReLU & Dropout
- ReLU - multiplies the input by 1 or 0
- Dropout - Randomly drops neurons

# Motivation

- Combine ReLU & Dropout
- ReLU - multiplies the input by 1 or 0
- Dropout - Randomly drops neurons

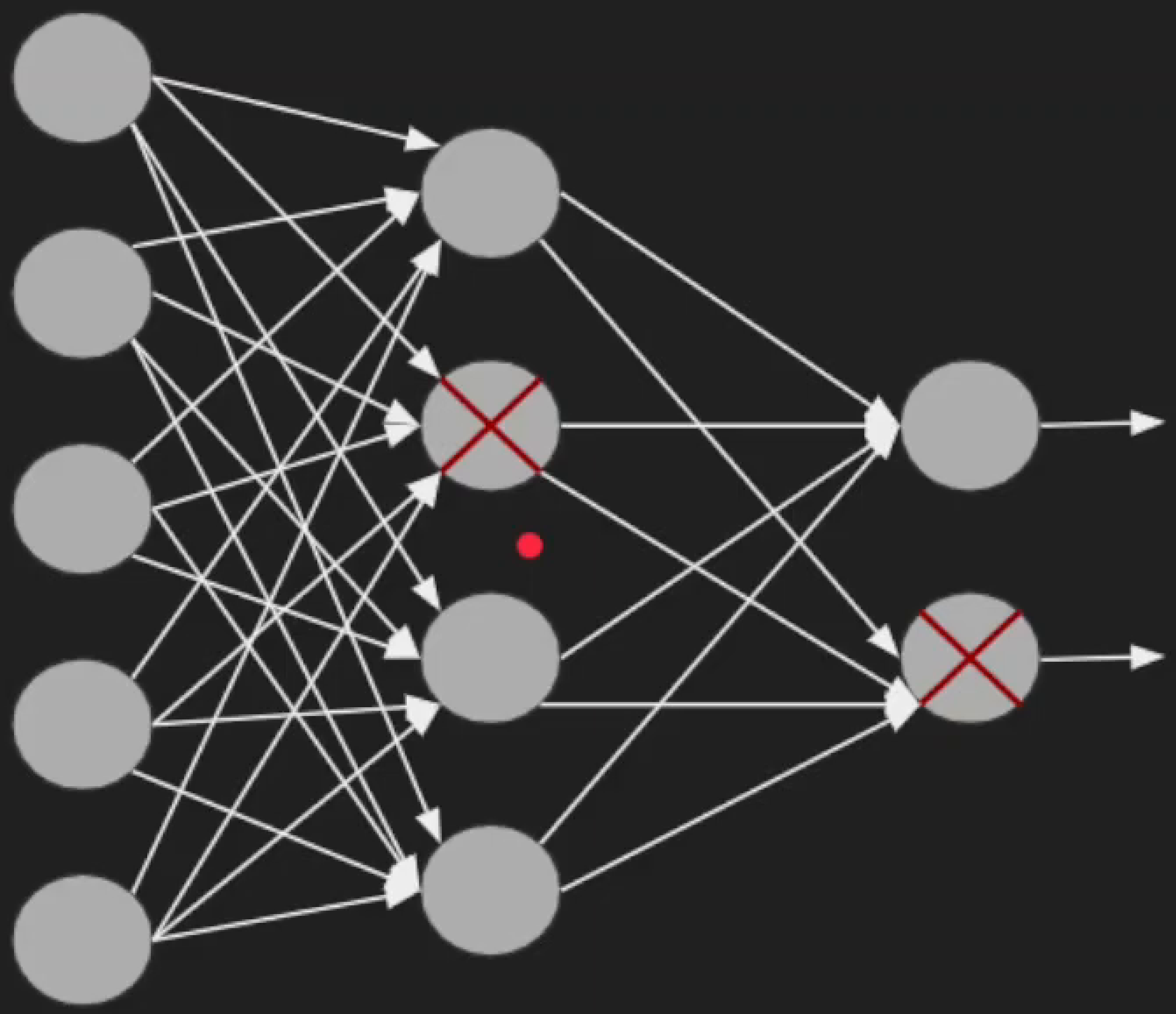
# Motivation

- Combine ReLU & Dropout
- ReLU - multiplies the input by 1 or 0
- Dropout - Randomly drops neurons



# Motivation

- Combine ReLU & Dropout
- ReLU - multiplies the input by 1 or 0
- Dropout - Randomly drops neurons



# Gaussian Error Linear Unit

Weighted sum - x

$$g(x) = GeLU(x) = x * m$$

# Gaussian Error Linear Unit

Weighted sum - x

g(x) = GeLU(x) = x * m

m = Cumulative Distribution Function(Normal Distribution)

# Intuition

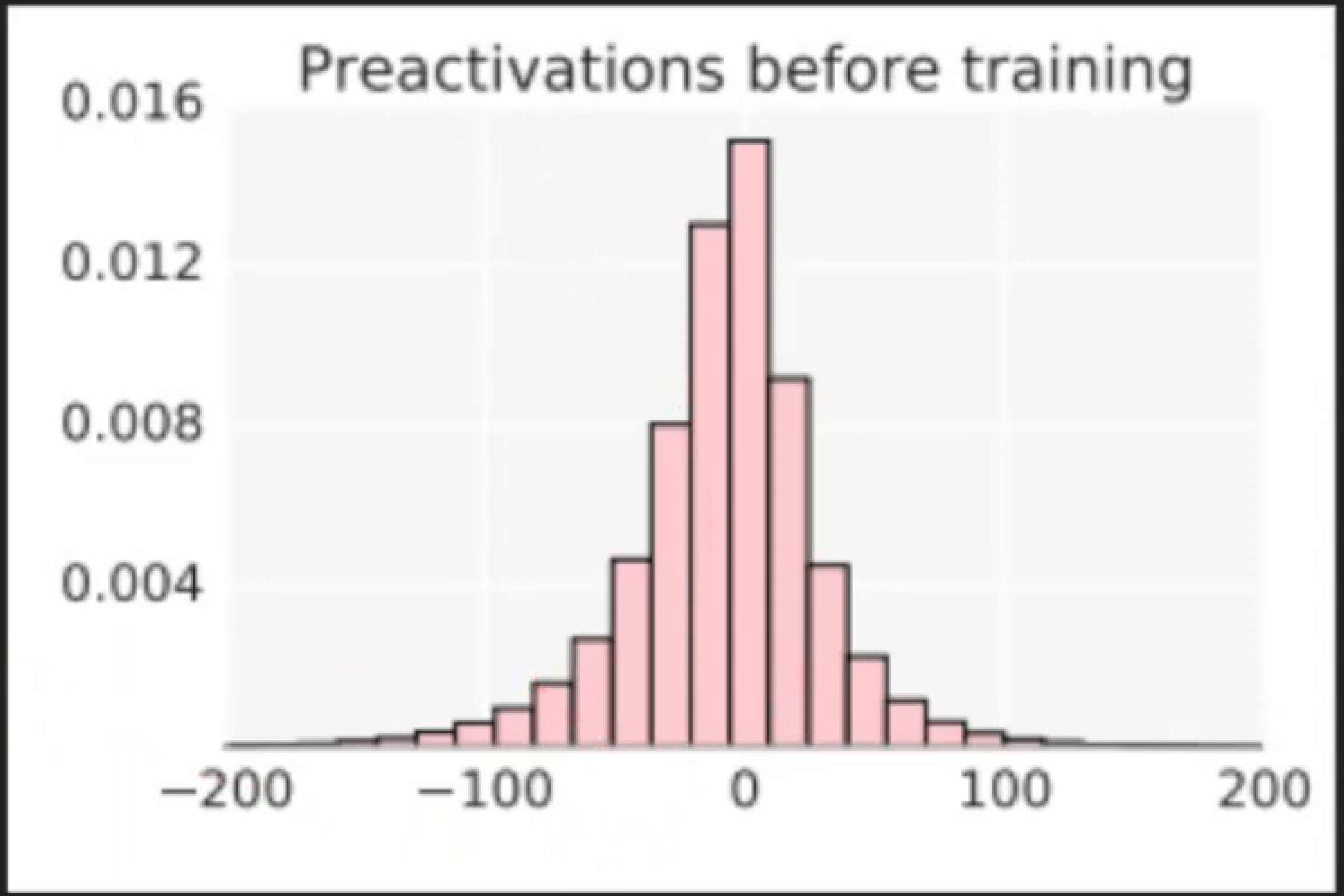
Input: $X \in R^{m,n}$

# Intuition

Input: $X \in R^{m,n}$

# Intuition

Input: $X \in R^{m,n}$

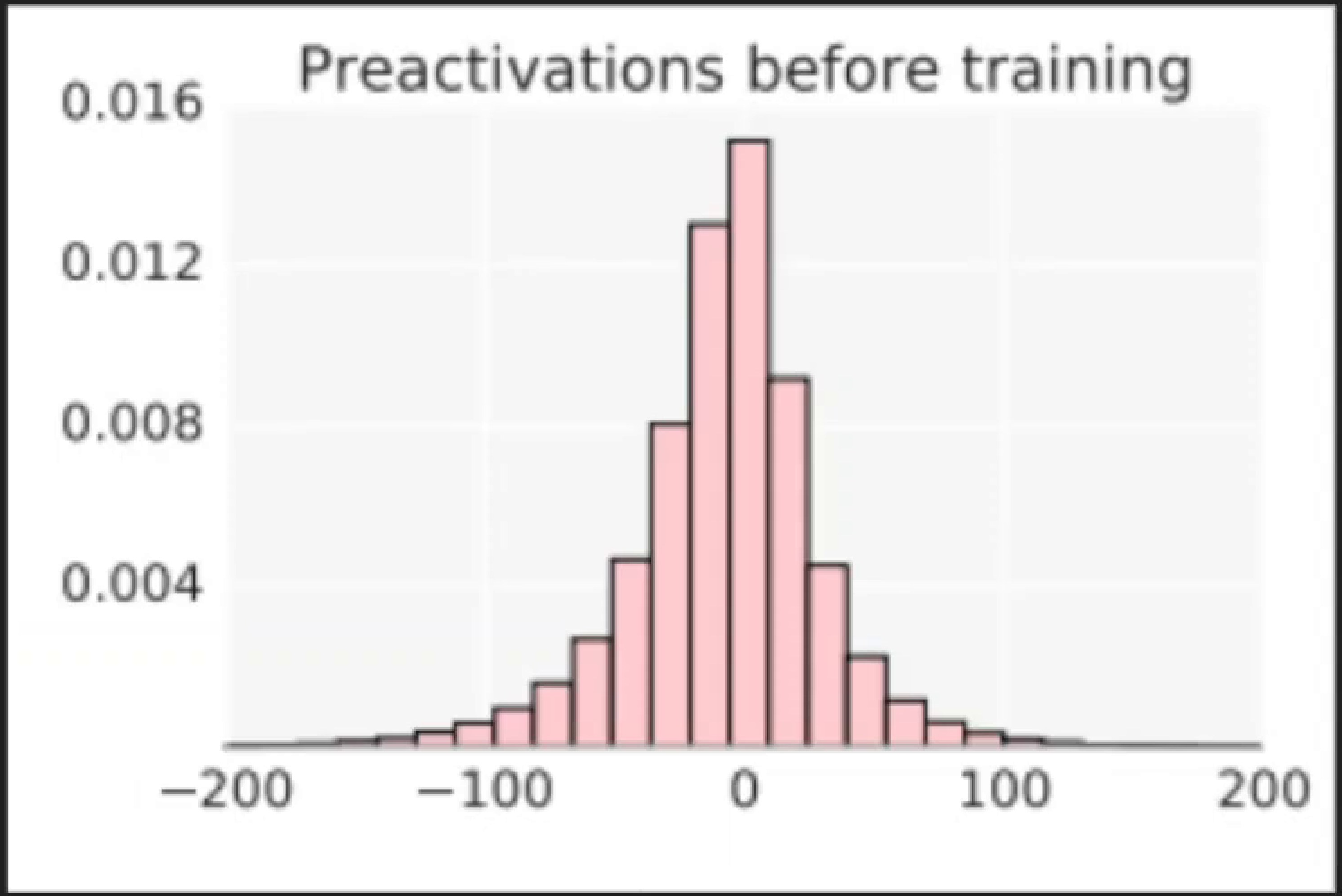



ResNet-50

# Intuition

Input: $X \in R^{m,n}$

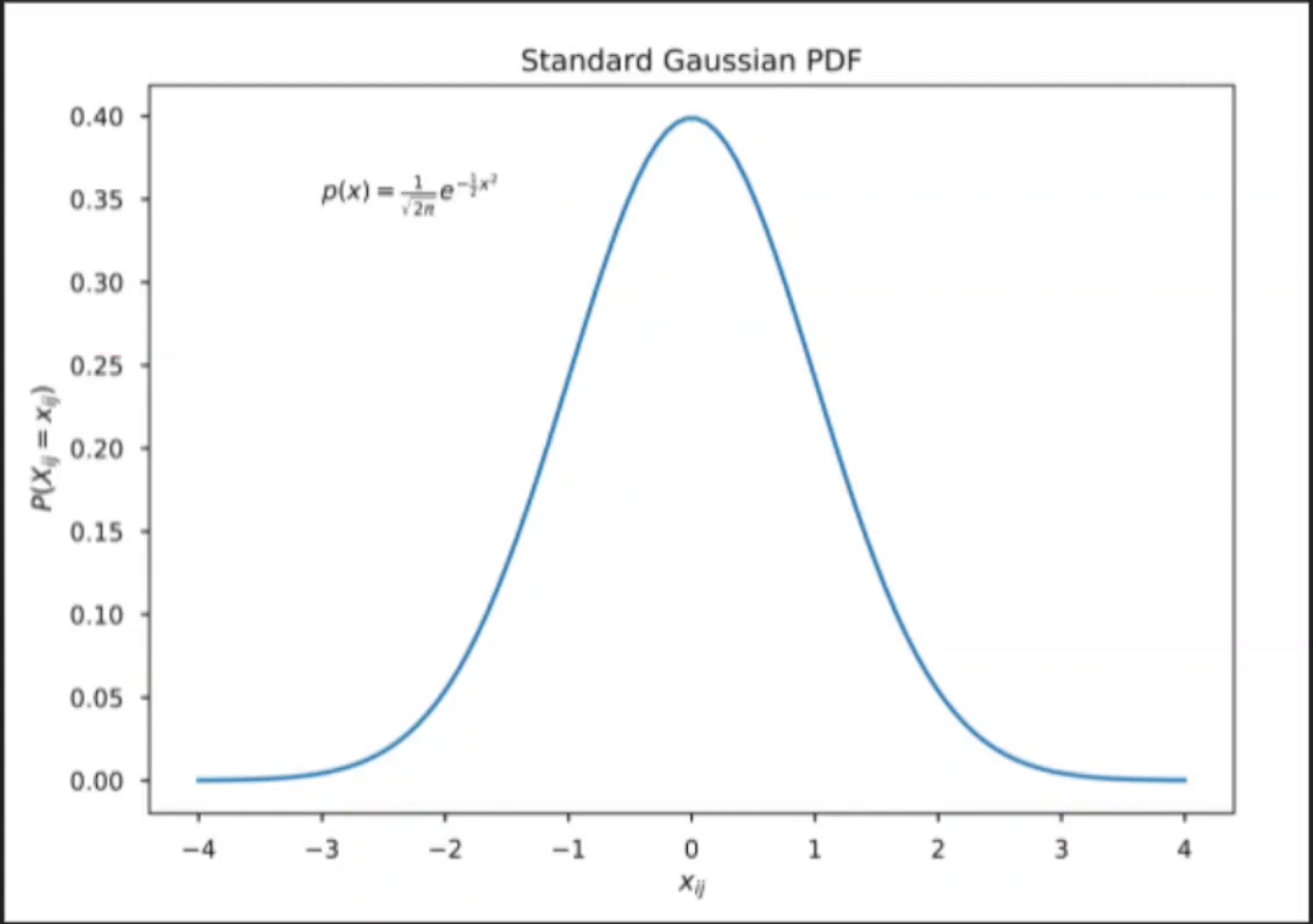
$$x_{ij} \sim N(0,1)$$




ResNet-50

# Intuition

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

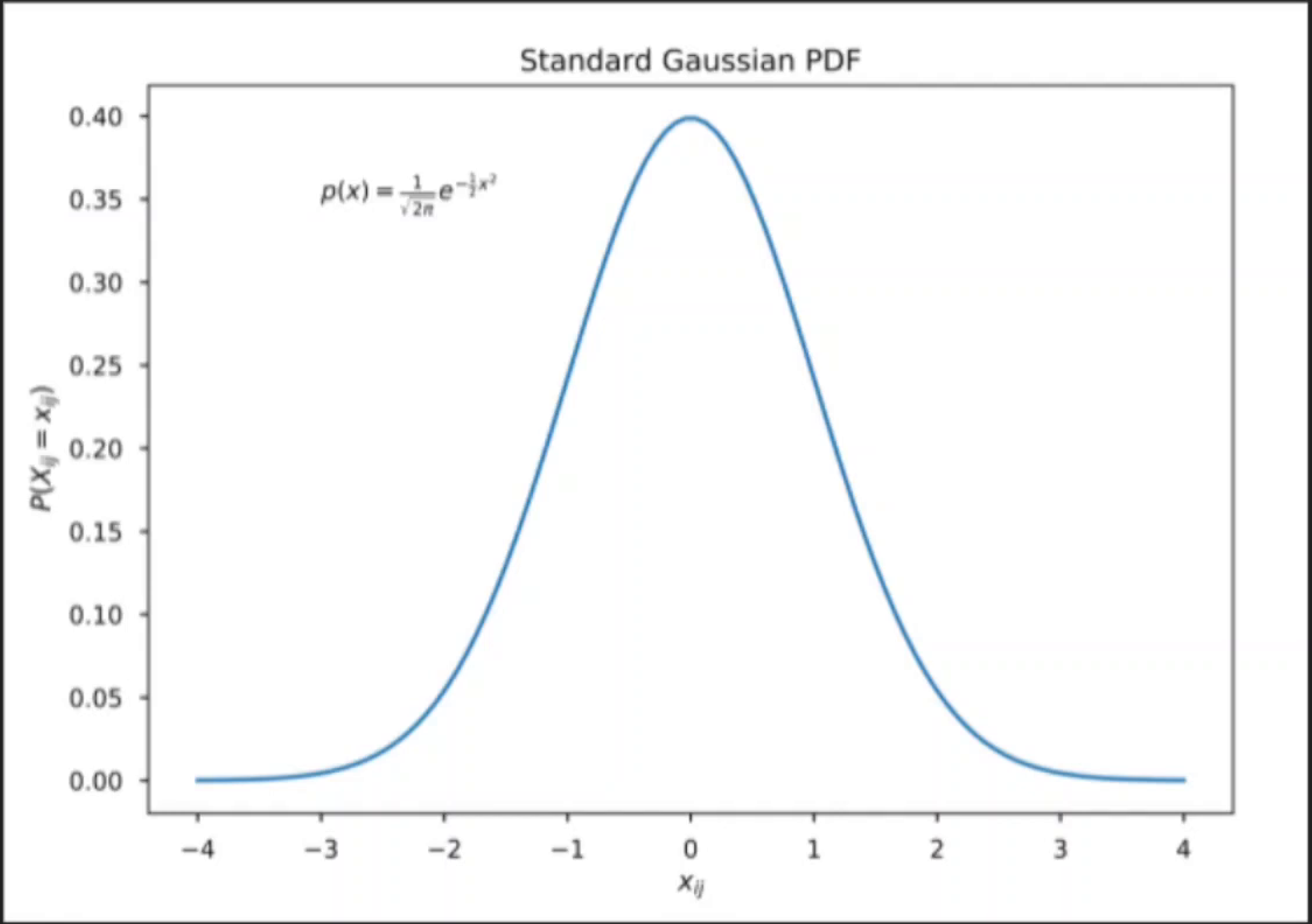


# Intuition

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$\mu = 0, \ \sigma = 1$



# Intuition

zero-to-Identity Mapping

0 for negative inputs

X for positive inputs

# Intuition

zero-to-Identity Mapping → 0 for negative inputs

zero-to-Identity Mapping → X for positive inputs

# Intuition

zero-to-Identity Mapping

- 0 for negative inputs
- X for positive inputs

# Intuition

zero-to-Identity Mapping → 0 for negative inputs

→ X for positive inputs

# Intuition

zero-to-Identity Mapping

- 0 for negative inputs
- X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq x_{ij})$$

# Intuition

zero-to-Identity Mapping

0 for negative inputs

X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq x_{ij})$$

# Intuition

zero-to-Identity Mapping

- 0 for negative inputs
- X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq \ x_{ij})$$

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

# Intuition

zero-to-Identity Mapping

0 for negative inputs

X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq \ x_{ij})$$

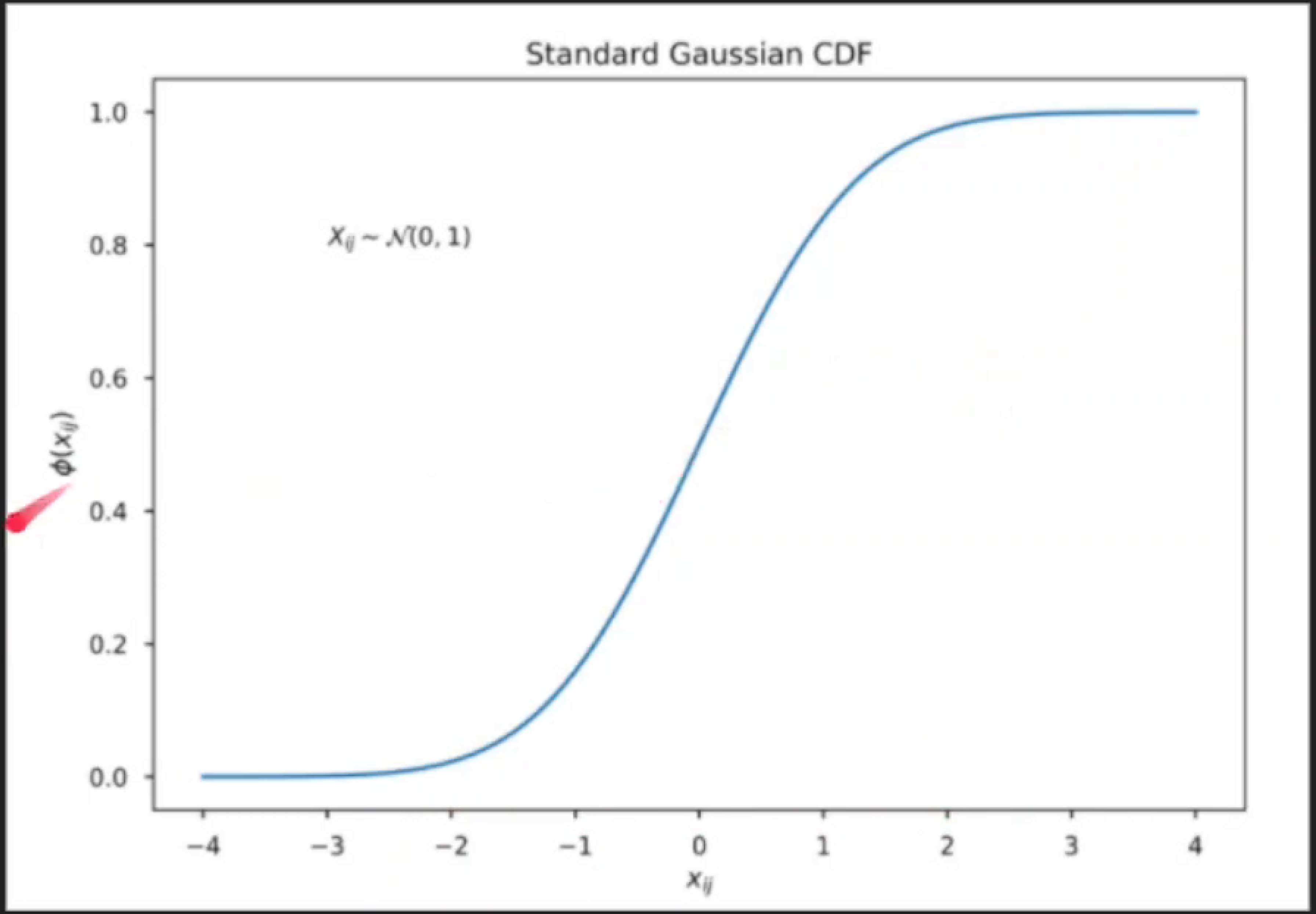
$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

# Intuition

zero-to-Identity Mapping → 0 for negative inputs / X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq \ x_{ij})$$

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$



# Gaussian Error Linear Unit

Weighted sum - x

g(x) = GeLU(x) = x * m

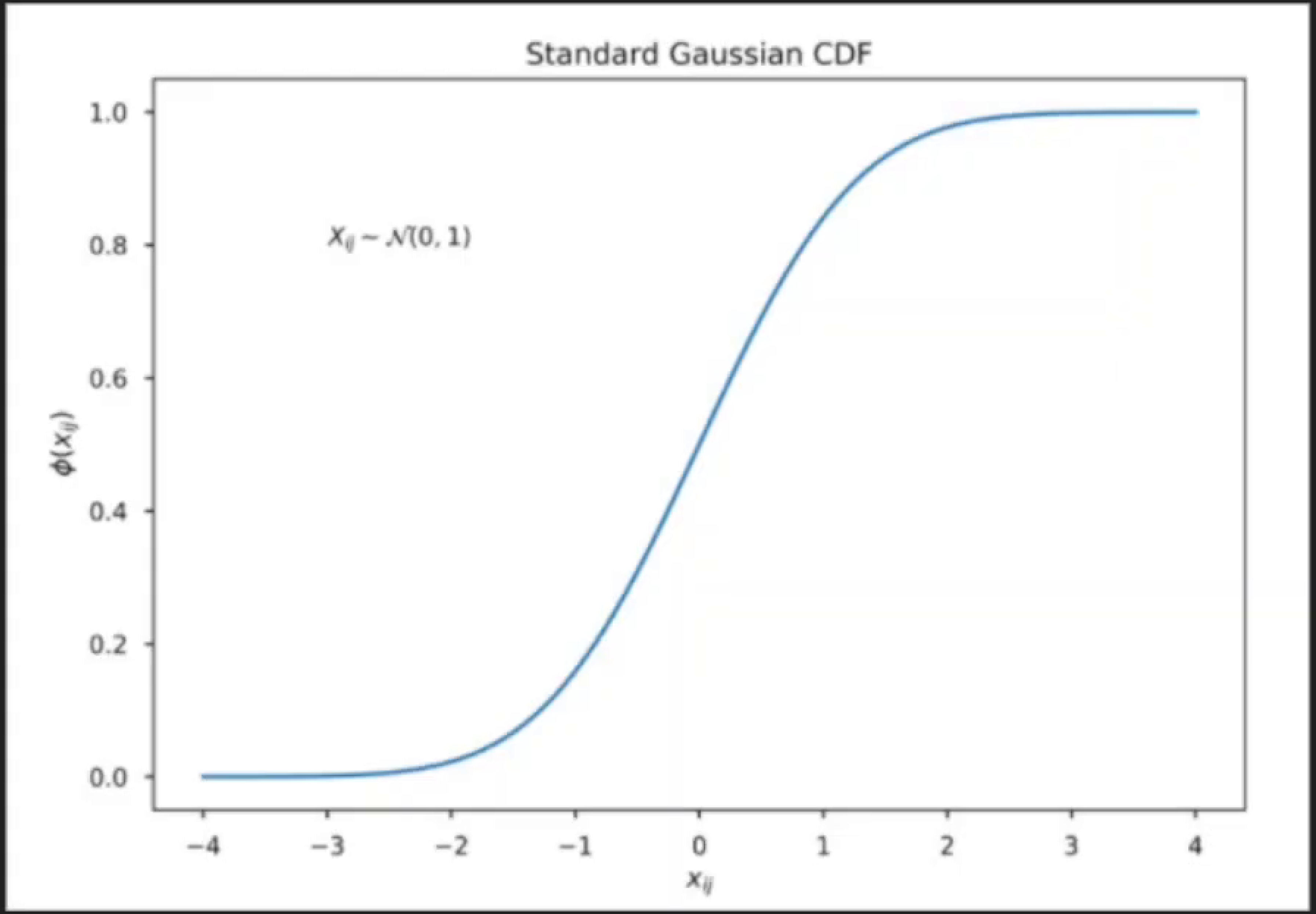
m = Cumulative Distribution Function(Normal Distribution)

# Intuition

zero-to-Identity Mapping → 0 for negative inputs; X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq \ x_{ij})$$

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

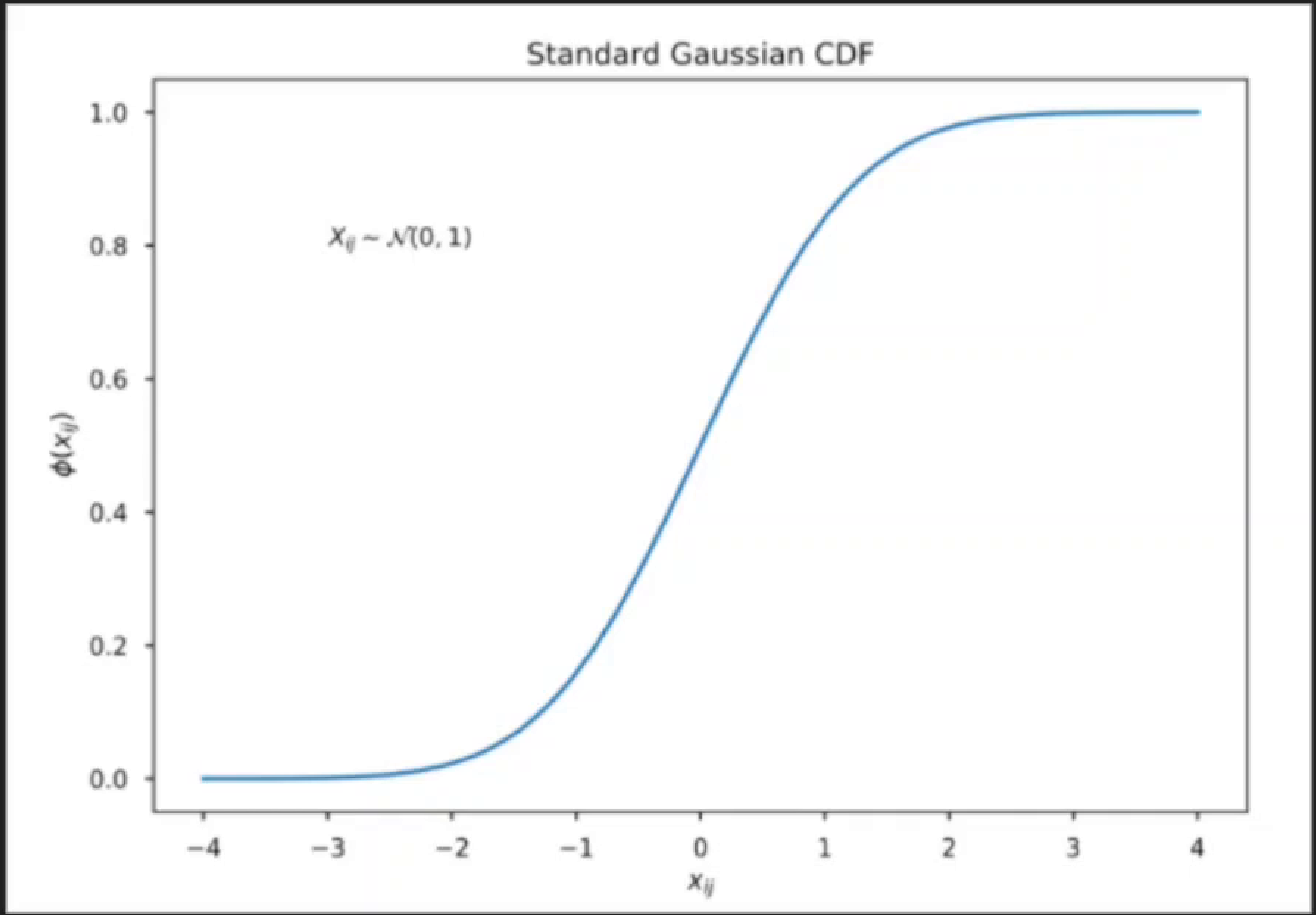


# Intuition

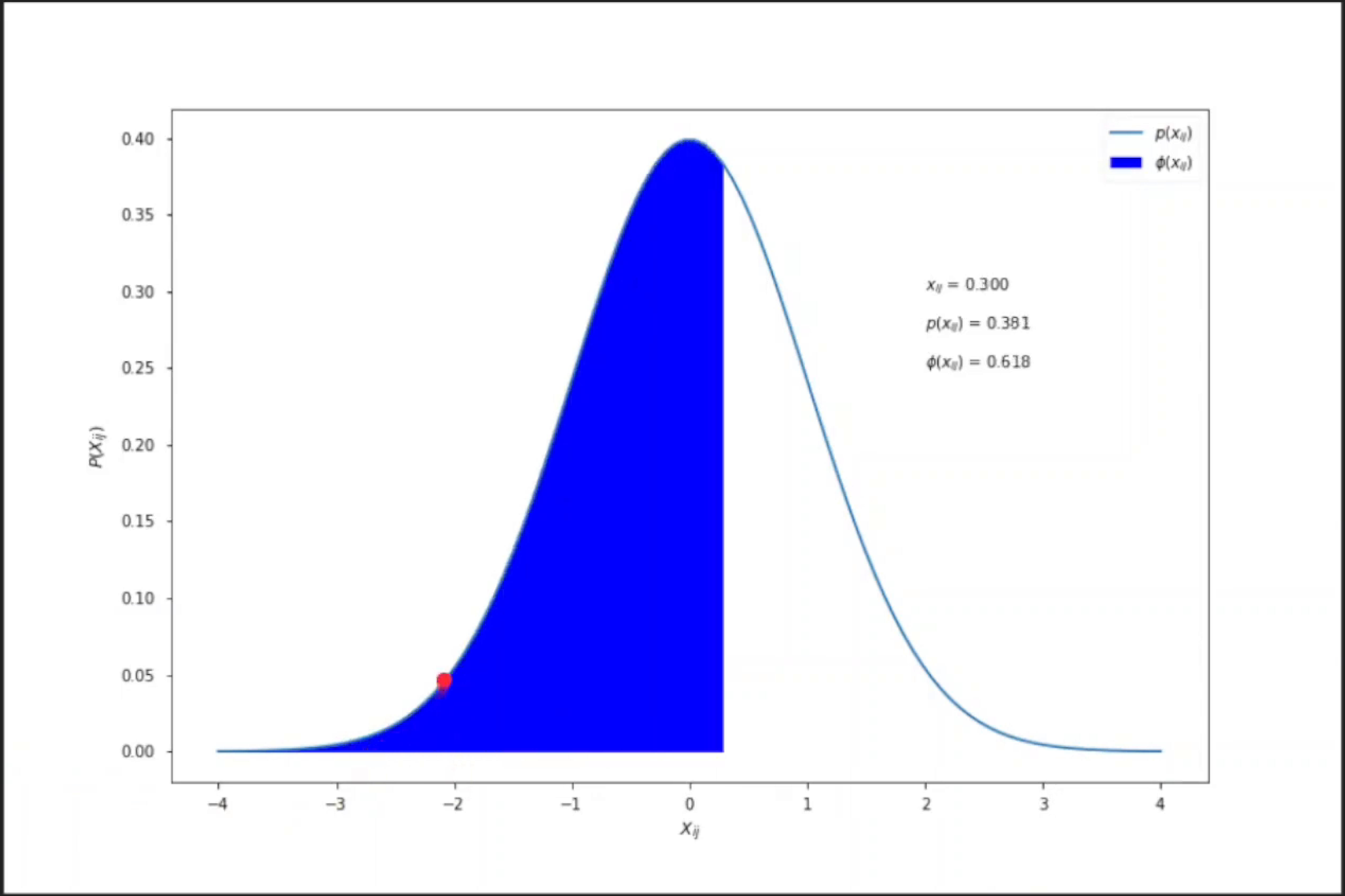
zero-to-Identity Mapping → 0 for negative inputs
→ X for positive inputs

$$\Phi(x_{ij}) = P(X_{ij} \leq \ x_{ij})$$

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

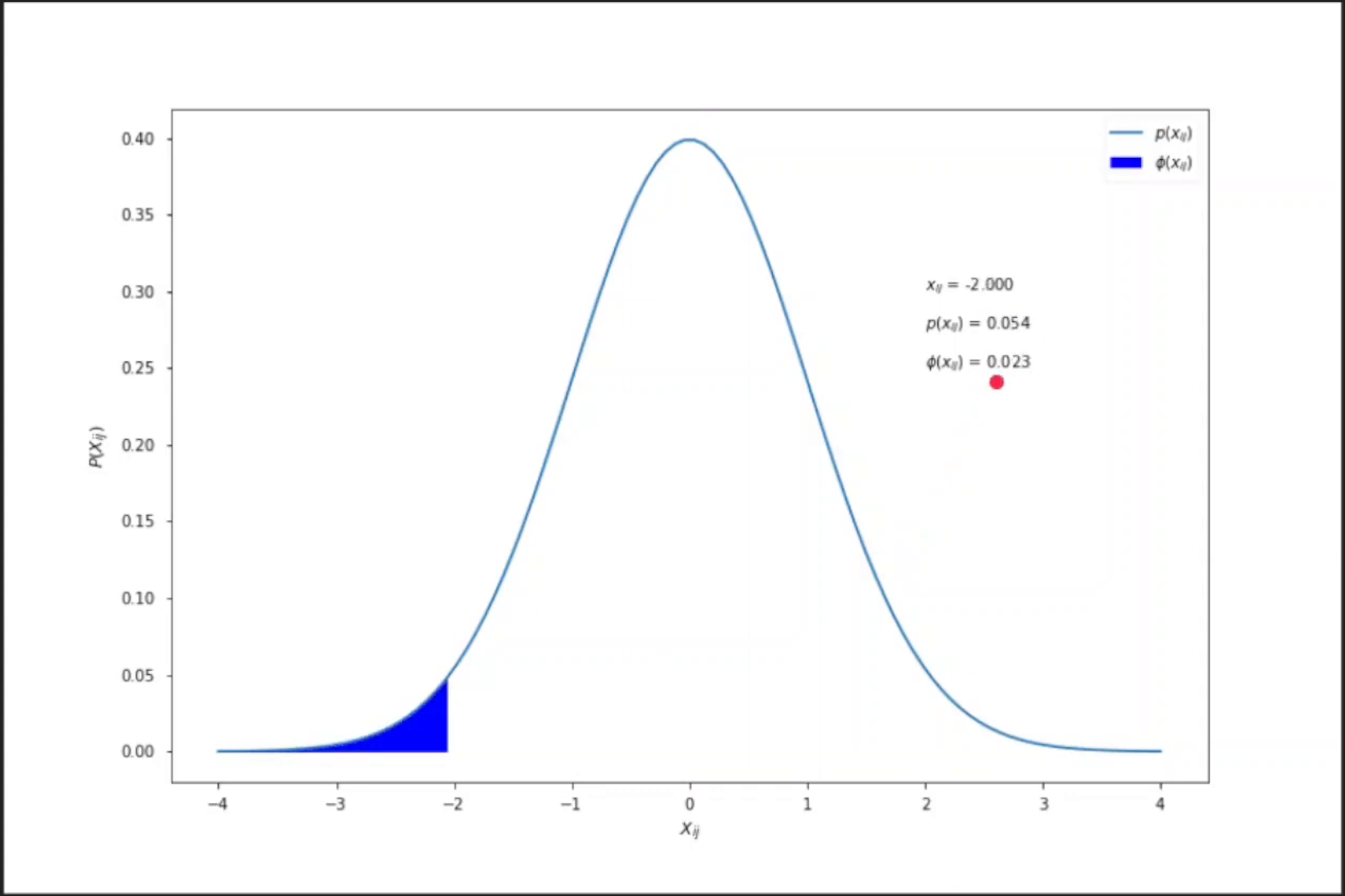


# Intuition



$p(x_{ij})$
$\phi(x_{ij})$
$x_{ij}$ = 0.300
$p(x_{ij})$ = 0.381
$\phi(x_{ij})$ = 0.618
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$P(X_{ij})$
−4
−3
−2
−1
0
1
2
3
4
$X_{ij}$

# Intuition



$p(x_{ij})$
$\phi(x_{ij})$
$x_{ij}$ = -2.000
$p(x_{ij})$ = 0.054
$\phi(x_{ij})$ = 0.023
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$P(X_{ij})$
−4
−3
−2
−1
0
1
2
3
4
$X_{ij}$

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

$$f(x) = x.\Phi(x)$$

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

Gauss Error Function - erf(x)

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

Gauss Error Function - erf(x)

$$erf(x) = \frac{2}{\Pi} \int_{0}^{x} e^{-t^2} dt$$

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

Gauss Error Function - erf(x)

$$erf(x) = \frac{2}{\Pi} \int_0^x e^{-t^2} dt$$

# Intuition

$$\Phi(x_{ij}) = \int_{-\infty}^{x_{ij}} \frac{1}{\sqrt[2]{2\Pi}} e^{-\frac{1}{2}t^2} dt$$

Gauss Error Function - erf(x)

$$erf(x) = \frac{2}{\Pi} \int_0^x e^{-t^2} dt$$

$$GeLU(x) = \frac{1}{2}x\left(1 + erf\left(\frac{x}{\sqrt[2]{2}}\right)\right)$$

$$\Phi(x) = \frac{1 + erf\left(\frac{x}{\sqrt[2]{2}}\right)}{2}$$

# GeLU Activation

$$\text{GELU}(x) = xP(X \leq x) = x\Phi(x) = x \cdot \frac{1}{2}\left[1 + \text{erf}(x/\sqrt{2})\right].$$

# GeLU Activation

$$\mathrm{GELU}(x) = xP(X \leq x) = x\Phi(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right].$$

# GeLU Activation

$$\mathrm{GELU}(x) = xP(X \leq x) = x\Phi(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}(x/\sqrt{2})\right].$$

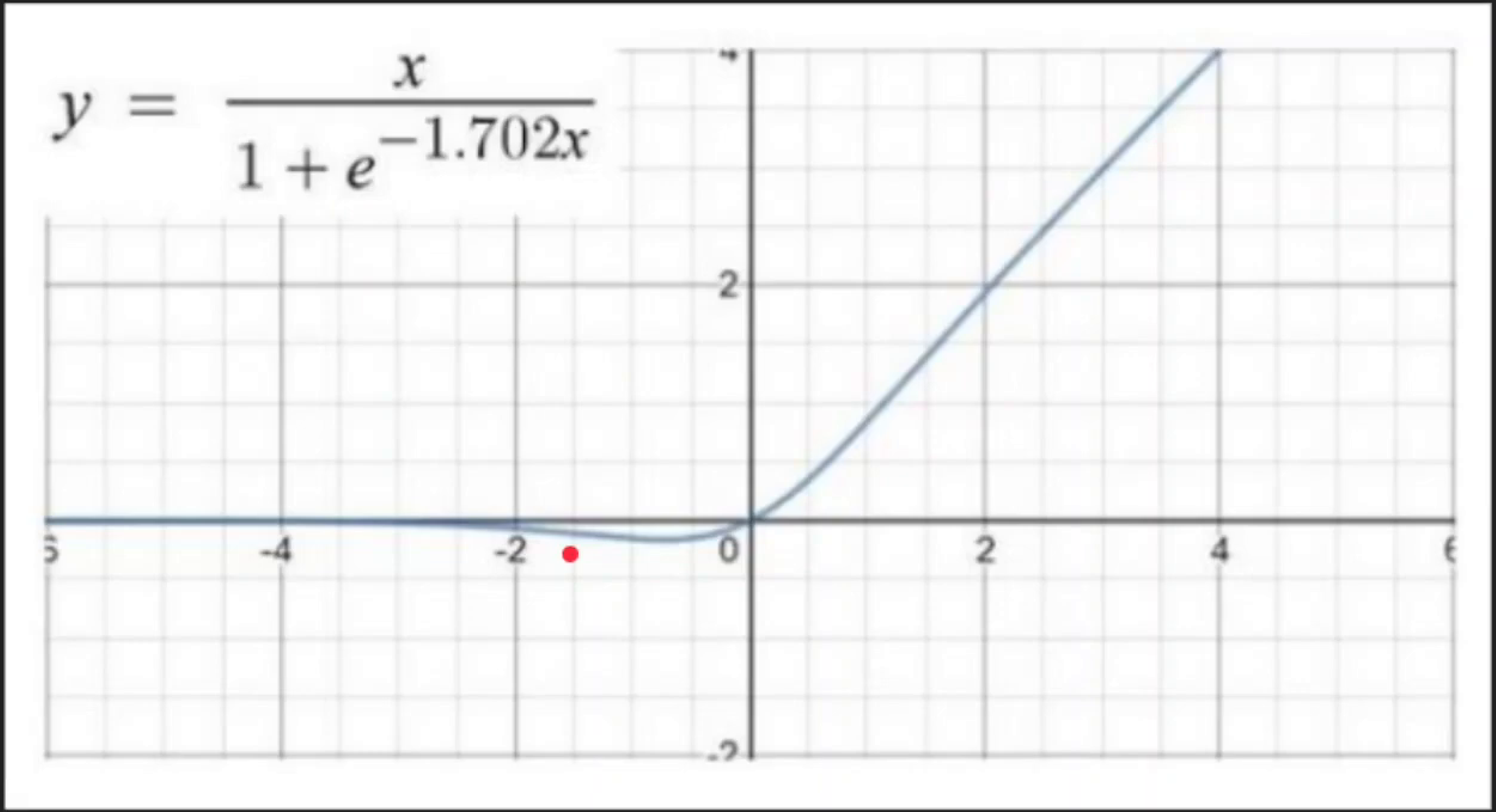
# GeLU Activation

$$\text{GELU}(x) = xP(X \le x) = x\Phi(x) = x \cdot \frac{1}{2}\left[1 + \text{erf}(x/\sqrt{2})\right].$$

$$0.5x(1 + \tanh[\sqrt{2/\pi}(x + 0.044715x^3)])$$

$$x\sigma(1.702x)$$

# Sigmoid Approximation



# Derivative

$$\frac{d}{dx}GELU(x) = \phi(x)\frac{dx}{dx} + x\phi'(x) = \phi(x) + xP(X = x)$$

P(X=x) is value of PDF at x

# Derivative

$$\frac{d}{dx}GELU(x) = \phi(x)\frac{dx}{dx} + x\phi'(x) = \phi(x) + xP(X = x)$$

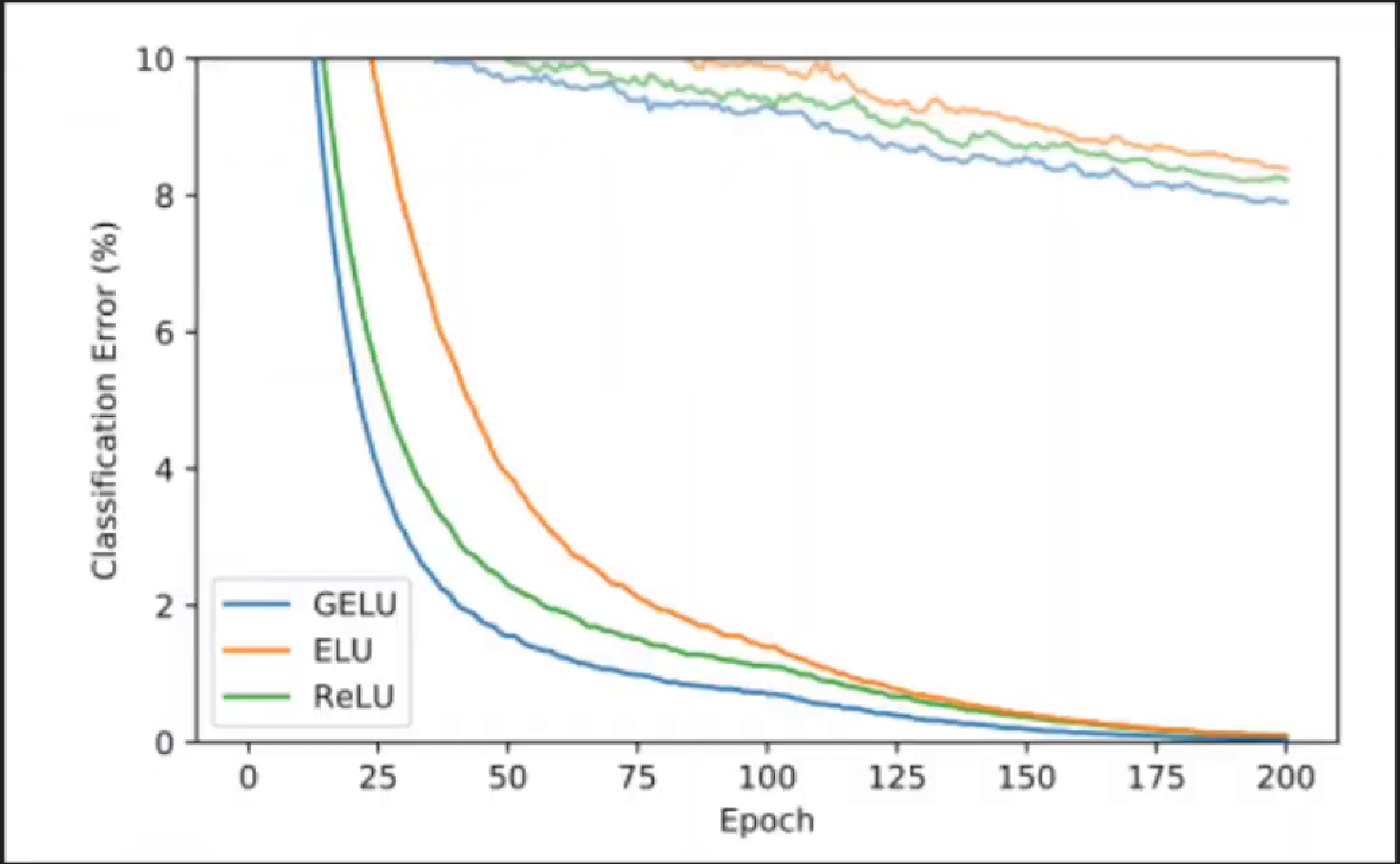
P(X=x) is value of PDF at x

# Derivative

$$\frac{d}{dx}GELU(x) = \phi(x)\frac{dx}{dx} + x\phi'(x) = \phi(x) + xP(X = x)$$

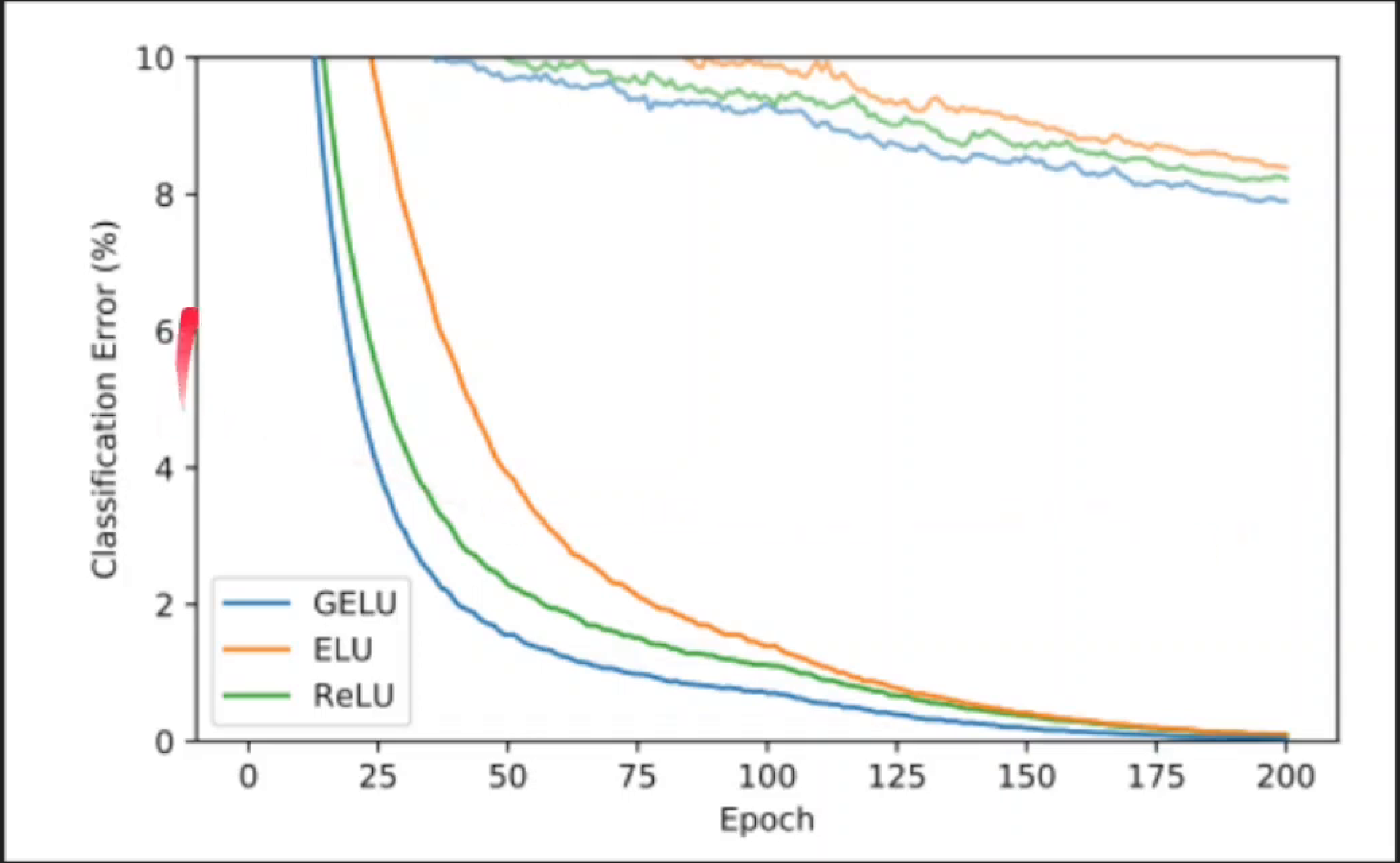
P(X=x) is value of PDF at x

# Experiments



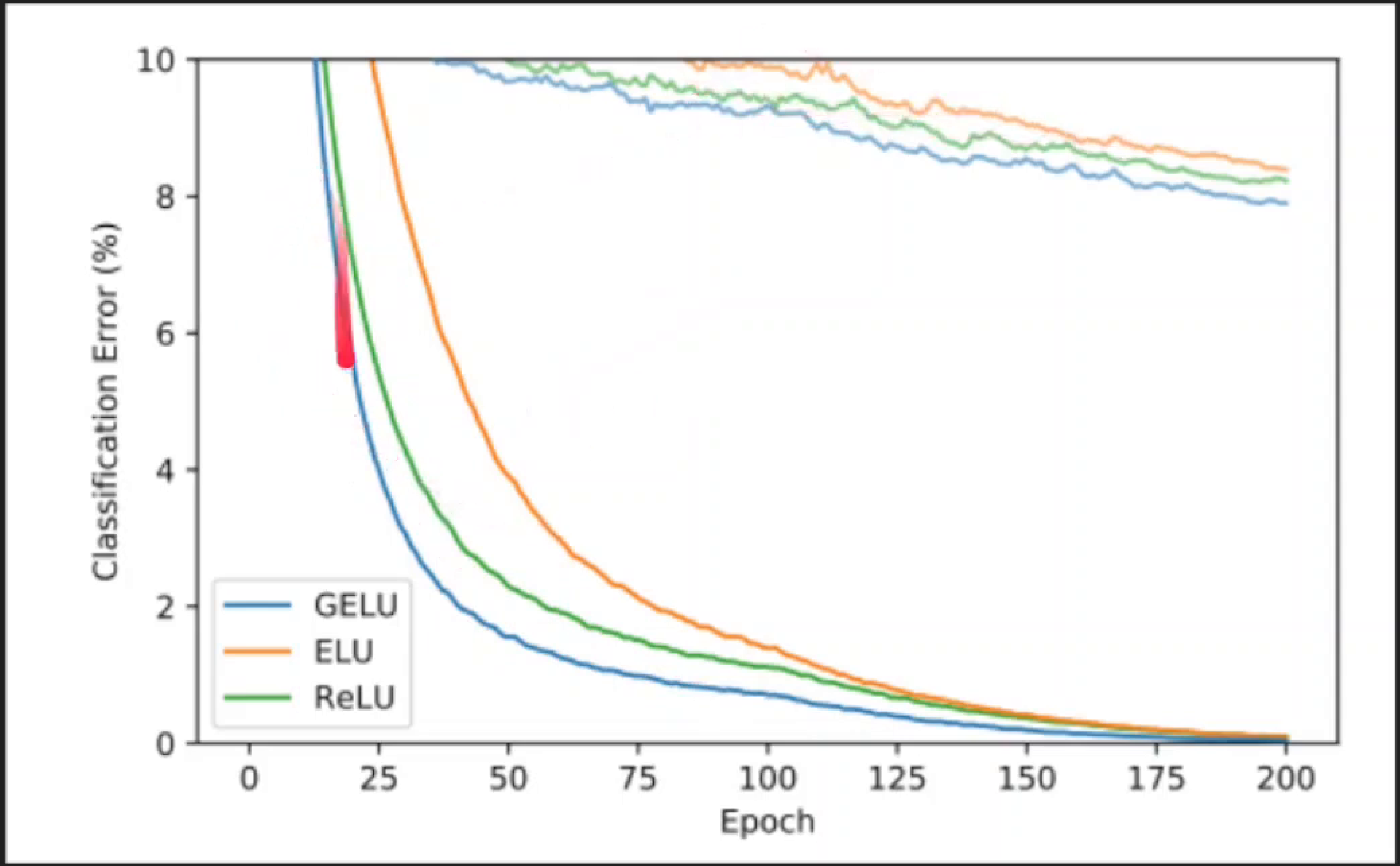

CIFAR-10

# Experiments



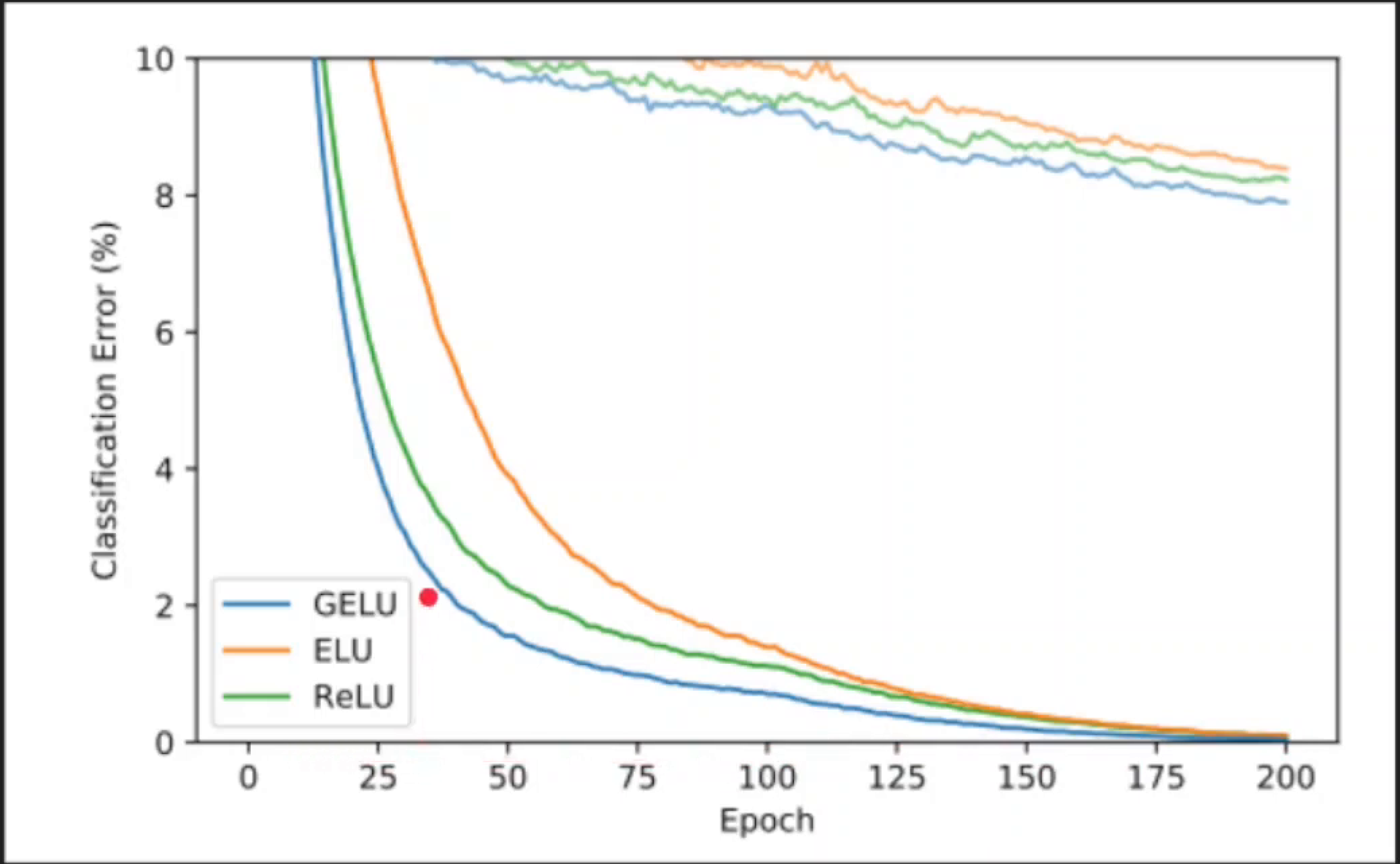

CIFAR-10

# Experiments




CIFAR-10

# Experiments




CIFAR-10

# Python Implementation

```
from scipy.specials import erf
def gelu(x):
    cdf = 0.5 * (1.0 + erf(x / np.sqrt(2.0)))
    return x * cdf
```

```
#Using TensorFlow
import tensorflow as tf
def gelu(x):
    cdf = 0.5 * (1.0 + tf.erf(x / tf.sqrt(2.0)))
    return x * cdf
```

# Python Implementation

```
from scipy.specials import erf
def gelu(x):
    cdf = 0.5 * (1.0 + erf(x / np.sqrt(2.0)))
    return x * cdf
```

```
#Using TensorFlow
import tensorflow as tf
def gelu(x):
    cdf = 0.5 * (1.0 + tf.erf(x / tf.sqrt(2.0)))
    return x * cdf
```

# Python Implementation

```
def sig(x):
  s = (1/(1+np.exp(-x)))   # sigmoid function
  return s

def gelu(x):
  g = x * sig(1.702*x)
  return g
```

# Python Implementation

```
def sig(x):
  s = (1/(1+np.exp(-x)))   # sigmoid function
  return s

def gelu(x):
  g = x * sig(1.702*x)
  return g
```

GeLU Activation - Google Slides

Neuron.ipynb - Colaboratory

colab.research.google.com/drive/109Wvb6_MPN9nDPOxf4wy2r6Ip7MZz4fs#scrollTo=vJK7kjvSbWhq

Neuron.ipynb

File Edit View Insert Runtime Tools Help

Comment Share

+ Code + Text

RAM Disk Editing

[2]

```
import numpy as np
import matplotlib.pyplot as plt

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

def gelu(x):
  g = x * sig(1.702*x)
  return g
```

```
z=np.arange(-4,4,0.01)

fig, ax = plt.subplots(figsize=(9, 5))
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.plot(z,gelu(z), color="#307EC7", linewidth=3, label="GeLU")
ax.legend(loc="upper left", frameon=False)
fig.show()
```

[ ]

Waiting for colab.research.google.com...

0s completed at 8:23 AM

Type here to search

21°C Clear

8:23 AM
2/24/2022

```
fig, ax = plt.subplots(figsize=(9, 5))
ax.xaxis.set_ticks_position('bottom')
ax.yaxis.set_ticks_position('left')
ax.plot(z,gelu(z), color="#307EC7", linewidth=3, label="GeLU")
ax.legend(loc="upper left", frameon=False)
fig.show()
```