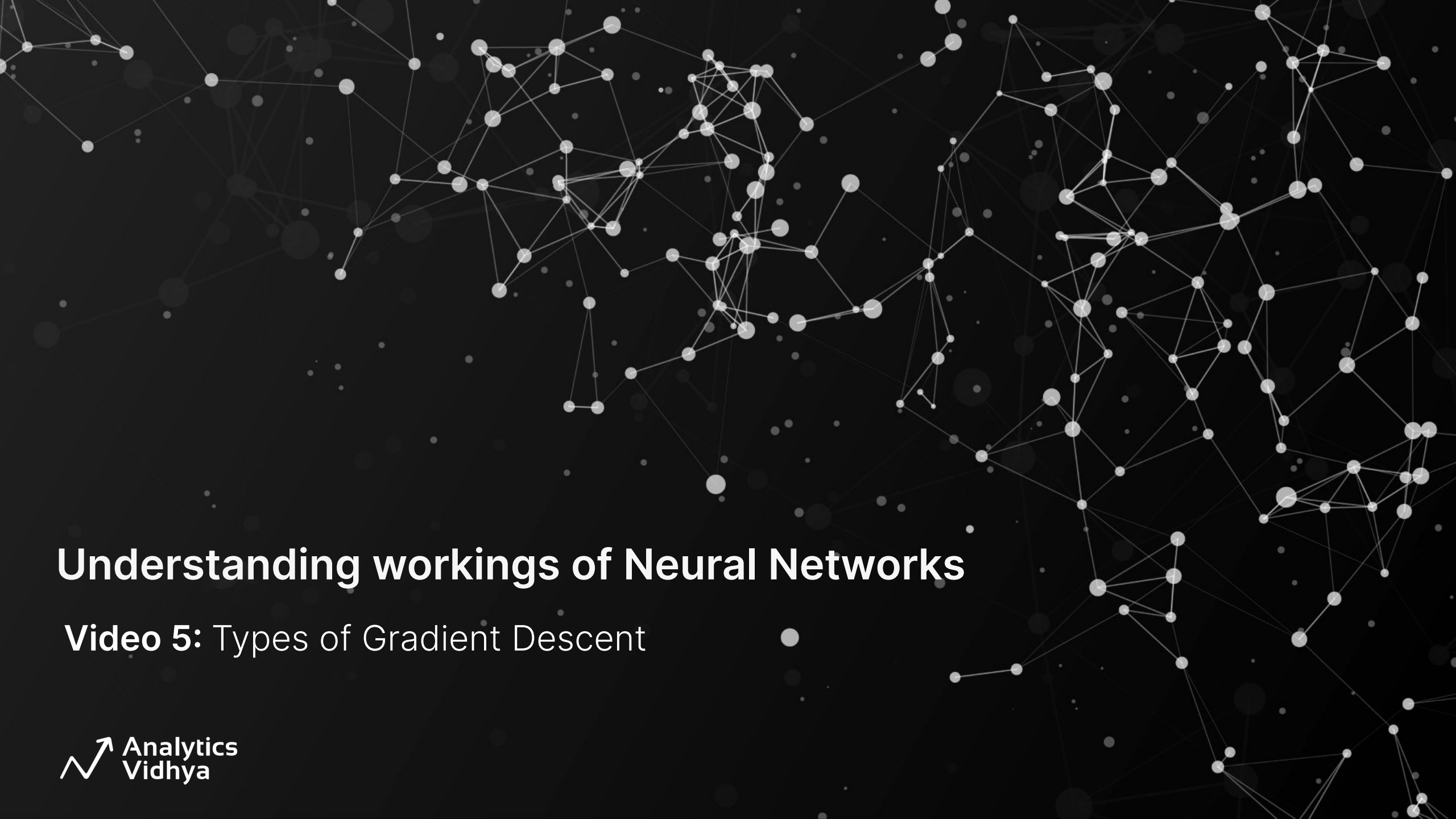


Understanding workings of Neural Networks

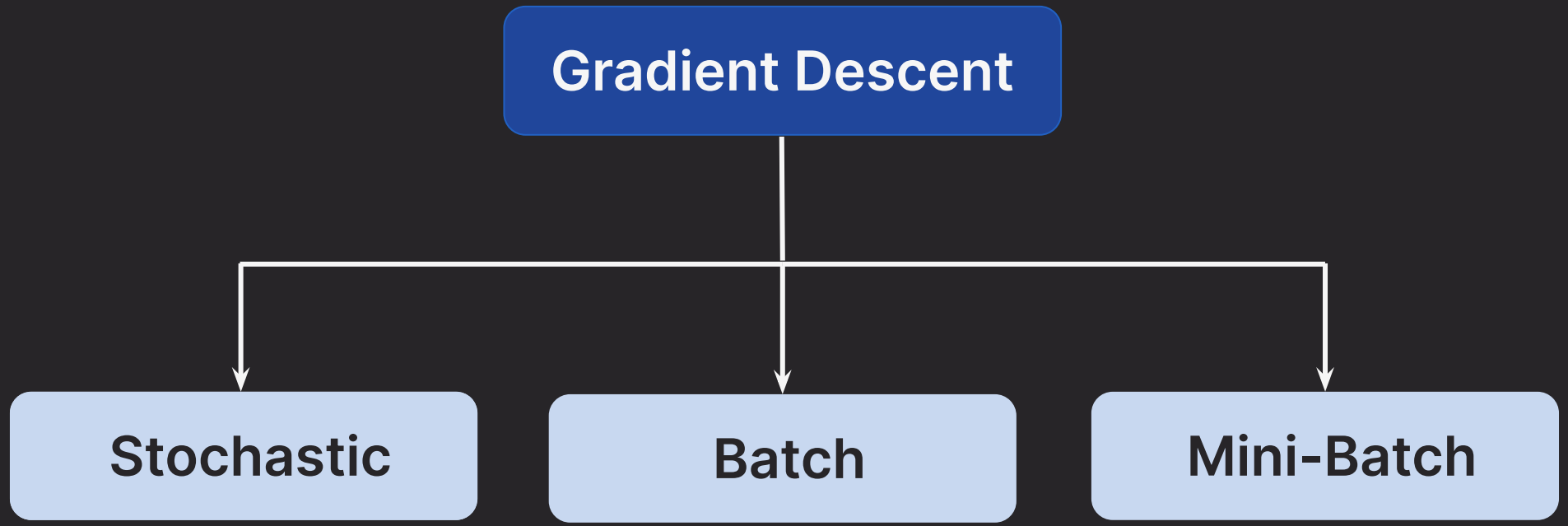
Video 5: Types of Gradient Descent

Analytics Vidhya

Analytics Vidhya

Gradient Descent: Classification

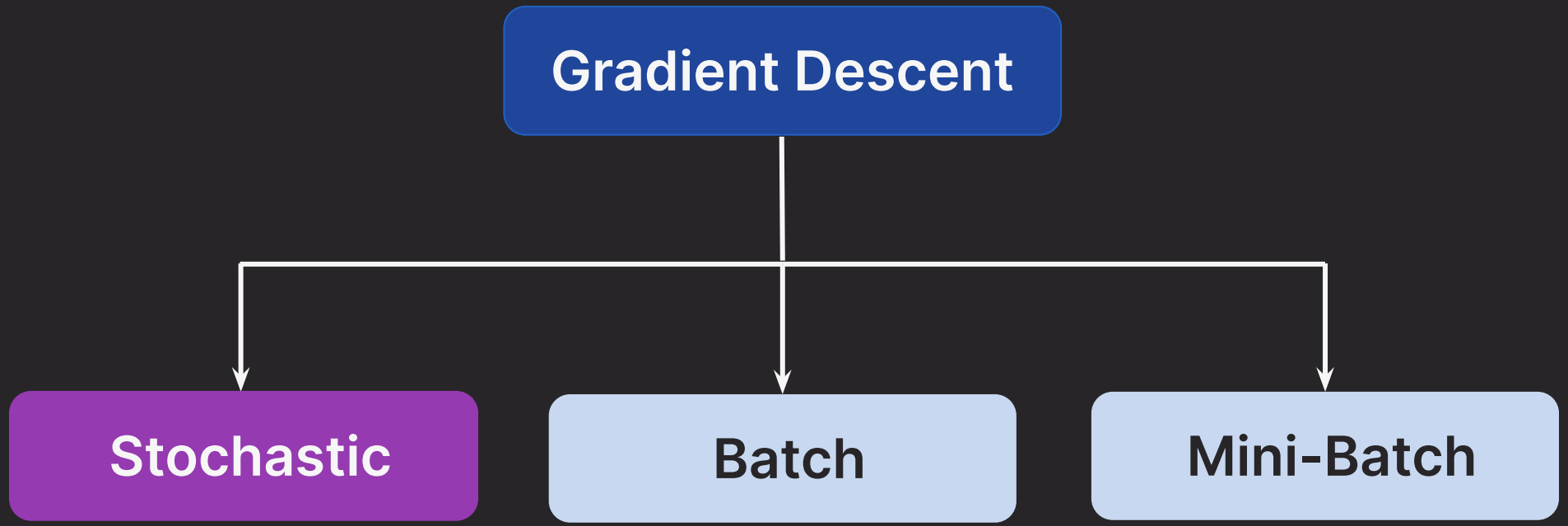
Categorized based on amount of data to compute the gradient of loss function



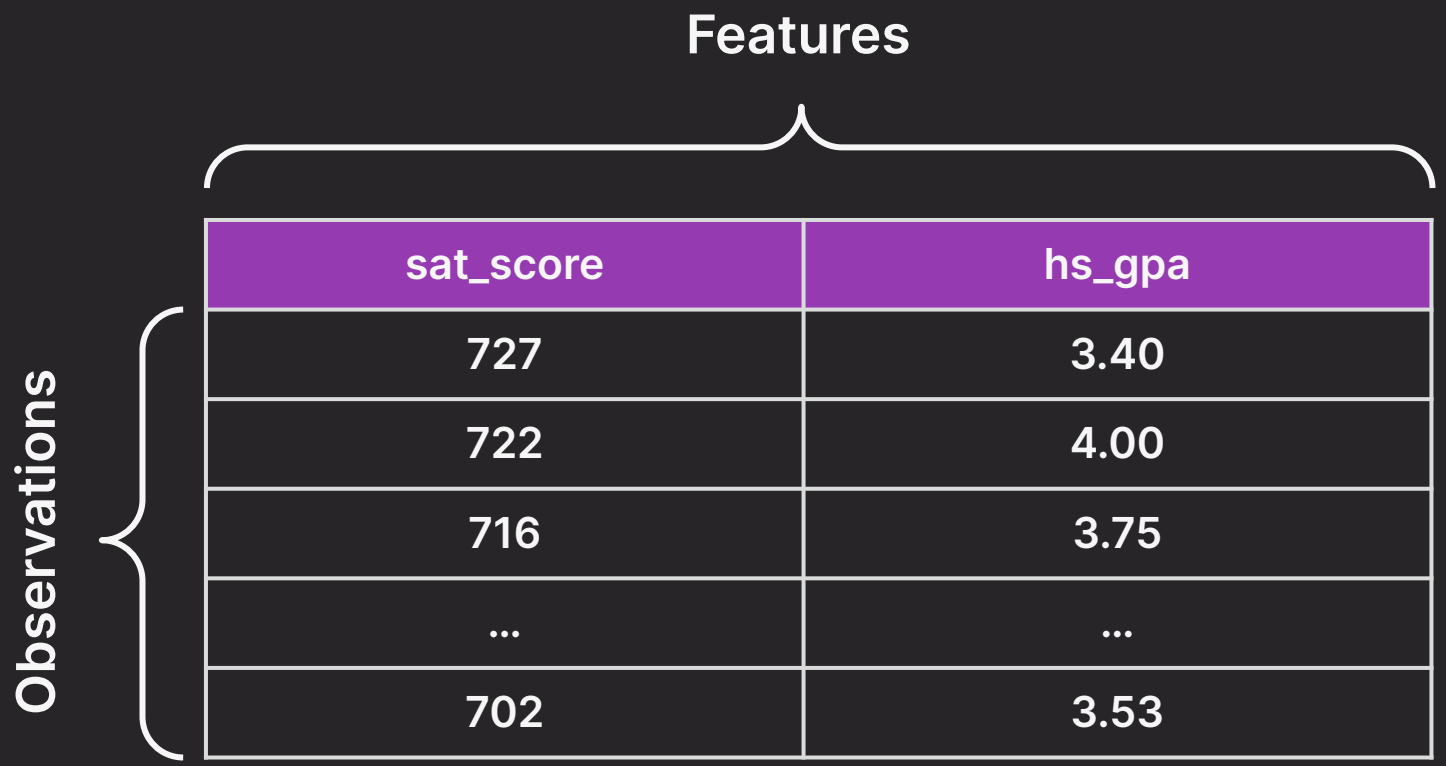
Analytics Vidhya

Gradient Descent: Classification

Categorized based on amount of data to compute the gradient of loss function



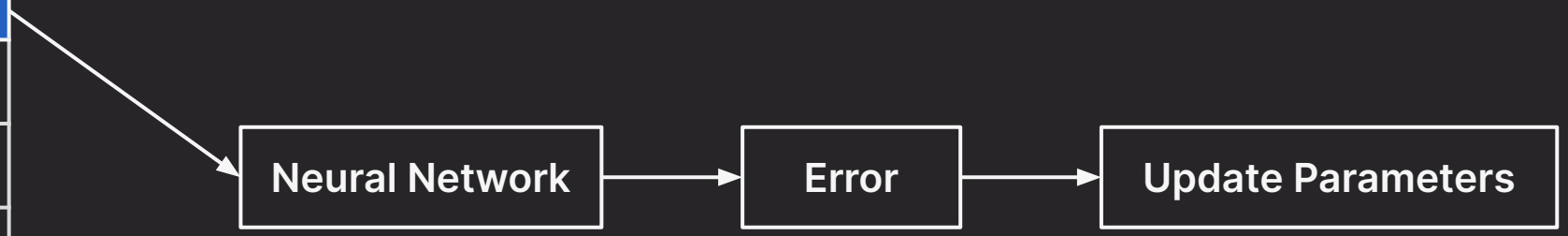
Stochastic Gradient Descent



Analytics Vidhya

Stochastic Gradient Descent

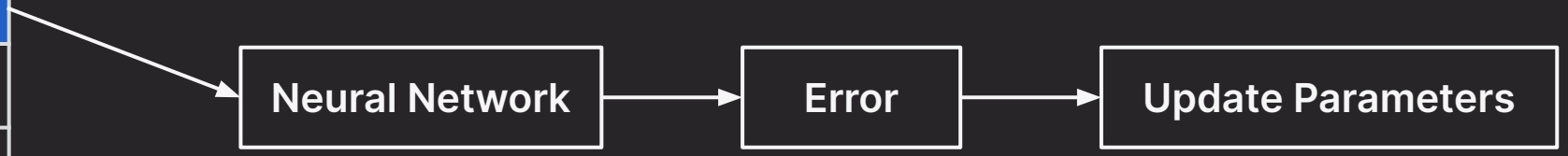
sat_score	hs_gpa
727	3.40
722	4.00
716	3.75
...	...
702	3.53



Analytics Vidhya

Stochastic Gradient Descent

sat_score	hs_gpa
727	3.40
722	4.00
716	3.75
...	...
702	3.53



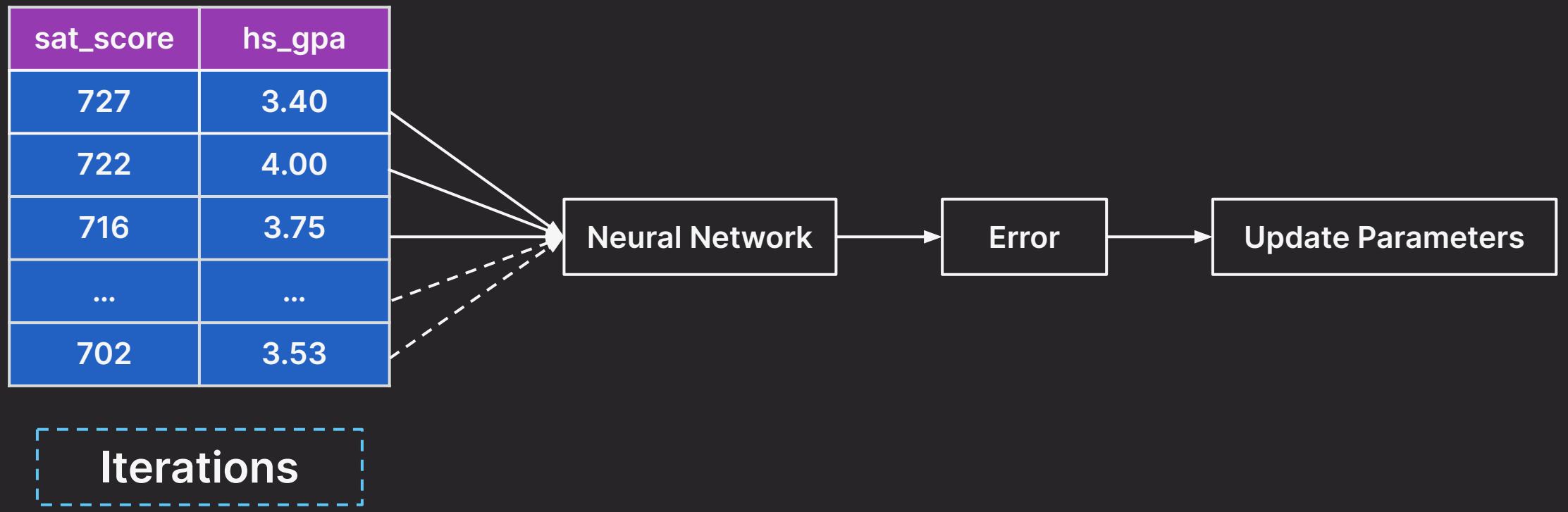
Analytics Vidhya

Stochastic Gradient Descent

sat_score	hs_gpa
727	3.40
722	4.00
716	3.75
...	...
702	3.53

Analytics Vidhya

Stochastic Gradient Descent



Analytics Vidhya

Stochastic Gradient Descent

'm' iterations per epoch

m_i Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
m_i	m_i	m_i

Stochastic Gradient Descent

One Epoch

Each sample has gone through one pass of forward and backward propagation.

Stochastic Gradient Descent

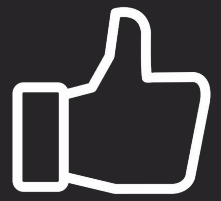
m_i Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
m_i	m_i	m_i

SGD Updation = 50*10 = 500 times

Stochastic Gradient Descent

Recommended Practice:



Add a shuffle parameter before selection of the records

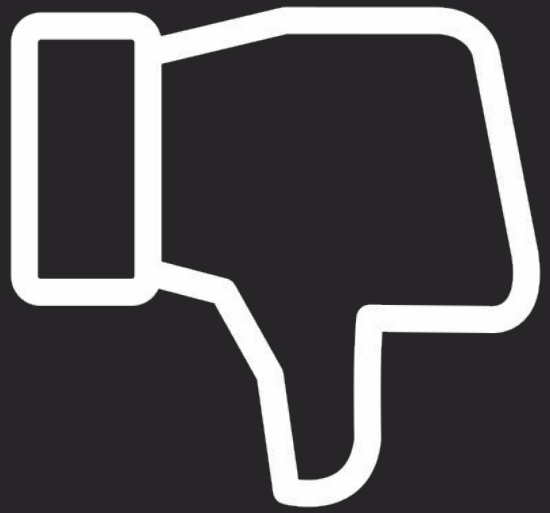
Analytics Vidhya

Stochastic Gradient Descent: Advantage

Improved model performance because of the frequent updation of weights and biases

Analytics Vidhya

Stochastic Gradient Descent: Advantage

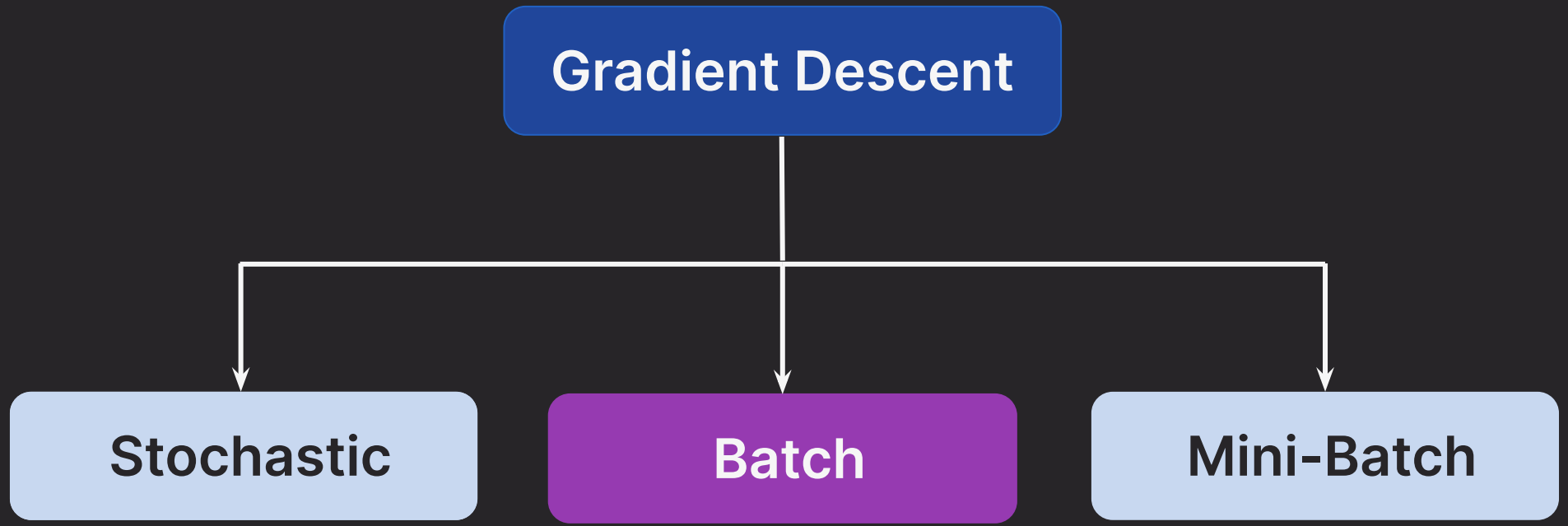


Updating the weights based on just one sample can result in fluctuations in loss values

Analytics Vidhya

Gradient Descent: Classification

Categorized based on amount of data to compute the gradient of loss function



Analytics Vidhya

Batch Gradient Descent

50 Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

Batch Gradient Descent Updation = 10 times for 10 epochs

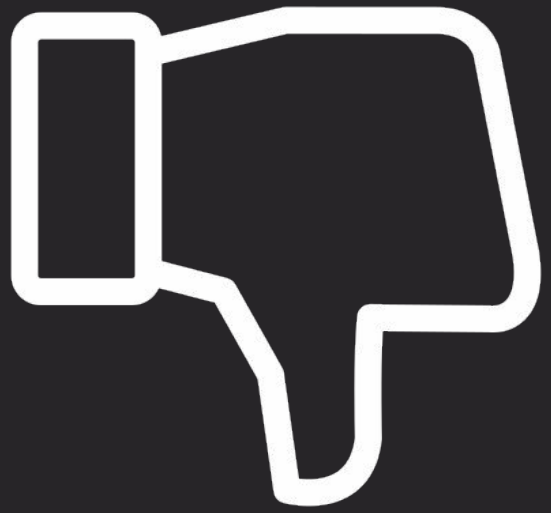
Analytics Vidhya

Batch Gradient Descent: Advantage

- Computational Efficiency
- Stable Error Gradient
- Faster Convergence

Analytics Vidhya

Batch Gradient Descent: Drawback

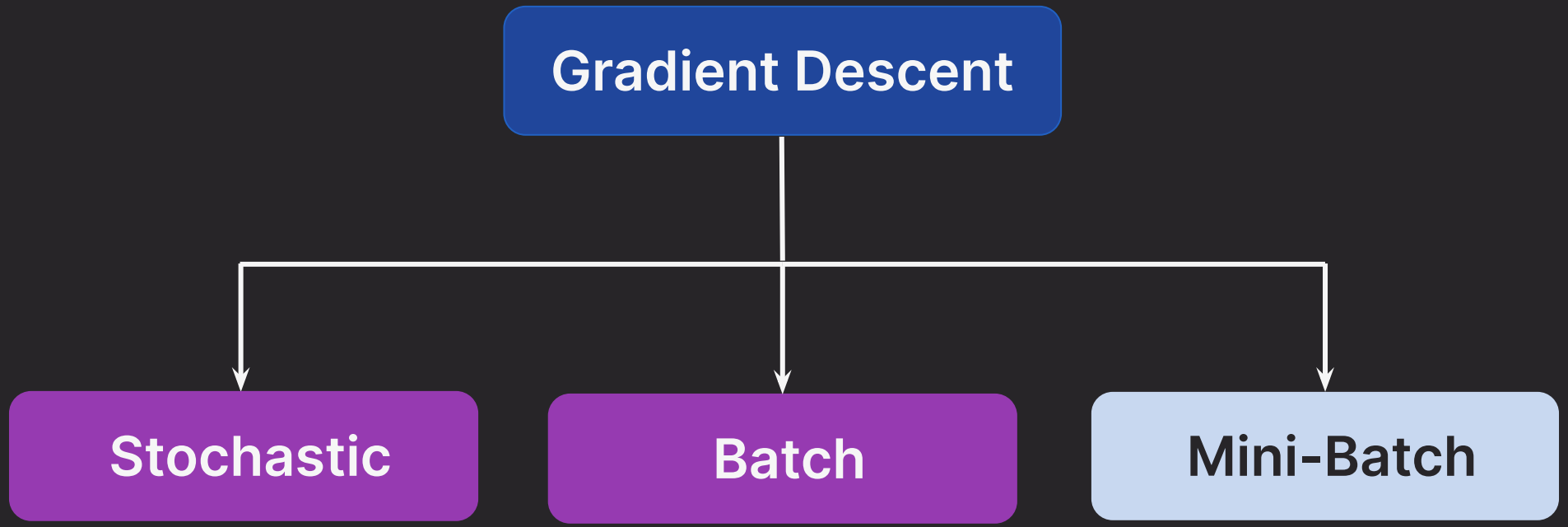


- Would need several epochs for training
- Requires the full dataset in memory, limiting scalability

Analytics Vidhya

Gradient Descent: Classification

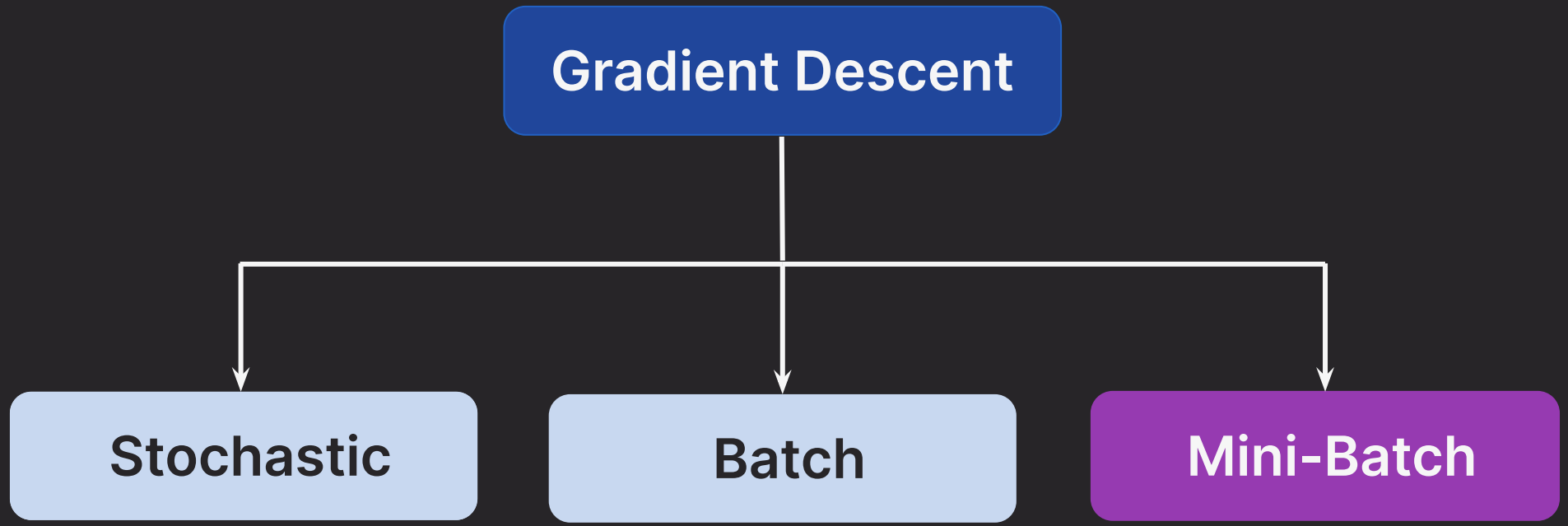
Categorized based on amount of data to compute the gradient of loss function



Analytics Vidhya

Gradient Descent: Classification

Categorized based on amount of data to compute the gradient of loss function



Analytics Vidhya

Mini-Batch Gradient Descent

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

Subset 1

Subset n

Analytics Vidhya

Mini-Batch Gradient Descent

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

Subset 1

Subset n

50 samples / 10 = 5 subsets

Number of epochs = 5

Analytics Vidhya

Mini-Batch Gradient Descent

800 Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

800 samples ; batch_size = 80

How many times will the weights be updated in 10 epochs?

Analytics Vidhya

Mini-Batch Gradient Descent

800 Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

800 samples ; batch_size=80 → 10 subsets

10 subsets * 10 epochs = 100 times

Analytics Vidhya

Mini-Batch Gradient Descent

Observations

sat_score	hs_gpa	fy_gpa
727	3.40	3.18
722	4.00	3.33
716	3.75	3.25
...	...	...
...	...	...
710	3.15	3.05

Subset 1

Subset n

Mini batch size of 2^n is preferred

eg: 16, 32, 64, 128, 256, 512, 1024.....

Analytics Vidhya

Hands-on