

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

1 1 1 1

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

ReLU

$\approx$

Normalize

μ

$-$

σ

$\div$

1 1 1 1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Next Layer

Batch Statistics

Σ

μ

σ^2

σ

Sum (Σ)

Mean (μ)

Variance (σ^2)

Std Dev (σ)

Trainable
Parameters

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

↓	↓	↓	↓
3	1	3	2
0	2	3	0
-2	5	2	0

ReLU

≈

Normalize

μ

-

σ

÷

1 1 1 1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Next Layer

Batch Statistics

Σ

μ

σ^2

σ

Sum (Σ)

Mean (μ)

Variance (σ^2)

Std Dev (σ)

Trainable
Parameters

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

↓	↓	↓	↓
3	1	3	2
0	2	3	0
-2	5	2	0

ReLU

3	1	3	2
0	2	3	0
0	5	2	0

Batch Statistics

Σ	μ	σ^2	σ

Sum (Σ)
Mean (μ)
Variance (σ^2)
Std Dev (σ)

Normalize

μ	
-	

σ	
÷	

1 1 1 1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Trainable
Parameters

Next Layer

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

1	1	1	1
3	1	3	2
0	2	3	0
-2	5	2	0

ReLU

3	1	3	2
0	2	3	0
5	2	0	

Batch Statistics

Σ	μ	σ^2	σ
9	2	1	1
5	1	1	1
7	2	4	2

Sum (Σ)
Mean (μ)
Variance (σ^2)
Std Dev (σ)

Normalize

μ	
-	

σ	
$\div$	

1 1 1 1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Trainable
Parameters

Next Layer

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

1	1	1	1
3	1	3	2
0	2	3	0
-2	5	2	0

ReLU

3	1	3	2
0	2	3	0
5	2	0	

Batch Statistics

Σ	μ	σ^2	σ
9	2	1	1
5	1	1	1
7	2	4	2

Normalize

μ	2
-	1
	2

1	-1	1	0
-1	1	2	-1
-2	3	0	-2

σ	
$\div$	

Sum (Σ)
Mean (μ)
Variance (σ^2)
Std Dev (σ)

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Trainable
Parameters

Next Layer

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

1	1	1	1
3	1	3	2
0	2	3	0
-2	5	2	0

ReLU

3	1	3	2
0	2	3	0
5	2	0	

Batch Statistics

Σ	μ	σ^2	σ
9	2	1	1
5	1	1	1
7	2	4	2

Sum (Σ)
Mean (μ)
Variance (σ^2)
Std Dev (σ)

Normalize

μ	2
-	1
	2

1	-1	1	0
-1	1	2	-1
-2	3	0	-2

σ	1
$\div$	1
	2

1	-1	1	0
-1	1	2	-1
-1	1	0	-1

1 1 1 1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

Trainable
Parameters

Next Layer

Batch Normalization

Mini-batch: X_1 X_2 X_3 X_4

1	0	3	0
0	3	1	1
2	1	0	2

Linear Layer

1	0	1	0
1	1	0	-1
0	2	-1	0

1	0	3	0
0	3	1	1
2	1	0	2

ReLU

3	1	3	2
0	2	3	0
5	2	0	0

Batch Statistics

Σ	μ	σ^2	σ
9	2	1	1
5	1	1	1
7	2	4	2

Sum (Σ)
Mean (μ)
Variance (σ^2)
Std Dev (σ)

Normalize

μ	2
-	1
	2

1	-1	1	0
-1	1	2	-1
-2	3	0	-2

σ	1
$\div$	1
	2

1	-1	1	0
-1	1	2	-1
-1	1	0	-1

Scale & Shift

2	0	0	0
0	3	0	0
0	0	-1	1

2	-2	2	0
-3	3	6	-3
2	0	1	2

Trainable
Parameters



Next Layer