

# 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

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

ReLU  
 $\approx$

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

Batch Statistics

$\Sigma$	$\mu$	$\sigma^2$	$\sigma$
9	2	1	1
5	1	1	1
7	2	4	2

Sum ( $\Sigma$ )  
Mean ( $\mu$ )  
Variance ( $\sigma^2$ )  
Std Dev ( $\sigma$ )

Normalize

$\mu$	2
-	1
	2

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

$\sigma$	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

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

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  
 $\approx$

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

Batch Statistics

$\Sigma$	$\mu$	$\sigma^2$	$\sigma$
9	2	1	1
5	1	1	1
7	2	4	2

Normalize

$\mu$	2
-	1
	2

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

$\sigma$	1
$\div$	1
	2

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

Sum ( $\Sigma$ )  
Mean ( $\mu$ )  
Variance ( $\sigma^2$ )  
Std Dev ( $\sigma$ )

Scale & Shift

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

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

Next Layer

Trainable  
Parameters

