Gradient Descent vs Stochastic Gradient Descent vs Mini Batch Gradient Descent

Let us have 5 training examples :
$\langle x_1^{(1)}, y^{(1)}\rangle$, $\langle x_1^{(2)}, y^{(2)}\rangle$, $\langle x_1^{(3)}, y^{(3)}\rangle$, $\langle x_1^{(4)}, y^{(4)}\rangle$, $\langle x_1^{(5)}, y^{(5)}\rangle$

predicted output be denoted by $o^{(i)}$

Let, $o^{(i)} = w_0 + w_1x_1^{(i)}$

$o^{(1)} = w_0 + w_1x_1^{(1)}$
$o^{(2)} = w_0 + w_1x_1^{(2)}$
$o^{(3)} = w_0 + w_1x_1^{(3)}$
$o^{(4)} = w_0 + w_1x_1^{(4)}$
$o^{(5)} = w_0 + w_1x_1^{(5)}$

Here, i goes from 1 to 5

Weight Update,
$w_j = w_j + \Delta w_j$

Here, j goes from 0 to 1

where,
$\Delta w_j = -\eta\,(\partial E / \partial w_j)$

In case of Gradient Descent,

$$
\begin{aligned}
E &= (1/2)*[\,[y^{(1)} - o^{(1)}]^2 + [y^{(2)} - o^{(2)}]^2 + [y^{(3)} - o^{(3)}]^2 + [y^{(4)} - o^{(4)}]^2 + [y^{(5)} - o^{(5)}]^2\,] \\
&= (1/2)*[\\
&\qquad [y^{(1)} - (w_0 + w_1x_1^{(1)})]^2 + \\
&\qquad [y^{(2)} - (w_0 + w_1x_1^{(2)})]^2 + \\
&\qquad [y^{(3)} - (w_0 + w_1x_1^{(3)})]^2 + \\
&\qquad [y^{(4)} - (w_0 + w_1x_1^{(4)})]^2 + \\
&\qquad [y^{(5)} - (w_0 + w_1x_1^{(5)})]^2 \\
&\qquad] \\
&= (1/2)* \Sigma\, [y^{(i)} - o^{(i)}]^2
\end{aligned}
$$

Again, i goes from 1 to 5. Same value of w0 and w1 is being used for all the 5 examples.

We update weight only after we have seen all the examples once.
Where,

$$
\begin{aligned}
\partial E / \partial w_0 &= (1/2)*[\\
&\qquad 2*[y^{(1)} - (w_0 + w_1x_1^{(1)})]*(-1) + \\
&\qquad 2*[y^{(2)} - (w_0 + w_1x_1^{(2)})]*(-1) + \\
&\qquad 2*[y^{(3)} - (w_0 + w_1x_1^{(3)})]*(-1) + \\
&\qquad 2*[y^{(4)} - (w_0 + w_1x_1^{(4)})]*(-1) + \\
&\qquad 2*[y^{(5)} - (w_0 + w_1x_1^{(5)})]*(-1) \\
&\qquad] \\
&= (-1/2)* 2*\Sigma[\, y^{(i)} - (w_0 + w_1x_1^{(i)})\,] \\
&= (-1/2)* 2*\Sigma[\, y^{(i)} - o^{(i)}\,]
\end{aligned}
$$

Again, i goes from 1 to 5.

$$
\begin{aligned}
\partial E / \partial w_1 &= (1/2)*[\\
&\qquad 2*[y^{(1)} - (w_0 + w_1x_1^{(1)})]*(-x_1^{(1)}) +
\end{aligned}
$$

$2*[y^{(2)} - (w_0 + w_1x_1^{(2)})]*(-x_1^{(2)})$ +
$2*[y^{(3)} - (w_0 + w_1x_1^{(3)})]*(-x_1^{(3)})$ +
$2*[y^{(4)} - (w_0 + w_1x_1^{(4)})]*(-x_1^{(4)})$ +
$2*[y^{(5)} - (w_0 + w_1x_1^{(5)})]*(-x_1^{(5)})$
]
$= (-1/2)* 2*\Sigma[y^{(i)} - (w_0 + w_1x_1^{(i)})]*(x_1^{(i)})$
$= (-1/2)* 2*\Sigma[y^{(i)} - o^{(i)}]*(x_1^{(i)})$

which is equivalent to
$\partial E / \partial w_j = (-1/2)*2*\Sigma[y^{(i)} - (w_0 + w_jx_j^{(i)})]*(x_j^{(i)})$

Again, i goes from 1 to 5.

Therefore,
In Gradient Descent we update the weights only after seeing all the examples.
$w_j = w_j + \Delta w_j$
where
Δw_j when $j = 0$ is $\eta\Sigma[y^{(i)} - o^{(i)}]$
$j = 1$ is $\eta\Sigma[y^{(i)} - (w_0 + w_1x_1^{(i)})]*(x_1^{(i)})$

Again, i goes from 1 to 5 (i.e., over all the training examples).

In Stochastic Gradient Descent, we update the weight after each training example:
$w_j = w_j + \Delta w_j$
where
Δw_j when $j = 0$ is $\eta[y^{(i)} - o^{(i)}]$
$j = 1$ is $\eta[y^{(i)} - (w_0 + w_1x_1^{(i)})]*(x_1^{(i)})$

Considering Batch Size of 2 examples,
According to Mini-batch Gradient Descent,
We would take
Batch_one = $\langle x_1^{(1)}, y^{(1)}\rangle, \langle x_1^{(2)}, y^{(2)}\rangle$
and apply gradient descent to it
Then take
Batch_two = $\langle x_1^{(3)}, y^{(3)}\rangle, \langle x_1^{(4)}, y^{(4)}\rangle$
and apply gradient descent to it
Then take
Batch_three = $\langle x_1^{(5)}, y^{(5)}\rangle$
and apply gradient descent to it.