In stochastic Gradient descent. We iterate each epoch for each Row, for example

If I have

	X_1	X_2	Y
P_1	X_{11}	X_{12}	$\hat{Y}_1$
P_2	X_{21}	X_{22}	$\hat{Y}_2$

dataset,

If I have 10 epochs & I have 2 Rows in the dataset, It iterate like this...

epoch: 1

Calculate weights & biase for P_1 & update

again " " " " " P_2 & update

epoch: 2

again calculate weights & bias for P_1 & update

" " " " " P_2 & update

epoch: 3

:

epoch: 10

calculate weight & bias for P_1 & update

" " " " P_2 & update.

so, Formulas are modified like,

For intercept / bias,

according to Batch Gradient Descent

Formula : ① $intercept\_slope = \frac{-2}{n}(Y_i - \hat{Y})$

according to

Stochastic GD : ① $intercept\_slope = \frac{-2}{1}(y[index] - \hat{y})$

∵ (y[index] means actual answer of $index^{th}$ row)

∵ ($\hat{y}$ is the prediction of that row)

So, $\hat{y} = (X[index].coef\_) + intercept\_$

② Updating intercept_ value,

intercept_ = intercept_ − intercept_slope.

according to

Stochastic GD : ② same as Batch GD.

Advantages :- ① Achieve performance in min NO of epochs.

② Suitable for large dataset where large memory required for Batch gradient descent as it take very less / constant memory.