a) **Batch**: When considering all the data examples when calculating the gradient we will have a std of zero as we will only get one lowest loss value and thus there is no std.

Mini-batch - batch size 6

Lowest loss occurred at: 0.53, -0.11, 0.53, 0.21
With standard deviation of 0.26
Comments:
By decreasing the batch size to 6 we get 4 different lowest loss values and thus the standard deviation increases from 0.0 to 0.26 as the w values differ.

Stochastic - batch size 1

Lowest loss occurred at: 2.11, 4.00, 4.00, 0.84, -2.00, 4.00, -2.00, 4.00, -2.00, 4.00, 3.37, 2.11, 2.74, 4.00, 1.79, 2.11, 2.74, -2.00, 4.00, -2.00
With standard deviation of 2.36
Comments:
By decreasing the batch size to 1 (stochastic) we get 20 different lowest loss values and thus the standard deviation increases even more from 0.26 to 2.36 as the w values have a larger spread.

b) **Batch gradient descent -** Computation cost is very high.
Stochastic gradient descent - Computation time is more.
Mini-batch gradient descent - Computation time is less than SGD and computational cost is less than BGD. This is a good option if both computational cost and time are equally important.