1. mini\_batch gradient descent

if n\_samples is very large, trainning time will be too long.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x1,x2,x3……xmini\_batch | xmini\_batch…x2\*mini\_batch | … | … | … |
| X1 | X2 | … | … | … |

Recommended value: 64,128,512

1. Exponentially weighted averages

Bias correction:

1. Momentum
2. RMSprop
3. Adam
4. Learning rate decay

….so on