

# Stochastic Gradient Descent

# Mini-Batch Gradient Descent

# Batch Gradient Descent

# Samples

1

batch  
size  
( $b$ )

all  
samples  
( $m$ )

Main  
Advantages

Faster than  
batch  
gradient  
descen

Random  
selection will  
help to avoid  
using redundant  
samples

Adds lots of noise to  
learning  
process which  
improves  
generalization  
error

Less  
computational  
burden than  
batch gradient  
descent

Faster than  
batch  
gradient  
descen

Random  
selection will  
help to avoid  
using redundant  
samples

Adds noise to  
learning process  
which improves  
generalization  
error

Less  
computational  
burden than  
batch gradient  
descent

Fixed  
learning  
rate

No  
learning  
rate decay  
needed

Guaranteed  
convergence to  
global (f is  
convex) or local (f  
is not convex)  
minimum

Straight  
trajectory  
towards the  
minimum

The more  
samples the  
lower the  
standard  
error

Main  
Disadvantages

Increased  
run time due  
to huge  
amount of  
noise

No vectorization  
can be used  
which makes  
the algorithm  
very slow

Variance  
becomes  
large

Only wanders  
around  
minimum region  
but never  
converges

Learning  
steps have  
oscillation  
due to noise

Adding Learning  
rate decay is  
needed as we  
approach the  
minimum

Slow  
especially  
for large  
datasets

Can get  
stuck in  
local  
minimum

Usage of  
redundant  
samples  
during each  
learning step

Visualization

