**Solving loading datasets into memory issue
by minibatch**

Load data through using a minibatch
approach
(utilizing small portions of data)

How much data should we get?
Set a batch amount by user
Set iteration amount by user

Batch Amount: How much
data do we get from each
file?

Iteration: How many times do we
get set number of batches from
the same file?

Starting Batch: Fetch the
required amount of data
through threading
for efficiency and allow for
"background work"

N amount of threads:
Each N thread should
process its own set file

After all N threads are done:
We add info about our
batches,
From which dataset did it
come from?

Through fetching:
"Remembering" where we
left off for next batch
Requiring to save the
progress of what line in the
file we left off after each
batch