Project: A Further Investigation on the Running Time

Last updated: April 20, 2020
Investigating why for the previous project some MATLAB operations are inefficient
Consider the following MATLAB code to run the same operation on CPU and GPU

```matlab
function test

m = 10000;

for gpu_use = 0:1

    A = gpu(randn(m,m), gpu_use);
    B = gpu(randn(m,m), gpu_use);
    a = rem(randperm(10*m)', m)+1;

end
```

f1 = @(() A*B);
f2 = @(() A(a,:));

if gpu_use == 1,
    gputimeit(f1)
    gputimeit(f2)
else
    timeit(f1)
    timeit(f2)
end
function M = gpu(M, gpu_use)
    if gpu_use == 1
        M = gpuArray(M);
    end

Results:
>> test
ans =
    5.6717
ans =
    2.9617
ans =
    4.2868
ans =
    0.3201
We conduct this experiment because both operations are used in our stochastic gradient implementation.

For example, in `padding_and_phiZ.m` we have:

\[ \text{phiZ} = \text{phiZ}(\text{net.idx_phiZ\{m\}}, :) ; \]

for generating

\[ \phi(Z^m_i), \forall i \]

This code can be run on MATLAB only. Neither `timeit` nor `gputimeit` is supported on Octave.
Complexity of the two operations

$10^{12}$ and $10 \times 10^8$

We do not expect a 1000-fold time difference because we already know that matrix products by optimized BLAS gets better data locality.

But the difference between CPU and GPU is surprising.

From CPU to GPU, the matrix product is shortened by less than half.
But for matrix expansion GPU is much faster
Let’s see if we can improve the matrix expansion on CPU as probably CPU is not fully utilized
Let’s write a C code on CPU to do the matrix expansion
Check if its running time is similar to MATLAB. Not that you want to exclude the time for data preparation
Try possible optimization. For example, use openmp or pthread to take the advantage of multi-core CPUs
See how much you can do better (or worse) than MATLAB

FYI, for matrix products, we have checked non-squared matrices. The speedup from CPU to GPU may be slightly better (but only slightly better)
Students with the following IDs (last three digits): ??
please do a ??-minute presentation (??-minute the contents and ??-minute Q&A)