homework1

January 29, 2019

1 Homework 1 - Berkeley STAT 157

Handout 1/22/2017, due 1/29/2017 by 4pm in Git by committing to your repository. Please ensure that you add the TA Git account to your repository.

1. Write all code in the notebook.
2. Write all text in the notebook. You can use MathJax to insert math or generic Markdown to insert figures (it's unlikely you'll need the latter).
3. **Execute** the notebook and **save** the results.
4. To be safe, print the notebook as PDF and add it to the repository, too. Your repository should contain two files: `homework1.ipynb` and `homework1.pdf`.

The TA will return the corrected and annotated homework back to you via Git (please give `rythei` access to your repository).

```
In [2]: import mxnet as mx
        from mxnet import ndarray as nd
        import time
```

1.1 1. Speedtest for vectorization

Your goal is to measure the speed of linear algebra operations for different levels of vectorization. You need to use `wait_to_read()` on the output to ensure that the result is computed completely, since NDArray uses asynchronous computation. Please see http://beta.mxnet.io/api/ndarray/_autogen/mxnet.ndarray.NDArray.wait_to_read.html for details.

1. Construct two matrices A and B with Gaussian random entries of size 4096×4096.
2. Compute $C = AB$ using matrix-matrix operations and report the time.
3. Compute $C = AB$, treating A as a matrix but computing the result for each column of B one at a time. Report the time.
4. Compute $C = AB$, treating A and B as collections of vectors. Report the time.
5. Bonus question - what changes if you execute this on a GPU?
    - It will be faster

```
In [ ]: # CPU, the time is ~5s, ~20s, ~5000s separately
        A=nd.random.normal(shape=[4096,4096])
        B=nd.random.normal(shape=[4096,4096])
```

```
        t1=time.time()
        C=nd.dot(A,B)
        C.wait_to_read()
        print('Direct dot-product time is: ', time.time()-t1, 's')

        t2=time.time()
        for i in range(4096):
            C[:,1]=nd.dot(A,B[:,i])
        C.wait_to_read()
        print('Treat B as column vectors, time is: ', time.time()-t2, 's')

        t3=time.time()
        for i in range(4096):
            for j in range(4096):
                C[i,j]=nd.dot(A[i,:],B[:,j])
        C.wait_to_read()
        print('Treat A, B as collection of vectors, time is ', time.time()-t3, 's')
```

```
Direct dot-product time is:  3.5399234294891357 s
Treat B as column vectors, time is:  15.46968412399292 s
```

```
In [ ]: # GPU, the time is ~.5s, ~10s,
        with mx.Context(mx.gpu()):
            A=nd.random.normal(shape=[4096,4096])
            B=nd.random.normal(shape=[4096,4096])
            A
            B
            t1=time.time()
            C=nd.dot(A,B)
            C.wait_to_read()
            print('[GPU] Direct dot-product time is: ', time.time()-t1, 's')

            t2=time.time()
            for i in range(4096):
                C[:,1]=nd.dot(A,B[:,i])
            C.wait_to_read()
            print('[GPU] Treat B as column vectors, time is: ', time.time()-t2, 's')

            t3=time.time()
            for i in range(4096):
                for j in range(4096):
                    C[i,j]=nd.dot(A[i,:],B[:,j])
            C.wait_to_read()
            print('[GPU] Treat A, B as collection of vectors, time is ', time.time()-t3, 's')
```

```
[GPU] Direct dot-product time is:  0.266162633895874 s
[GPU] Treat B as column vectors, time is:  9.777193069458008 s
```

1.2 2. Semidefinite Matrices

Assume that $A \in \mathbb{R}^{m \times n}$ is an arbitrary matrix and that $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix with nonnegative entries.

1. Prove that $B = ADA^\top$ is a positive semidefinite matrix.
2. When would it be useful to work with B and when is it better to use A and D?

1. We can prove B to be positive semidefinite by showing that there exists some matrix Q, s.t. $B = Q^\top Q$. Since $B = ADA^\top$ and D is a diagonal matrix with nonnegative entries, we can set $Q = \sqrt{D}A^\top$, thus $Q^\top = A\sqrt{D}$. Hence $Q^\top Q = A\sqrt{D}\sqrt{D}A\top = ADA^\top = B$. By definition, we have proved that B is positive semidefinite.
2. When it comes to multiply B with another matrix or vector, it is better to use B. But if we want to calculate B to its power (like $B \cdot B \cdot B \cdots$), then we can do faster with A and D if m is much larger than n.

1.3 3. MXNet on GPUs

1. Install GPU drivers (if needed)
2. Install MXNet on a GPU instance
3. Display `!nvidia-smi`
4. Create a 2×2 matrix on the GPU and print it. See http://d2l.ai/chapter_deep-learning-computation/use-gpu.html for details.

1.3.1 Note

I managed to install cuda on my Surfacebook and run the code below on it.

```
In [1]: import mxnet as mx
        from mxnet import ndarray as nd
        !nvidia-smi
        x = nd.ones((2, 2), ctx=mx.gpu())
        x
```

```
Mon Jan 28 22:09:32 2019
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 417.71       Driver Version: 417.71       CUDA Version: 10.0     |
|-------------------------------+----------------------+----------------------+
| GPU  Name            TCC/WDDM | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  GeForce GPU        WDDM  | 00000000:01:00.0 Off |                  N/A |
| N/A   37C    P8     1W /  N/A |     36MiB /  1024MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
```

```
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

```
Out[1]:
        [[1. 1.]
         [1. 1.]]
        <NDArray 2x2 @gpu(0)>
```

1.4 4. NDArray and NumPy

Your goal is to measure the speed penalty between MXNet Gluon and Python when converting data between both. We are going to do this as follows:

1. Create two Gaussian random matrices A, B of size 4096×4096 in NDArray.
2. Compute a vector $\mathbf{c} \in \mathbb{R}^{4096}$ where $c_i = \|AB_{i.}\|^2$ where $\mathbf{c}$ is a **NumPy** vector.

To see the difference in speed due to Python perform the following two experiments and measure the time:

1. Compute $\|AB_{i.}\|^2$ one at a time and assign its outcome to $\mathbf{c}_i$ directly.
2. Use an intermediate storage vector $\mathbf{d}$ in NDArray for assignments and copy to NumPy at the end.

```
In [33]: import numpy as np
         c=np.ndarray([4096,1])
         A=nd.random.normal(shape=[4096,4096])
         B=nd.random.normal(shape=[4096,4096])
         t1=time.time()
         for i in range(4096):
             c[i]=nd.norm(nd.dot(A, B[:, i])).asnumpy()
         print('Assign outcome to c_i directly: ', time.time()-t1)

         d=nd.zeros(shape=[4096,1])
         t2=time.time()
         for i in range(4096):
             d[i]=nd.norm(nd.dot(A, B[:, i]))
         c=d.asnumpy()
         print('Assign to intermediate storage d: ', time.time()-t2)
```

```
Assign outcome to c_i directly:  16.207096815109253
Assign to intermediate storage d:  12.968188047409058
```

1.5 5. Memory efficient computation

We want to compute $C \leftarrow A \cdot B + C$, where A, B and C are all matrices. Implement this in the most memory efficient manner. Pay attention to the following two things:

1. Do not allocate new memory for the new value of C.

2. Do not allocate new memory for intermediate results if possible.

```
In [9]: # Random initialize some matrices A, B and C
        A=nd.random.normal(shape=[4096,4096])
        B=nd.random.normal(shape=[4096,4096])
        C=nd.random.normal(shape=[4096,4096])
        # print(nd.dot(A, B)-nd.elemwise_mul(A, B))
        # print(id(C))
        C[:]+=nd.dot(A, B)
        # print(id(C))
```

1.6 6. Broadcast Operations

In order to perform polynomial fitting we want to compute a design matrix A with

$$A_{ij} = x_i^j$$

Our goal is to implement this **without a single for loop** entirely using vectorization and broadcast. Here $1 \leq j \leq 20$ and $x = \{-10, -9.9, \ldots 10\}$. Implement code that generates such a matrix.

```
In [9]: # Suppose we implement the calculation as a function
        # The input parameters are vector x and power j, and the return value is the design ma
        def design_mat(j):
            e=nd.arange(1, 21).reshape(1, 20)
            x=nd.arange(-10.0, 10.1, step=0.1).reshape(201, 1)
            A=x**e
            return A.reshape(21, j)
        design_mat(4)
```

```
Out[9]:
        [[-1.0000000e+01  1.0000000e+02 -1.0000000e+03  1.0000000e+04]
         [-1.0000000e+05  1.0000000e+06 -1.0000000e+07  1.0000000e+08]
         [-1.0000000e+09  1.0000000e+10 -9.9999998e+10  1.0000000e+12]
         [-9.9999998e+12  1.0000000e+14 -9.9999999e+14  1.0000000e+16]
         [-9.9999998e+16  9.9999998e+17 -1.0000000e+19  1.0000000e+20]
         [-9.8999996e+00  9.8009995e+01 -9.7029889e+02  9.6059590e+03]
         [-9.5098984e+04  9.4147994e+05 -9.3206510e+06  9.2274440e+07]
         [-9.1351693e+08  9.0438175e+09 -8.9533784e+10  8.8638449e+11]
         [-8.7752059e+12  8.6874538e+13 -8.6005787e+14  8.5145724e+15]
         [-8.4294265e+16  8.3451318e+17 -8.2616803e+18  8.1790629e+19]
         [-9.8000002e+00  9.6040001e+01 -9.4119208e+02  9.2236826e+03]
         [-9.0392086e+04  8.8584250e+05 -8.6812570e+06  8.5076312e+07]
         [-8.3374790e+08  8.1707295e+09 -8.0073155e+10  7.8471692e+11]
         [-7.6902260e+12  7.5364211e+13 -7.3856929e+14  7.2379793e+15]
         [-7.0932201e+16  6.9513558e+17 -6.8123289e+18  6.6760824e+19]
         [-9.6999998e+00  9.4089996e+01 -9.1267297e+02  8.8529277e+03]
         [-8.5873391e+04  8.3297194e+05 -8.0798275e+06  7.8374320e+07]
         [-7.6023091e+08  7.3742397e+09 -7.1530127e+10  6.9384221e+11]
         [-6.7302693e+12  6.5283608e+13 -6.3325098e+14  6.1425345e+15]
```

```
 [-5.9582582e+16  5.7795107e+17 -5.6061250e+18  5.4379413e+19]
 [-9.6000004e+00  9.2160004e+01 -8.8473608e+02  8.4934668e+03]]
<NDArray 21x4 @cpu(0)>
```