

6/10

# 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 [80]: from mxnet import ndarray as nd
```

## 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](http://beta.mxnet.io/api/ndarray/_autogen/mxnet.ndarray.NDArray.wait_to_read.html)

([http://beta.mxnet.io/api/ndarray/\\_autogen/mxnet.ndarray.NDArray.wait\\_to\\_read.html](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  $4096 \times 4096$ .
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?

```
In [81]: a = nd.random.normal(0, 1, shape=(4096, 4096))
         b = nd.random.normal(0, 1, shape=(4096, 4096))
```

```
In [82]: import time
         tic = time.time()
         c = nd.dot(a, b)
         print(time.time() - tic)
```

0.01248311996459961

```
In [83]: import numpy as np
c.wait_to_read()
total = 0
for i in np.arange(4096):
    c[i] = (nd.dot(a, b[i]))
print(time.time() - tic)
```

3.7118330001831055

In [ ]:

```
In [84]: c.wait_to_read()
total = 0
for i in np.arange(4096):
    c[i] = (nd.dot(a[i], b[i]))
print(time.time() - tic)
```

26.64480495452881

- | this only gets diagonal,  
need  $c_{ij} = \text{dot}(A[i,:], B[:,j])$ .

## 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  $B = ADA^T$  is a positive semidefinite matrix.
2. When would it be useful to work with  $B$  and when is it better to use  $A$  and  $D$ ?

- | not sure what you mean, need to show  $x^T B x \geq 0$   
 $\forall x \in \mathbb{R}^m$

1. Because matrix multiplication is associative, we can move the  $A$ ,  $D$ , and  $A$ 's transpose around as we please and it will be the same as the layout in the problem. We know that  $A$  and its transpose, when multiplied together, create a positive semidefinite matrix. A matrix multiplied by a diagonal matrix preserves symmetry because a diagonal matrix is symmetric and because  $A$  multiplied by its transpose is symmetric. A symmetric matrix multiplied by another symmetric matrix is a symmetric matrix.
2. The shape of  $B$  is a square symmetric positive semidefinite matrix because we multiply  $(m \times n)(n \times n)(n \times m)$  which results in an  $(m \times m)$  matrix.

## 3. MXNet on GPUs

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

OK

In [ ]:

## 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  $4096 \times 4096$  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 [85]: a = nd.random.normal(0, 1, shape=(4096, 4096))
         b = nd.random.normal(0, 1, shape=(4096, 4096))
```

```
In [87]: c = np.empty(shape= 4096, dtype=object)
         for i in np.arange(4096):
             c[i] = (nd.norm(nd.dot(a, b[i])))**2
         print(time.time() - tic)
```

36.830349922180176

```
In [89]: c = np.empty(shape= 4096, dtype=object)
         for i in np.arange(4096):
             c[i] = (nd.norm(nd.dot(a, b[i])))**2
         print(time.time() - tic)
```

47.646074056625366

```
In [90]: c = np.empty(shape= 4096, dtype=object)
         d = nd.empty(4096)
         for i in np.arange(4096):
             d[i] = (nd.norm(nd.dot(a, b[i])))**2
             c[i] = d[i]
         print(time.time() - tic)
```

50.9998459815979

should convert to  
numpy at end:  
 $c = d.asnumpy()$

## 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.

— 1

In [ ]:

## 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  $x = \{-10, -9.9, \dots, 10\}$ . Implement code that generates such a matrix.

In [ ]: