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]:
```
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) 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?

In [27]:
```
import time
##Q1
A = nd.random.normal(0, 1, (4096, 4096))
B = nd.random.normal(0, 1, (4096, 4096))

##Q2
start2 = time.time()
C2 = nd.dot(A, B)
C2.wait_to_read()
print("time for Q2 is", time.time() - start2)

##Q3

C3 = nd.zeros((4096, 4096))
start3 = time.time()
for i in range(4096):
    C3[:, i] = nd.dot(A, B[:,i])
C3.wait_to_read()
print("time for Q3 is", time.time() - start3)
```

```
time for Q2 is 2.7555549144744873
time for Q3 is 20.58580207824707
```

In [28]:
```
##Q4
C4 = nd.zeros((4096, 4096))
start4 = time.time()
B = B.T
for i in range(4096):
    for k in range(4096):
        C4[i, k] = nd.dot(A[i], B[k]).asscalar()
C4.wait_to_read()
print("time for Q4 is", time.time() - start4)
```

```
time for Q4 is 2494.8225951194763
```

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. To prove positive semidefinite, want to show that
$$\forall x \in \mathbb{R}^m, x^T Bx \geq 0$$
Since $B = ADA^\top$,
$$x^T Bx = (x^T A)D(A^T x)$$
And we know that $(x^T A)$ is a $1 * n$ matrix, and $(A^T x) = (x^T A)^T$ is a $n * 1$ vector. Let $x^T A = v = [v_1, \ldots, v_n]$, v has n arbitrary entries since both A and x are arbitrary. Now let $d_{11}, d22, \ldots, dnn$ denote the non-negative entries in D, then
$$vDv^T = \Sigma_{i=1}^n v_i^2 * d_{ii}$$
for each i, $v_i^2 \geq 0, d_{ii} \geq 0$, thus $\Sigma_{i=1}^n v_i^2 * d_{ii} \geq 0$ Thus by definition, $B = ADA^\top$ is a positive semidefinite matrix.
2. I suppose that when no additional operation is needed, B is better because you can just compute it once and save running time; but if you want A to contain some information or want the option to adjust A or D during operations, it's better to use both A and D.

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 (http://d2l.ai/chapter_deep-learning-computation/use-gpu.html) for details.

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 $c_i = ||AB_{i\cdot}||^2$ 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 $||AB_{i\cdot}||^2$ 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 [14]:
```
import numpy as np
import time

A = nd.random.normal(0, 1, (4096, 4096))
B = nd.random.normal(0, 1, (4096, 4096))
##Method1
B = B.T
c = np.ones(4096)
start1 = time.time()
for i in range(4096):
    vector = nd.dot(A, B[i])
    vector.wait_to_read()
    toAdd = vector.sum().asscalar()
    c[i] = toAdd**2

print("time for Method#1 is", time.time() - start1)
```

```
time for Method#1 is 23.370030164718628
```

In [16]:
```
##Method2
start2 = time.time()
d = nd.ones((1, 4096))
for i in range(4096):
    vector = nd.dot(A, B[i])
    toAdd = vector.sum().asscalar()
    d[0][i] = toAdd**2
c = d.asnumpy()
d.wait_to_read()
print("time for Method#2 is", time.time() - start2)

## Observation: method two is obviously faster
```

```
time for Method#2 is 19.857491970062256
```

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 [19]:
```
C = nd.random.normal(0, 1, (500, 500))
print("ID of C before is", id(C))
A = nd.random.normal(0, 1, (500, 500))
B = nd.random.normal(0, 1, (500, 500))
for i in range(500):
    C[i] = nd.dot(A, B)[i] + C[i]
print("ID of C after is", id(C))
```

```
ID of C before is 4570709688
ID of C after is 4570709688
```

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 [62]:
```
import numpy as np
x = np.arange(-10, 10, 0.1)
x.size
x = nd.array(x)
x.reshape((1, 200))
result = nd.zeros((200, 20))
result[:, 0] = x
result[:, 1] = result[:, 0]*x
result[:, 2] = result[:,1]*x
```

```
[[1. 1. 1. 1. 1.]
 [2. 2. 2. 2. 2.]
 [3. 3. 3. 3. 3.]
 [4. 4. 4. 4. 4.]
 [5. 5. 5. 5. 5.]]
<NDArray 5x5 @cpu(0)>

[[ 1.]
 [ 4.]
 [ 9.]
 [16.]
 [25.]]
<NDArray 5x1 @cpu(0)>

[[  1.   1.]
 [ 16.  16.]
 [ 81.  81.]
 [256. 256.]
 [625. 625.]]
<NDArray 5x2 @cpu(0)>
```