Data Intensive systems for Machine Learning

Coding Assignment 3
Dense Matrix Multiplication

Nihaarika Agarwal
nagarw48@asu.edu (1233370023)

Hardware Configuration

Processor	M3 Chip
Architecture	ARM based (Apple-Silicon)
Core type/ number of cores	CPU Cores: 8 Cores GPU Cores: 8 Cores
Core frequency	4.05 GHz
Memory	8 GB

Software Configuration

Operating system	MacOS Sonoma (MacOS 14)
Scala	2.12.18
Spark core	3.4.0
Spark mllib	3.5.5
Netlib	1.1.2
Breeze	0.13
Sbt	1.10.10

Performance Comparison

1000 Iterations

	Handcoded scala	Spark	Breeze
Iterations	1000	1000	1000
Average time-taken for one iteration (seconds)	0.978008	0.12642799999999998	0.234707

100 Iterations

	Handcoded scala	Spark	Breeze
Iterations	100	100	100
Average time-taken for one iteration (seconds)	0.10231100000000001	0.12547	0.22743

Things I learned

Working with large dense vectors for numerical operations is computationally expensive. Optimized libraries like Breeze and Spark significantly improve performance by providing efficient implementations for linear algebra operations.
Additionally, I gained hands-on experience with Scala and the sbt build tool. I also experimented with compiling code using sbt to generate a fat jar, as spark-submit requires one. While creating thin jars in sbt is straightforward, generating fat jars can be challenging due to the need to resolve merge strategies for conflicting packages during assembly and handle numerous version dependencies. In comparison, Maven and Gradle offer a more streamlined approach to building packages.

```
Time for 1000 runs: 234.707 s

Average time per run: 0.234707 s

[success] Total time: 236 s (0:03:56.0), completed Mar 7, 2025, 11:41:59 AM
Nihaarikas-Laptop:Breeze nihaarikaagarwal$
```

Fig: Run on Breeze for 1000 runs.

```
Nihaarikas-Laptop:Handcoded nihaarikaagarwal$ scala Handcoded.scala

Time for 1000 runs: 978.008 s

Average time per run: 0.978008 s
```

Fig: Run for handcoded for 1000 runs.

```
[info] Time for 1000 runs: 126.428 s
[info] Average time per run: 0.12642799999999998 s
[success] Total time: 130 s (0:02:10.0), completed Mar 7, 2025, 5:51:57 PM
```

Fig: Run on Spark for 1000 runs.

```
Time for 100 runs: 22.743 s

Average time per run: 0.22743 s

[success] Total time: 25 s, completed Mar 7, 2025, 11:49:07 AM
```

Fig: Run on Breeze for 100 runs.

```
^CNihaarikas-Laptop:Handcoded nihaarikaagarwal$ scala Handcoded.scala

Time for 100 runs: 102.311 s

Average time per run: 0.10231100000000001 s
```

Fig: Run of handcoded for 100 runs.

```
[info] Time for 100 runs: 12.547 s
[info] Average time per run: 0.12547 s
[success] Total time: 16 s, completed Mar 7, 2025, 5:49:07 PM
```

Fig: Run on Spark for 100 runs.