Comp528 report

1. Explain each OMP directive used

```
#pragma omp parallel for schedule(dynamic)
for (int chunk_start = 0; chunk_start < test_rows; chunk_start += chunk_size) {
    int current_chunk_size = (chunk_start + chunk_size > test_rows) ? (test_rows - chunk_start) : chunk_size;
    processChunk(train_data, test_data, train_rows, test_rows, train_cols, test_cols, k, chunk_start, current_chunk_size);
}
```

This is the only OMP directive used in my solution, but quite a bit of explanation can go into why it is as effective as it is. Firstly all test data points are perfectly data parallel, meaning that the processing from start to finish can be done entirely independently of any other test point. This includes the distance calculations, the finding the k nearest, and the writing of the solution into the array storing all of the test points. This provides near perfect parallelism to all of the processing of each point. The points are handled in chunks in order to assist with cache memory allocation and also so that no memory limitations exist when handling the points in sequential solutions.

2. For each thread provide the runtime, the speedup, and how they are calculated.

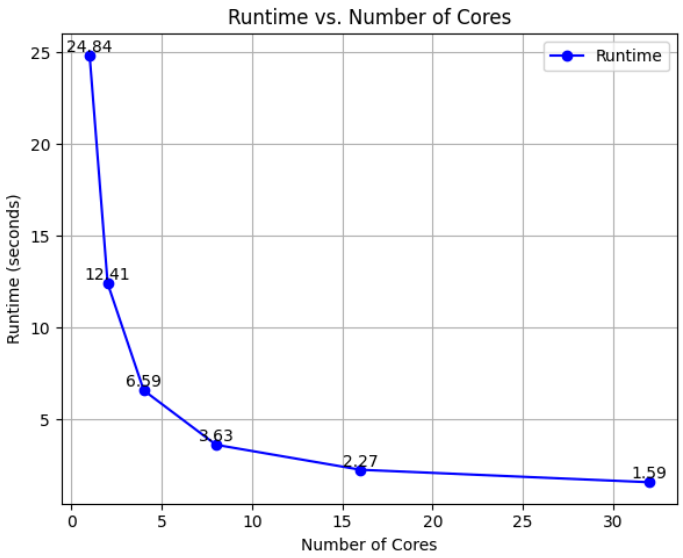
The calculation of the run time for my code has been done using the fairest method possible, I am taking the average runtime from 5 runs of the code at each number of threads for my main benchmark. This was done using gcc -O3 as I found that to be the fastest for my code. This was done using a simple bash script:

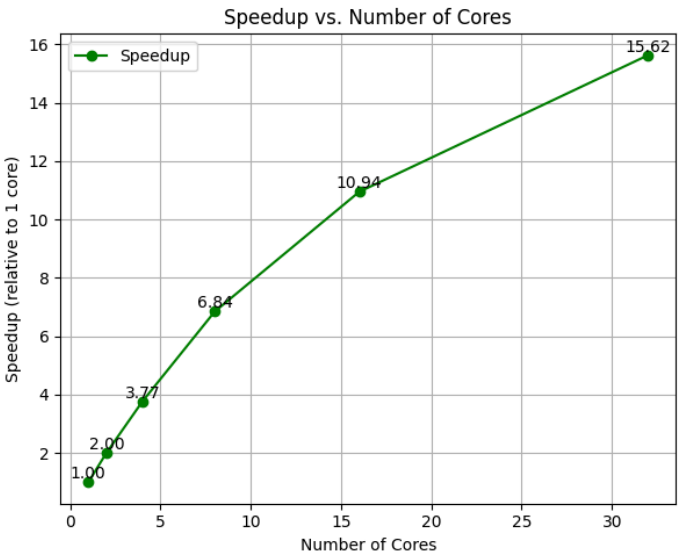
```
#!/bin/bash -l
#Written by Dr Maryam Abo Tabik
# Specify the current working directory as the location for executables/files
# This is the default setting.
#SBATCH -D ./

# Export the current environment to the compute node
# This is the default setting.
#SBATCH --export=ALL

# Specific course queue, exclusive use (for timings), max 1 min wallclock time
#SBATCH -p lowpriority
#SBATCH -t 5:00
#SBATCH -N 1
#SBATCH -n 40

# load modules
module load compilers/intel/2019u5

# just 1 thread to run on
export OMP_NUM_THREADS=1
# The program to run (replace with your actual executable)

# Number of times to run the program
NUM_RUNS=5

# Variable to accumulate total time
total_time=0
gcc -fopenmp -march=native -O3 first-opt.c -std=c99 -o knn.out
# Loop to run the program NUM_RUNS times
for ((i = 1; i <= NUM_RUNS; i++)); do
    # Measure the time taken to run the program
    start_time=$(date +%s.%N) # Get start time in seconds
    ./knn.out "data/asteroids_train.csv" "data/asteroids_test.csv" "out.csv" 3
    end_time=$(date +%s.%N)   # Get end time in seconds

    # Calculate elapsed time
    elapsed_time=$(echo "$end_time - $start_time" | bc)

    # Add elapsed time to total
    total_time=$(echo "$total_time + $elapsed_time" | bc)

    echo "Run $i: $elapsed_time seconds"
done

# Calculate average time
average_time=$(echo "scale=2; $total_time / $NUM_RUNS" | bc)
echo "Average time over $NUM_RUNS runs: $average_time seconds"
```

Here is the table of results (the output for each run is in Appendix A):

Num of threads	Average runtime (s)	Fastest runtime (s)
1	24.84	24.595
2	12.41	12.308
4	6.59	6.497
8	3.63	3.584
16	2.27	2.223
32	1.59	1.552

3. Here is a plot of the runtime and parallel efficiency:





4. An explanation as to why I did or didn't achieve linear speedup

My program starts off with a near linear speedup, this is as all of the processing for each point is done in a perfectly parallel setup. However once the number of threads starts to increase to higher numbers the speedup becomes sublinear, this is as at high thread numbers more factors have a larger impact on the runtime of the program. At runtimes as low as 1.59s factors such as data access patterns and the limiting sequential portion of the code have a larger impact.

Appendix A

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 1.564941222 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 1.618428974 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 1.646995181 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 1.552306007 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 1.570837101 seconds
Average time over 5 runs: 1.59 seconds
```

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 2.361577190 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 2.223955251 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 2.263147210 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 2.266855056 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 2.263679922 seconds
Average time over 5 runs: 2.27 seconds
~
~
~
~
~
~
~
~
~
~
"slurm-49762005.out" 51L, 1459C
```

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 3.669054286 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 3.584552417 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 3.615199683 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 3.645355244 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 3.670120704 seconds
Average time over 5 runs: 3.63 seconds
~
~
~
~
~
~
~
~
~
~
"slurm-49762009.out" 51L, 1459C
```

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 6.799169212 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 6.512310345 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 6.567507038 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 6.497141405 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 6.580472867 seconds
Average time over 5 runs: 6.59 seconds
~
~
~
~
~
~
~
~
~
~
"slurm-49762010.out" 51L, 1459C
```

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 12.647400620 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 12.407273042 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 12.349143483 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 12.308397885 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 12.362946847 seconds
Average time over 5 runs: 12.41 seconds
~
~
~
~
~
~
~
~
~
"slurm-49762011.out" 51L, 1465C
```

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 24.746190672 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 24.780791287 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 25.143547748 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 24.595452866 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 24.980869757 seconds
Average time over 5 runs: 24.84 seconds
~
~
~
~
~
~
~
~
~
"slurm-49762013.out" 51L, 1465C
```

And if you're curious about the runtime on 40 cores

```
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 1: 1.412443245 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 2: 1.410929912 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 3: 1.483165107 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 4: 1.394498340 seconds
Reading num of points: 153364
Reading num of features: 9
Reading data points: 153365
Reading num of points: 38342
Reading num of features: 9
Reading data points: 38343
Writing output data to file: out.csv:
number of data points: 38342
number of features: 9
Run 5: 1.470590752 seconds
Average time over 5 runs: 1.43 seconds
~
~
~
~
~
~
~
~
~
~
"slurm-49762118.out" 51L, 1459C
```