**UCI**[**Optical Recognition of Handwritten Digits**](http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of_Handwritten+Digits)

For CST3170 Artificial Intelligence module

Geraldas Kastauna (M00651656)

Middlesex University

London, United Kingdom

GK468@live.mdx.ac.uk

**Description:**

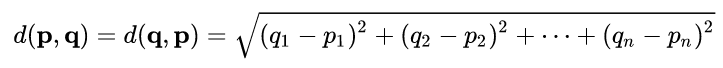
For this coursework students are required to build a machine learning system to categorise one of the UCI digit tasks. Student should develop the system on your own from scratch in JAVA programming language, run a two-fold test, and report the results.

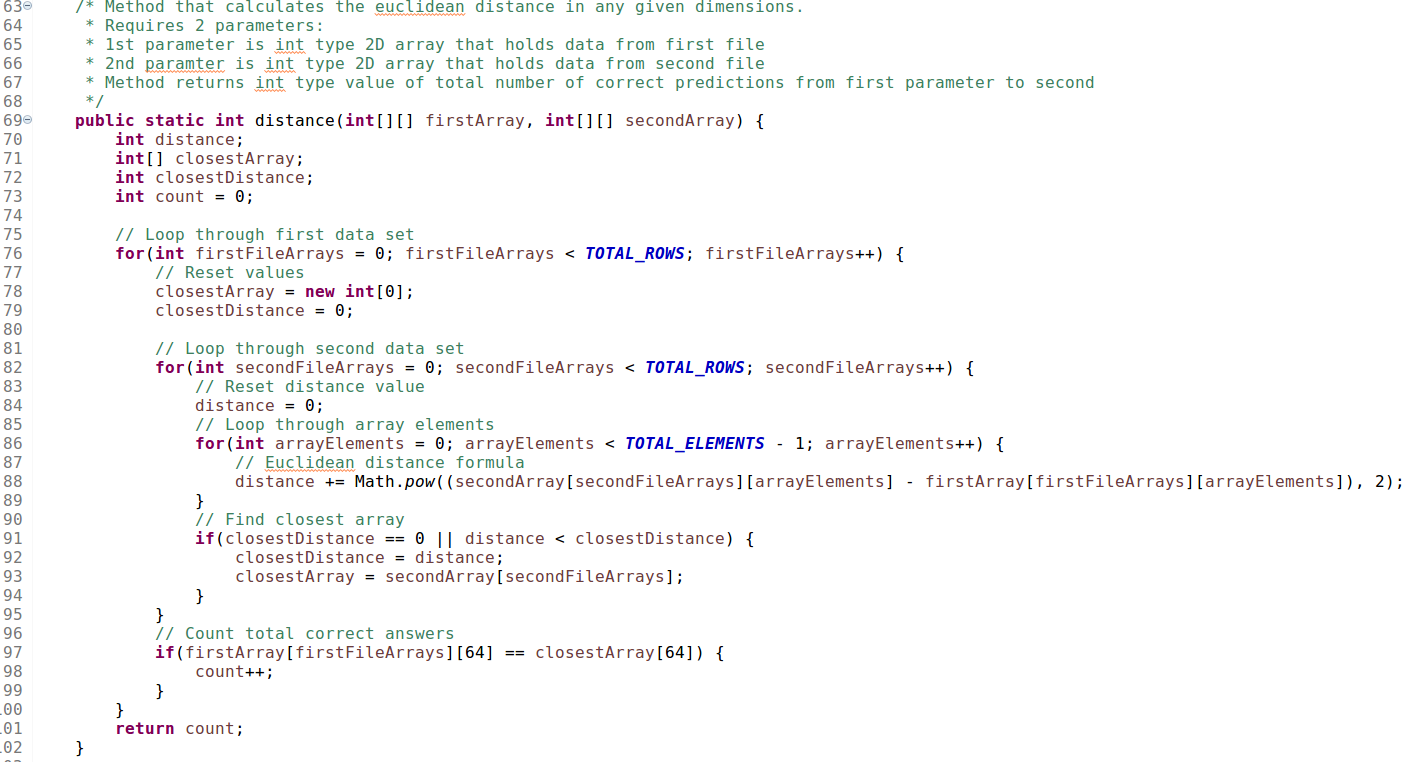
The data is from the University of California at Irvine's Machine Learning Repository. It's the [Optical Recognition of Handwritten Digits Data Set](http://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of_Handwritten+Digits). This gives two data sets, training set and a test set. Module leader converted them to two data sets - [data set 1](http://www.cwa.mdx.ac.uk/cst3170/coursework/cw2DataSet1.csv), and [data set 2](http://www.cwa.mdx.ac.uk/cst3170/coursework/cw2DataSet2.csv)that should be used by students system.

**More info: http://www.cwa.mdx.ac.uk/cst3170/coursework/CourseWork2.html**

**Algorithm:**

In this task I used euclidean distance because it was the easiest one to implement as it does not do any training. It helped me get familiar with the data that I was working with and this was the first suggestion from the module leader. Algorithm is pretty straigth forward – it calculates the euclidean distance in 64 dimensions using this formula:

where *d* is distance *p* and *q* are points in *n* space.

Here is the code snippet from my system:

*ALL THIS PROCESS IS WELL COMMENTED INSIDE THE SYSTEM*

**Results:**

