COMP4601 Project

Presenter:
Zhixin TANG (z5142033)
Shuyi WANG (z5190769)

Project objectives

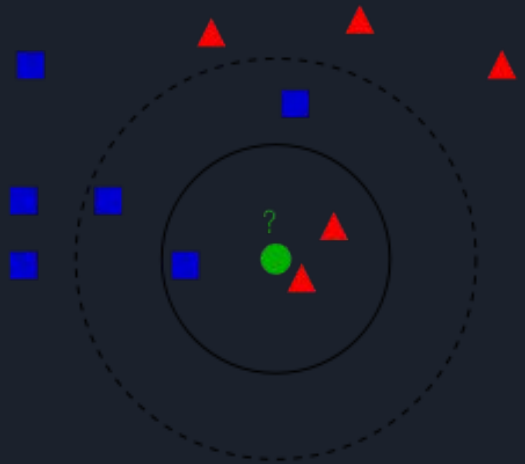
Accelerating a k-nearest neighbours (KNN) algorithm in Machine Learning by using high-level synthesis by given a dataset.

- When the size and dimension of dataset increase, KNN algorithm becomes slower. Therefore, it has the potential to be accelerated by FPGA hardware.

KNN Algorithm

KNN is used for classification and regression. The input consists k closest training examples in dataset.

Example:



KNN Algorithm

In out project, we assume each vector in the dataset only contains four features including vector, label, distance and result.

The KNN Algorithm used in our project can be divided into 3 steps:

1: Distance Calculation

2: K-Nearest Neighbors Finder

3: Class Determination

```
Function KNN:
    Initialization
    For each element in the dataset
        Compute the distance (euclidean metric)
    Sort the elements from the traning dataset by distance in increasing order
    For each element in the first k elements
        Count the times each label observed
    Find the label of highest voting
```

Hardware Pseudo Code and possible acceleration

The C++ code contains following features:

- Loop to calculate the distance for each element in the dataset <- Pipelining
- Sort the elements in the dataset by distance in increasing order

Array_partition & Unroll ->

- Loop to count how many times each label has been observed for the first *k* elements
- Determine the class label

Project plan

Stream a

Week 7:

- Background research
- Planning Phase

Week 8:

- Writing the c++ code and verifying its accuracy
- Partition the code to accelerate

Week 9 & 10:

- Adding directives, comparing the performance and utilization of different designs
- Demo video
- Report writing

Roles

Together:

- Writing c++ code & verify the code's correctness
- Determine loops/functions to accelerate
- Add directives and make improvements
- Report writing
- Final Presentation

Zhixin:

- Zedboard testing (if available)

Shuyi:

- Demo video

Risks

- Distance calculation - Computational intensive
- Loading dataset - Memory-access intensive

When the size and dimensionality of the dataset increase, hardware resources may be a limitation of performance.

Thank you