PATTERN RECOGNITION

1   Introduction

In Mission 3, we split the incoming beam into two parts and project its polarisation with two separate polarizers. The beam intensities were measured with two separate photodetectors, giving voltages from detector 1 and detector 2. From the size of the signals measured, we try to identify the incoming polarisation to be: **H-Horizontal, V-Vertical, D-Diagonal or A-Antidiagonal**

A sample of the signal voltage from both detectors is plotted below.
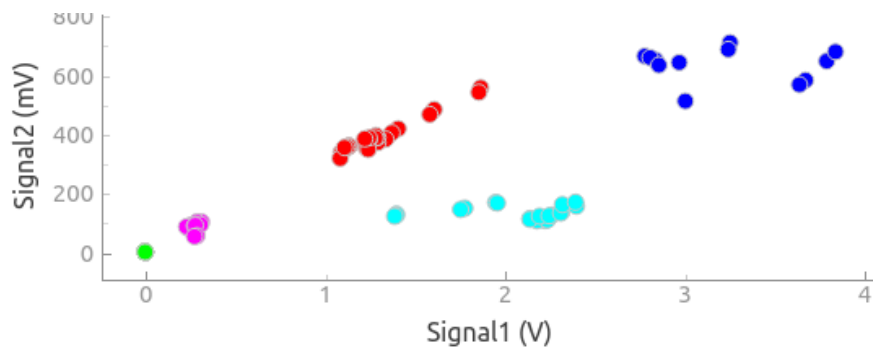Using a 'clustering' algorithm, we identify 5 clusters.

Figure 1: Signals from detectors 1 and 2.

Task 1 [1 pts]   Why are 5 clusters identified, when only 4 polarisations are being measured?

1.1 Motivation

Task 2a [2 pts] From Figure 1, sketch what you think the histogram for Signal 1 and Signal 2 might look like.

Just a rough sketch will do.

Task 2b [3 pts] Why do you think we made the effort to measure the beam using two detectors?

[Hint: observe the spread of the signals for each cluster using your answer in Task 2a.]

1.2 K-Means Clustering Algorithm

K-means clustering is an algorithm that finds groups in data, with the number of groups represented by K.[1] The algorithm works by iteratively assigning each point to one of the K groups based on the data provided.

The result of the algorithm are:

* The means of the K clusters, which can be used to label new data points.

* Labels for the old data.

Task 3 [4 pts] Make a simple guess of how you would design a clustering algorithm.

How would you determine if a data point would belong to one cluster instead of another?

Hint:

How would you kick-start the algorithm? eg. randomly assign the cluster centers, etc...

What calculations would you make using the data?

Do you think you will get the answer on the first iteration of your algorithm?

---

[1] https://www.datascience.com/blog/k-means-clustering