

# ML Exercise 1

---

## 1.5.1

It really depends on the context, what is the document needed for. The main advantage of using OCR would be the reduced file size, as you simply have a string of characters instead of a bitmap representing the entire “image” of the document. Having a plain-text document is also easier to work with computationally, allowing for searching in the document, filtering, compression etc.

If the document is of big importance, e.g. legal stuff, perhaps one wouldn't risk getting characters confused, such as in numbers and would thus prefer an image of the document. If formatting (incl. graphs, tables etc.) plays an important role, it may add too much complexity for an OCR to separate the various sections.

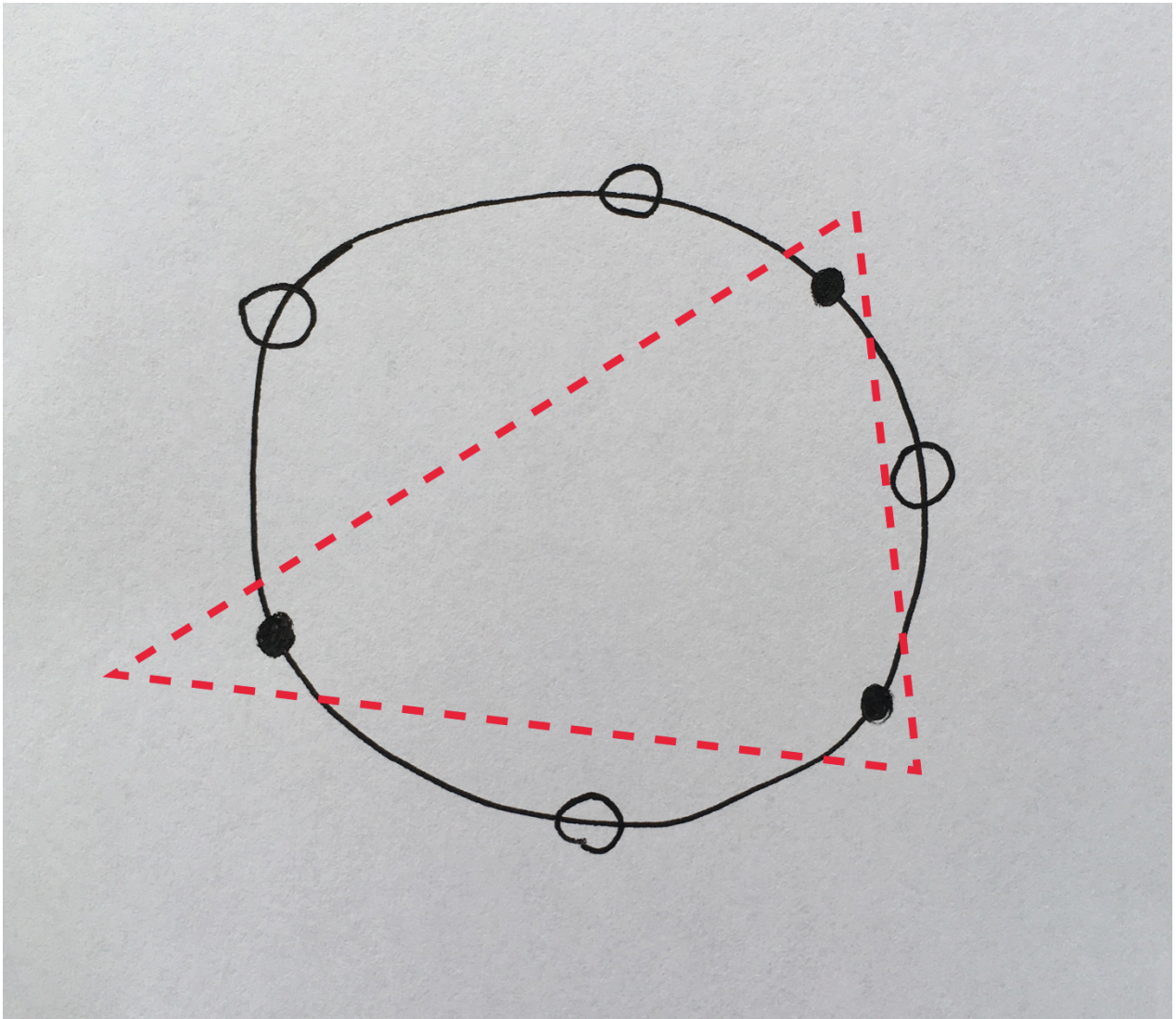
## 1.5.5

Such a problem could easily be formulated in terms of a graph, for instance building a bi-partite graph on the customer-item relationship and from there use various methods of network analysis to find dependencies between the items.

In more simple terms, you could probably build a probabilistic model by looking at how often item X appears when item Y is there and vice-versa.

## 2.10.9 (is this a real answer?)

With all points being placed equidistantly on a circle, it is easy to see that no matter the arrangement, it will always be possible to separate 3 (or fewer) points from the rest, thus the triangle can shatter 7 points.



2.10.11