ML Excercise 1

1.5.1

Imagine we have two possibilities: We can scan and email the image, or we can use an optical character reader (OCR) and send the text file. Discuss the advantage and disadvantages of the two approaches in a comparative manner. When would one be preferable over the other?

It really depends on the context, what is the document needed for. The main advantage of using OCR would 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

In basket analysis, we want to find the dependence between two items Xand Y. Given a database of customer transactions, how can we find these dependencies? How would we generalize this to more than two items?

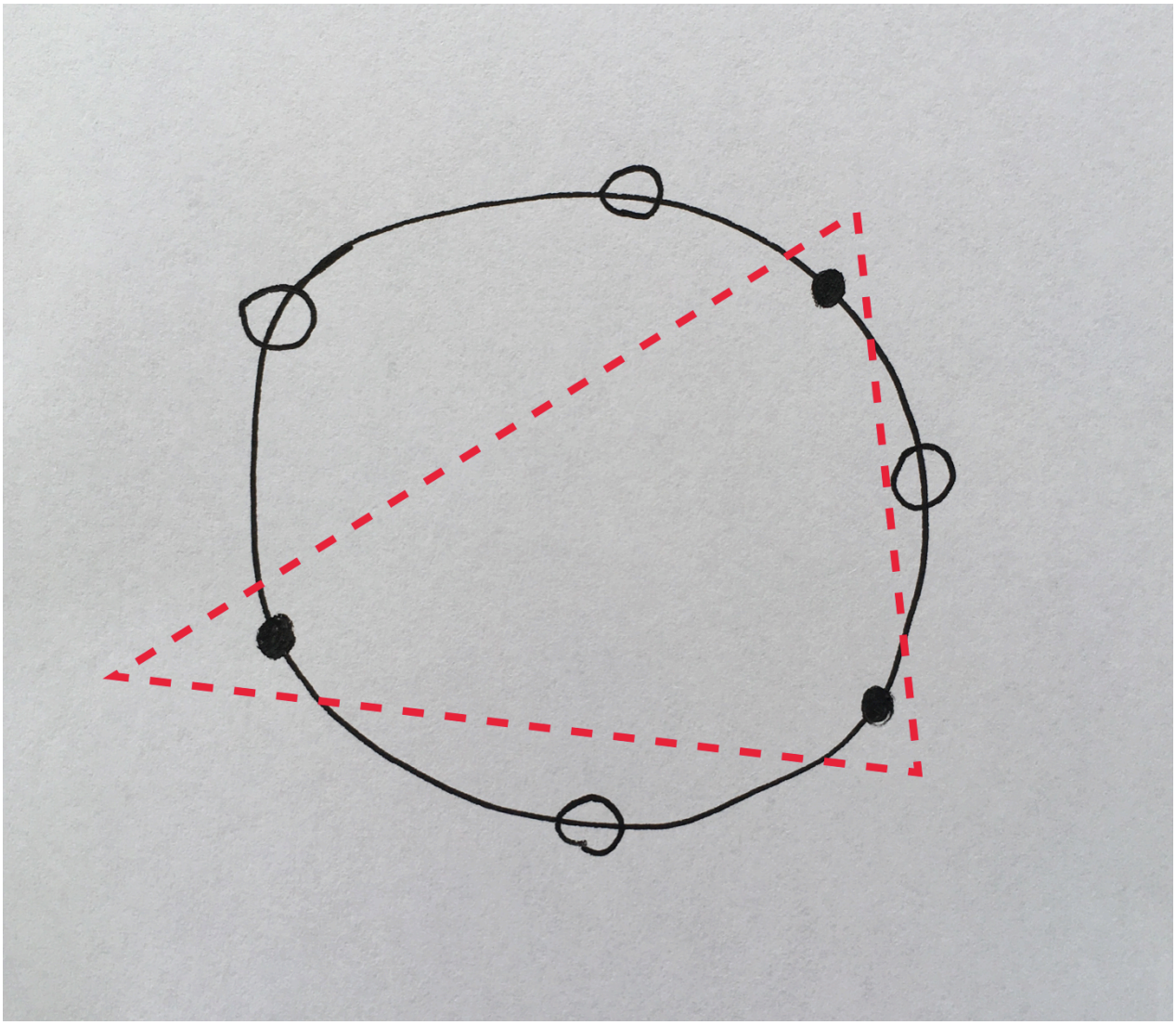
Such a problem could easily be formulated in terms of a graph, for instance building a bi-partite graph on the customer-item relationship and from threre use various methods of network analysis to find dependencies betweeen the items.

In more simple terms, you could probably build a probabalistic model, looking at the conditional probability for when item X was bought, how often item Y also was present. You could build on this by looking at the association rules mentioned in chapter 3, such as *support*, which shows how significant a relation is.

2.10.9 (is this a real answer?)

Show that the VC dimension of the triangle hypothesis class is 7 in two di- mensions. (Hint: For best separation, it is best to place the seven points equidistant on a circle.)

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.



With points arrannged in an alternating order, it wound't be possible to seperate the two groups.

Programing Exercise 1.1

B) You can't really describe a trigonometric function be a polynomial.