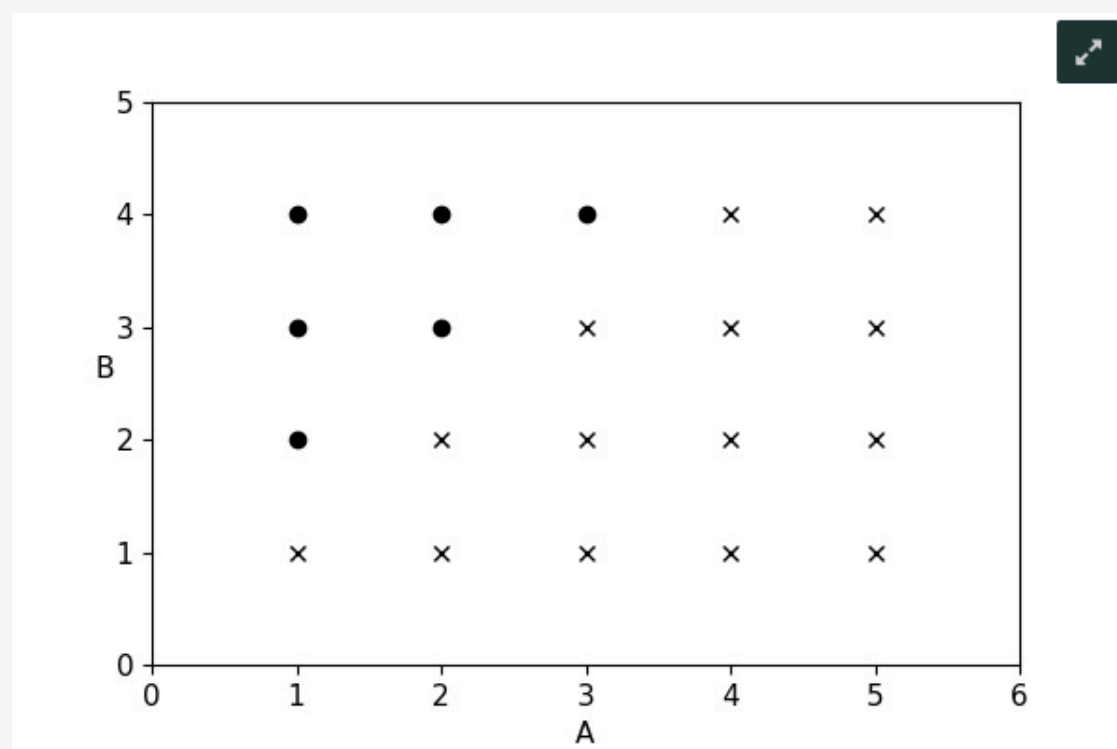


## Q2 Classification diagram

3 Points

This question is about classification. In the figure, each data point is a record in a dataset, the class of a record is shown by a data point's shape, and A and B are two discrete integer variables.



### Q2.1

1 Point

What is the entropy of the whole dataset to three decimal places?

### Q2.2

1 Point

Imagine that you used the criterion  $A > 2$  to split the data. Calculate the information gain to three decimal places?

### Q2.3

1 Point

To three decimal places, what is the minimum total entropy that can be achieved by two splits of the data?