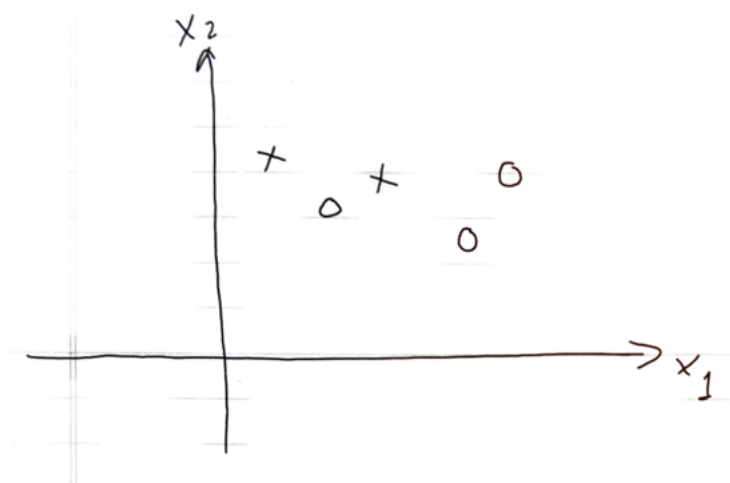# Exam 1

Problem 1

Decide the best vertical first split by impurity gain for the following dataset. Impurity can be measured by entropy or Gini-Index.

Problem 2

- Given the **training** data.

Class	Sex	Survived
1	0	1
1	0	1
1	0	1
1	1	1
1	0	1
1	0	1
1	0	1
3	1	1
3	0	1
2	1	0
3	0	1
2	1	0
2	1	0
3	1	0
2	1	0
3	0	0
1	0	0

Class	Sex	Survived
2	0	0
1	1	0
2	1	0

Using Gini Index as the measure for impurity to:

1. Grow the **maximum tree** with three leaves (a stopping rule!) on the data. Draw the (diagram of) tree.
2. Find the misclassification rate on training data of the maximal tree
3. Draw all the possible **subtrees**
4. Validate the maximal tree and all the subtrees on the following data to select the **optimal tree**. Note: if the chance of Survived is 1/2, predict `Survived`

Class	Sex	Survived
1	0	1
1	1	1
3	1	0
2	0	0