Exercise

VC-Dimension of other hypothesis spaces, e.g. intervals in R :

$h(x) = +1$ if $a \le x \le b$, $h(x) = -1$ otherwise.

Definition

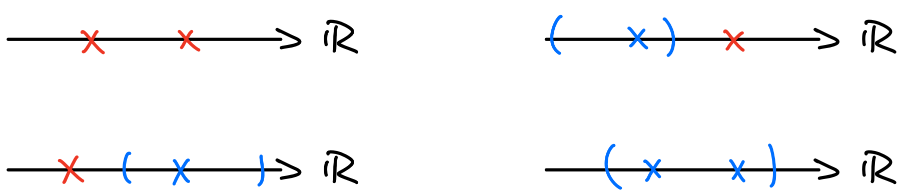
The VC-dimension of a hypothesis space H defined over an instance space X is the size of the largest finite subset of X shattered by H: $VC(H) = \max|S|$: S is shattered by H $S \subseteq X$

If arbitrarily large finite sets of X can be shattered by H, then $VC(H) = \infty$.

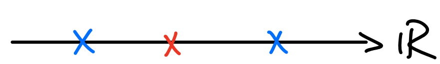
Solution

We want to find the size of the largest set that can classify all different combinations of positive (1) and negative (-1) labels by an interval on the real line.

For a set of size 2, we can classify all different combinations as follows, where blue represents points within the interval, i.e. labelled positive (1).



For a set of size 3, we simply cannot classify all different combinations with an interval on the real line. Suppose we have 3 points in ascending order, we cannot label the first and last points with the same label by 1 single interv , as shown below.



Therefore, the VC-Dimension of intervals in R is 2.