

# If a classifier has 75% accuracy, it means that ...

- A. It correctly classifies 75% of the data items in the training set
- B. It correctly classifies 100% of the data items in the training set but only 75% in the test set
- C. It correctly classifies 75% of the data items in the test set
- D. It correctly classifies 75% of the unknown data items

**Given the distribution of positive and negative samples for attributes  $A_1$  and  $A_2$ , which is the best attribute for splitting?**

$A_1$	P	N
a	2	2
b	4	0
$A_2$	P	N
x	3	1
y	3	1

- A.  $A_1$
- B.  $A_2$
- C. They are the same
- D. There is not enough information to answer the question

# When splitting a continuous attribute, its values need to be sorted ...

- A. to avoid overfitting
- B. to prune the data
- C. to define a binary split condition
- D. to accelerate tree induction

**The computational cost for constructing a RF with K as compared to constructing K decision trees on the same data**

- A. is identical
- B. is on average larger
- C. is on average smaller