

03/02/2023

feature selection

filter-based

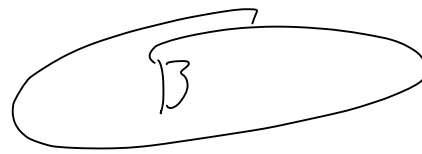
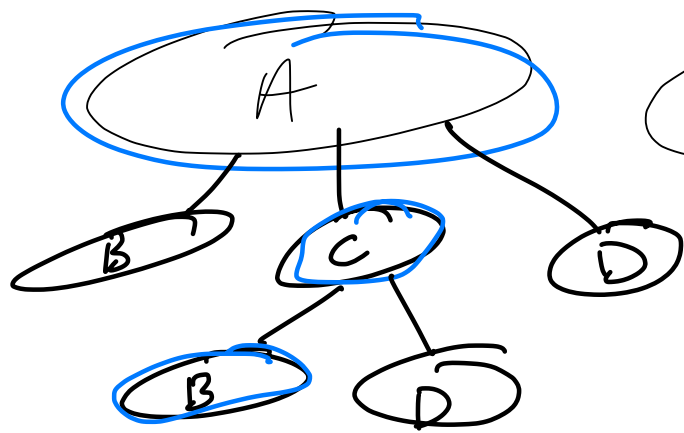
- features ranked prior to ML model

wrapper-based

- feature subset determined by using ML model to score

embedded method

- predictor builds with subset of features selected by model



A

ABC

B

ABD

C

ACD

D

BCD

AB

ABCD

AC

AD

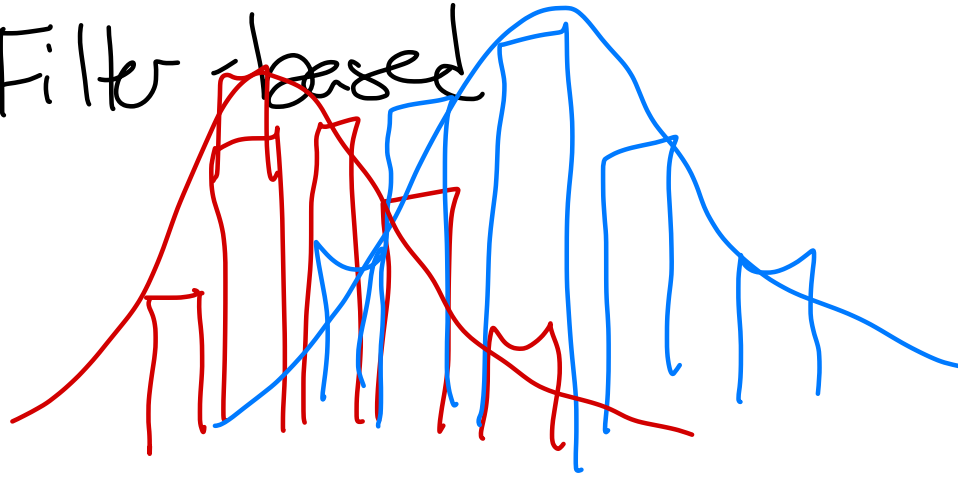
BC

BD

CD

$$2^p - 1$$

Filter based



male

female

	survived	died
male		
female		

$$\chi^2$$

$$\chi^2 = \sum_{v \in \text{Facts}} \sum_{c \in \text{Classes}} \frac{(O_{vc} - E_{vc})^2}{E_{vc}}$$

- permutation feature importance

train model on all feats
evaluate on validation

repeat:

shuffle vales across all samples of
1 feature

predict and evaluate model