

# module 5: features

## DATA PROBLEM

RAW DATA  
spambase  
20 newsgroups

## LABELS

DATA  
PROCESSING  
missing values

## REPRESENTATION

FEATURES  
extraction  
sparsity  
correlation

SELECTION

DIMENSIONS  
PCA

## LEARNING

CLUSTERING

SUPERVISED  
LEARNING  
naive bayes  
regul. regression

## PERFORMANCE

EVALUATION

ANALYSIS  
boosting margins  
feature analysis

TUNING

