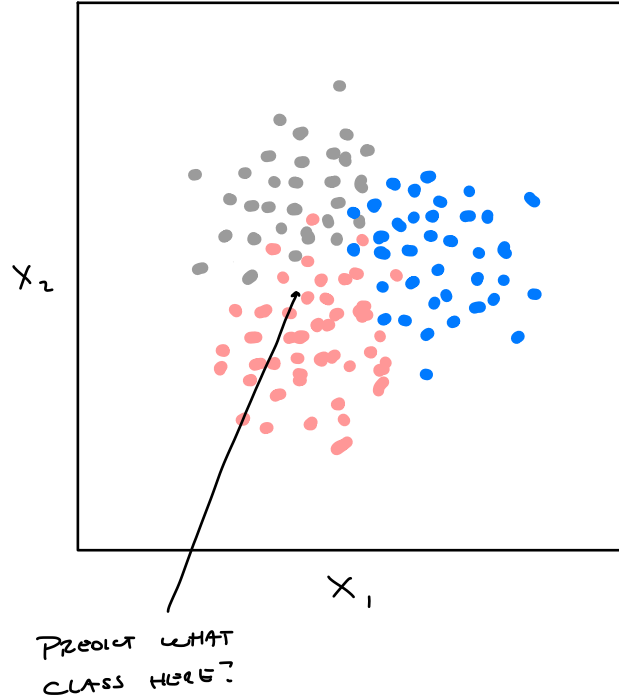
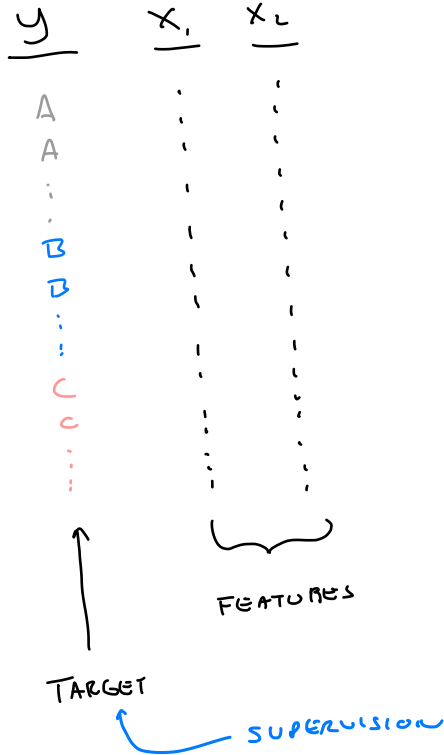


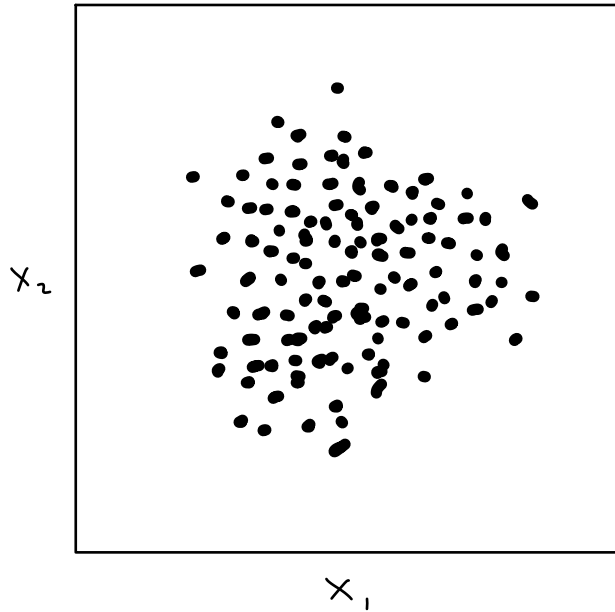
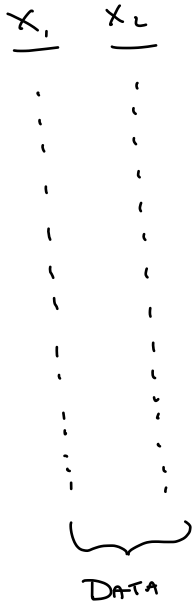
CS 307

UNSUPERVISED
LEARNING

SUPERVISED LEARNING



UNSUPERVISED LEARNING



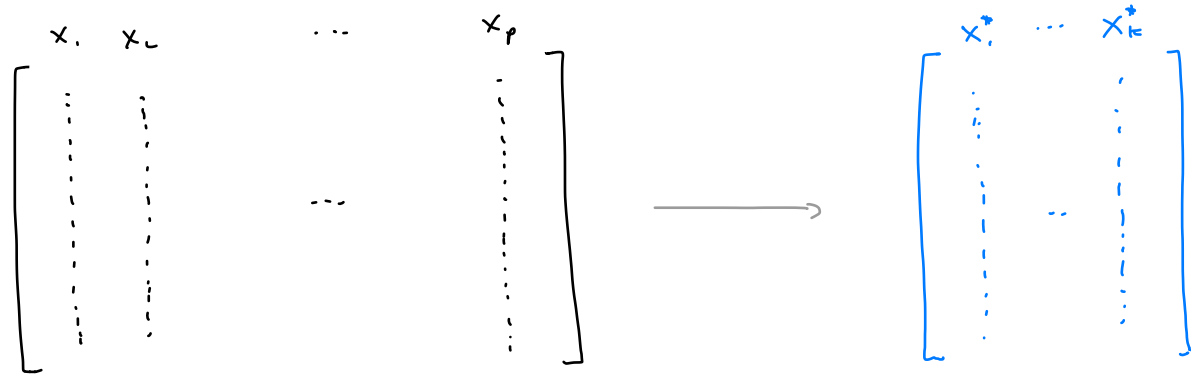
???

UNSUPERVISED LEARNING

- DIMENSION REDUCTION PCA
- CLUSTERING K-MEANS, HIERARCHICAL CLUSTERING, DBSCAN
- DENSITY ESTIMATION KDE, MIXTURES
- OUTLIER DETECTION ONE-CLASS SVM, ISOLATION FOREST

DIMENSION REDUCTION

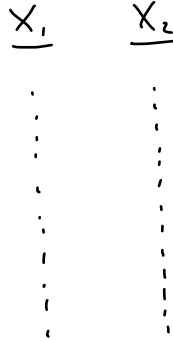
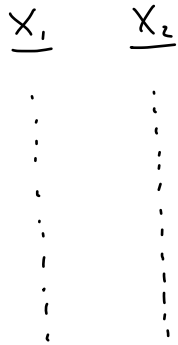
$$1 \leq k \leq p$$



OFTEN USED AS PREPROCESSING FOR SUPERVISED LEARNING

EXAMPLE: PCA └─┬─┘ CAN BE USED IN A PIPELINE w/ SKLEARN

CLUSTERING



CLUSTER ASSIGNMENT

A
B
B
C
B
A
A
C

NOT A TARGET

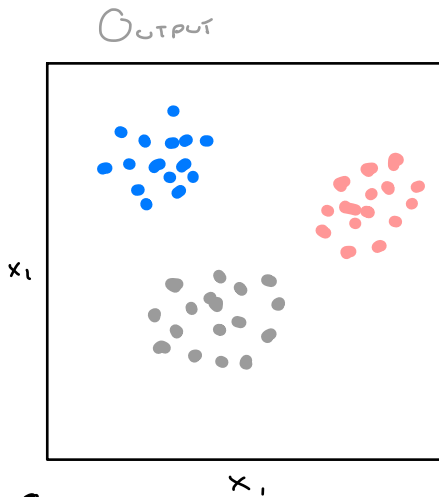
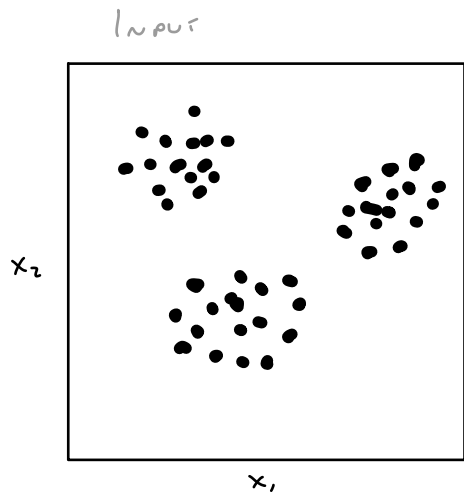
HARD TO VALIDATE

- NO STANDARD TO COMPARE TO IN PRACTICE
- ORDER / LABELS MEANINGLESS

EXAMPLES

K-MEANS
HIERARCHICAL

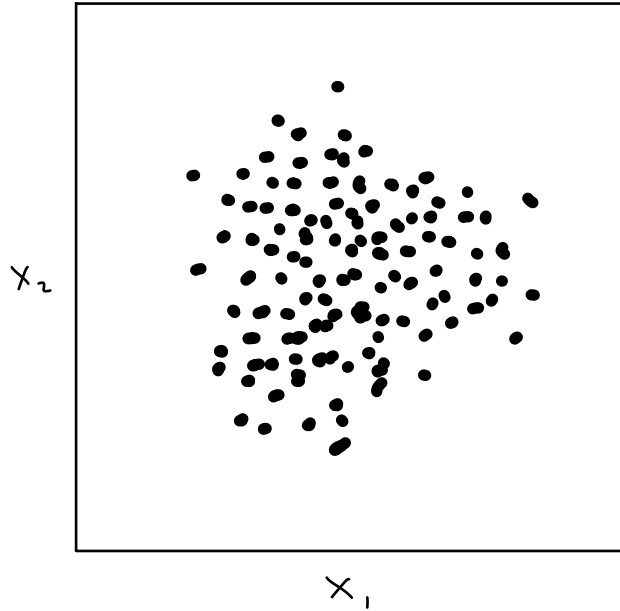
CLUSTERING



↑
COLORS HAVE NO MEANING!
COULD BE SWAPPED!

ONLY GROUPINGS MATTER

CLUSTERING

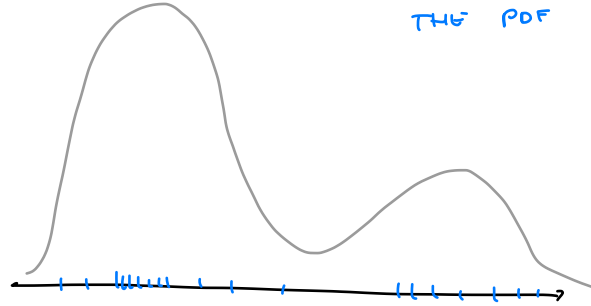
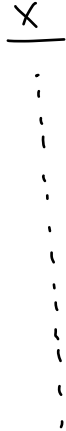


???

• How many clusters?

OFTEN NEEDS TO BE SET
BY USER BEFORE FITTING

DENSITY ESTIMATION



↑
RUG PLOT (OF GIVEN DATA)

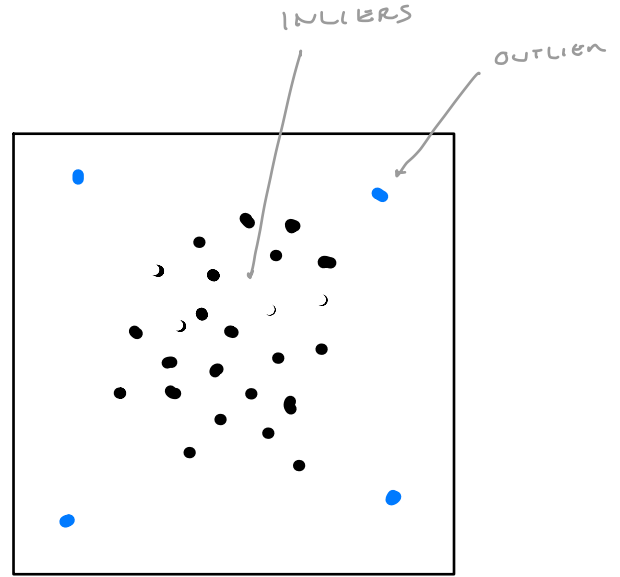
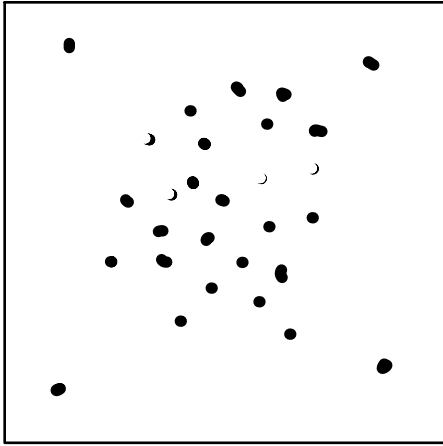
↙ DENSITY (LEARNED FROM DATA)
↳ SPECIFICALLY $f(x)$
THE PDF

EXAMPLES

KDE
MIXTURE MODELS

COULD BE USED TO GENERATE
NEW DATA

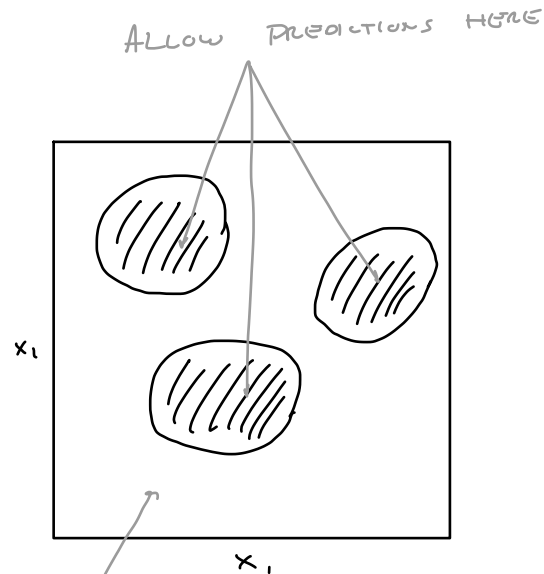
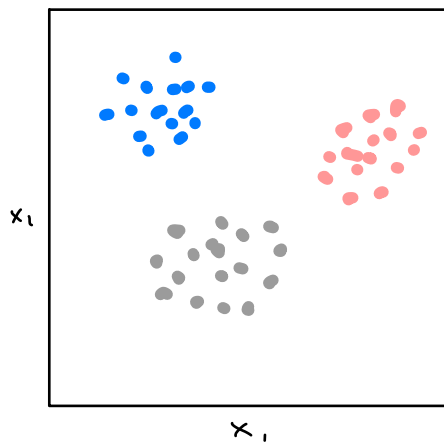
OUTLIER DETECTION



OUTLIER DETECTION

- WHEN SHOULD OUTLIERS BE REMOVED FROM TRAINING DATA? ONLY VERY CAREFULLY!
- ELIMINATE "OUTLIERS" AT TEST TIME?
 - ↳ NOVELTY DETECTOR
 - ↳ USEFUL WITHIN SUPERVISED PIPELINES

NOVELTY DETECTION



REFUSE TO PREDICT HERE