

25 SEP 2015

~~MODSAB. FILTERING~~

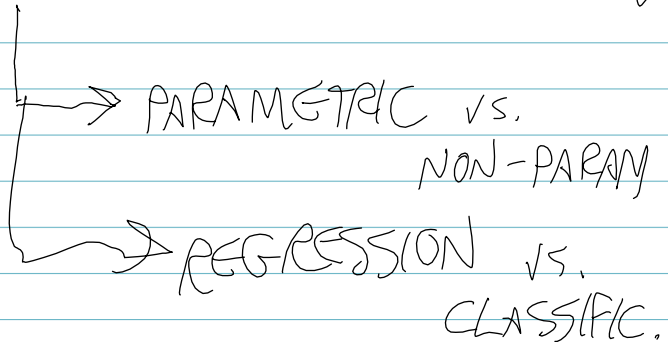
REV. LAST TIME

BASIC ML CONCEPTS

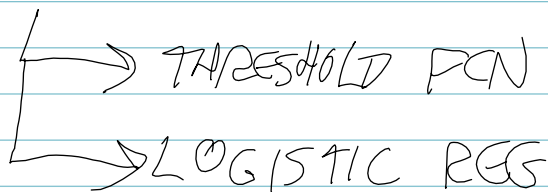
- FEATURE

- DISTANCE FUNCTIONS

SUPERVISED VS. UNSUPERVISED



LIN. REGRESSION FOR CLASS



DECISION TREE

RANDOM FOREST

SVM

TESTING ML ALG

CROSS-FOLD VAL

"VALIDATION" SETS

RMSQ, PREC/RECALL, ACCUR, ETC.

COLLAB FILTERING

CONTENT-BASED K-NN

WHY K-NN AND NOT
~~SUPERVISED~~ ~~ML~~ PARAMETRIC
APPROACH?

- ONLY KEEPING TRACK
OF WHAT USER
PURCHASED
- FIND PRODUCTS SIMILAR
TO WHAT USER
IS LOOKING AT.

ASSN-RULE CF

NEW INPIV:

BOUGHT ITES a, b, c

DATABASE

$$a, b, e \rightarrow z \quad 0.9$$

$$a, c, q \rightarrow w \quad 0.87$$

•
z
b
•
c

$w = 4$

ITEM = ABFW

SEARCH TO DEPTH 4 \rightarrow NOTHING
STOPPED AT 3

NEXT ITER GO TO DEPTH 2

FOUND AB !

CHILDREN OF NODE ~~AB~~
ARE

ABC (4) ~~ABE~~ (5)

AB \rightarrow C

AB \rightarrow E

MATRIX FACTORIZATION

HAVE MATRIX T

USERS \times MOVIES

EACH USER VECTOR
IS A SERIES
OF RATINGS OR
P? FOR EA
MOVIE

FOR SOME VALUE K
(EMPIRICALLY DETERMINED)
FIND MATRICES F
AND G S.T.

$$T \approx \cancel{G} G^T F$$

$$P_{u,j} = \sum_{k=1}^K G_{k,j} \cdot F_{k,j}$$

USER MOVIE
 PREDICTED RATING
 MATRIX REL. MOVIE, TO ATTRIB.
 MATRIX REL. USERS TO ATTRIB.