**features:**

iid,

gender,(only keep iid’s)

piid,

match,(only keep iid’s)

like,(only keep iid’s)

prob,(only keep iid’s)

age,

field\_cd,(change to categorical)

race, (change to categorical)

imprace,

imprelig,

from,

goal, (1,2,5,6--0; 3,4---1)

date,

go out,

career\_c,(change to categorical)

sports(\*17),

attr1\_1(\*6),

attr2\_1(\*6),

attr3\_1(\*5),

attr,

sinc,

intel,

fun,

amb,

shar,

undergrad,

pink : index

yellow: label for classification/ prediction

green: keep both / diff

blue: keep both / inner product

black: unchanged

**remove:**

id,

idg,

condtn,

wave,

pid,

age\_o,

race\_o,

pf\_o\_att,

pf\_o\_sin,

pf\_o\_int,

pf\_o\_fun,

pf\_o\_amb,

pf\_o\_sha,

dec,

met,

satis\_2,

length,

numdat\_2,

mn\_sat,

tuition,

uni\_id,

uni\_pid,

round,

int\_corr,

income,

exphappy,

samerace