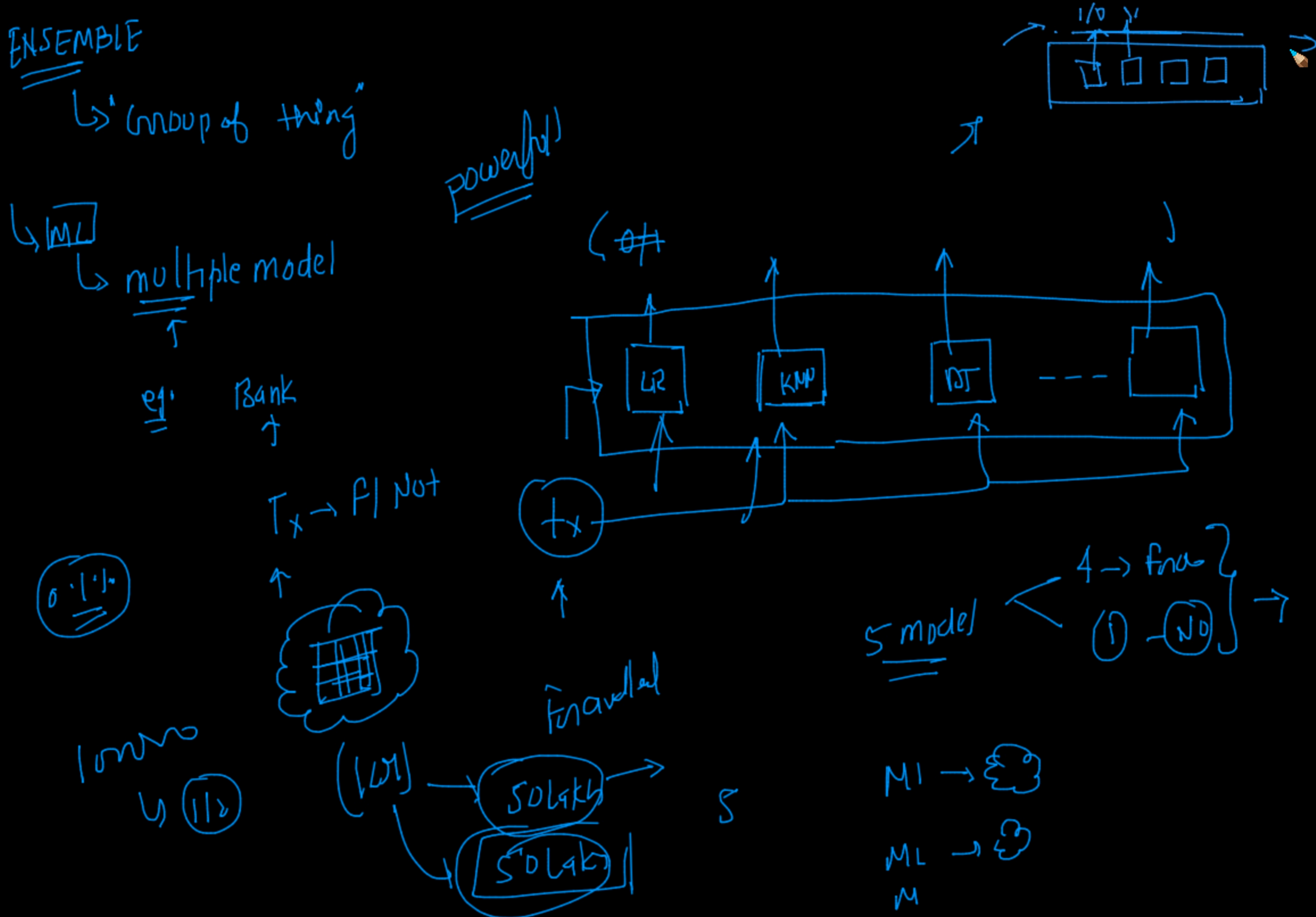
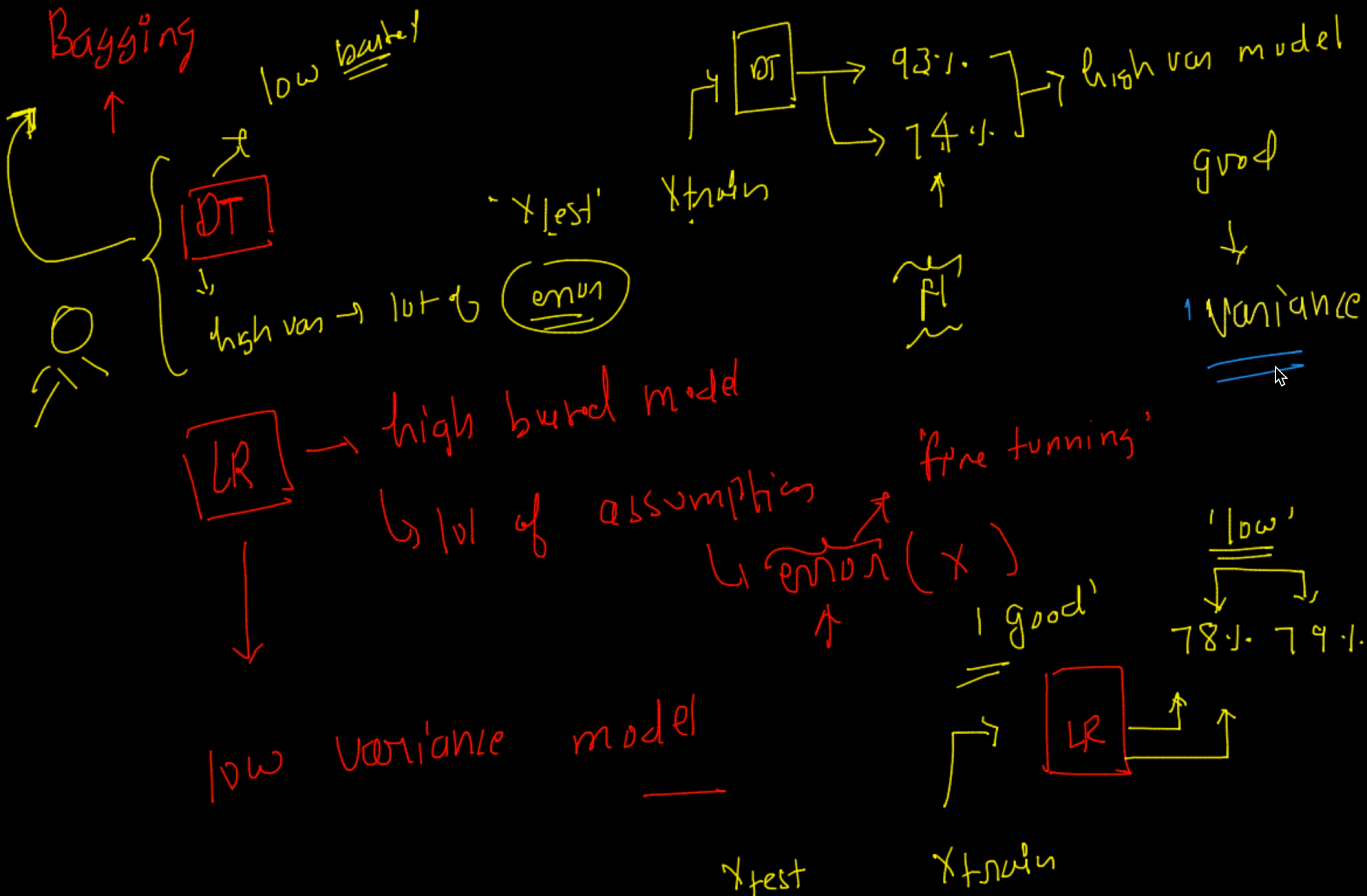
Safari File Edit View History Bookmarks Window Help Thu Dec 11 10:39PM



1) Bagging
2) Boosting
3) Stacking
4) cascading
(Acc ↑)
proven
compilit
Ds hack
SDl's
Dx → ML → f( ) → 88%
LR( )
DT( )
Random Model
accept
Industry
1) Black Box
2)
high performing
very powerfull


Bagging
sklearn
eg.
sampling
Bootstrap Aggregation
'Random Forest'
HUO
Techs
fn
D_n
{x, y}
Sample
m pts
D_n^1
D_n^2
D_n^3
D_n
Bootstrap sampling
M1
M2
M3
Mk
final output
?
final model
Mi
is built using D_n' of size m
m ≤ n
3 / 4

X
Y
income
CS
Status
17
Yes
16
20K
50K
env
DT
M1
M2
M3
'Regression'
[income, CS]
Agg(0, 0, 1)
750
650
74%
93%
86%


Bagging
low bias
DT
high var → lot of error
'X test'
X train
DT
93%
74%
high var model
good
variance
LR
high bias model
lot of assumptions
error (x)
'fine tunning'
'low'
78% 79%
'good'
LR
low variance model
X test
X train
5 / 9

Bagging + take bunch of low bias model and high var model

↑ ↓↓ ↑

Combine using bagging

↓↓

(low bias + low variance)

on based on tech → Random Forest ← sklearn

Bagging tec

follow the Bagging tech

Random Forest

Interview

D. Tree

Sample Bootstrap

RF ⇒ DT + Bagging + Col. sampling

model

Row sampling with Replacement.

Custom

python is


Training
n
m
D'
RS + CS
Feature sampling
f1 f2 f3 f4
100
20
RS
DT
n-estimators
10/5

d = 4
M
100
d = 3
CV
fine
DT
CV
80
CV
find out
20
DT
CV
bag of point
test
DT(depth=2)
100%
93%
fine tune
depth = ?
CV
train
test
CV
test
d=1
DT(train)
f( )
90%
d=2
DT(train) → f'( )
94
test