feature:

1. submission\_list( former submissions of a image)

2. title\_len\_list (the title length)

3. raw\_time\_list ( four time periods during a day, and the timezone(-7,-8, a vector containing 6 binary elements)

4. date\_list(month1-11, day(Mon-Sat))

dataset:

1. train set : test set = 3 : 1 (99230 :33077)

2. 10 fold cross validation on train set

3. Normalize: StandardScalar

predict label :comment number of a post

A. binary classification:

**a) whether or not the submission is good**

1 : data[i]['number\_of\_comments']\*0.2 + data[i]['total\_votes']\*0.8 > 36 (large)

0. otherwise (not large)

**b) model used**

# 

# 3. LR

**validation**

Accuracy: 0.66 (+/- 0.01)

Recall: 0.66 (+/- 0.01)

Precision: 0.66 (+/- 0.01)

**test**

precision: 64.67%,

recall: 69.77%

accuracy: 66.05%

Theta :

[-0.37144436, -0.02347957, 0.04094704, 0.00194088, 0.02391902,

0.01820491, -0.01376737, 0.38655975, 0.07524588, 0.12291697,

0.14388276, 0.10014861, 0.08800185, 0.05255167, -0.07999241,

-0.12938848, -0.16938931, -0.14619977, -0.01037853, -0.03111783,

-0.05102258, -0.03895545, -0.0582152 , -0.03594115, 0.02645251]

# 5. DT

**validation**

Accuracy: 0.66 (+/- 0.01)

Recall: 0.66 (+/- 0.01)

Precision: 0.66 (+/- 0.01)

**test**

precision: 66.87%, recall: 62.95%

accuracy: 66.11%

# 6. ADA

**validation**

Accuracy: 0.66 (+/- 0.01)

Recall: 0.67 (+/- 0.01)

Precision: 0.67 (+/- 0.01)

**test**

precision: 67.53%, recall: 65.45%

accuracy: 67.21%