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))

5. subreddit\_list (subreddit submissions)

dataset:

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

2. 10 fold cross validation on train set

3. Normalize: StandardScalar

predict label : whether or not the submission is good(large)

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: 0.6654 (+/- 0.0126)

Recall: 0.6653 (+/- 0.0126)

Precision: 0.6655 (+/- 0.0126)

**test**

precision: 66.25%,

recall: 68.62%

accuracy: 67.05%

Theta :

[-0.37396174, -0.02265814, 0.04470522, 0.00250932, 0.02213239,

0.01592305, -0.01753274, 0.3815239 , 0.08349381, 0.12348568,

0.14289216, 0.09816523, 0.08920768, 0.05048658, -0.08792127,

-0.13319385, -0.18189006, -0.14015651, 0.007246 , -0.03192661,

-0.0500869 , -0.03870152, -0.05920357, -0.03730283, 0.02765636,

-0.29692625]

# 5. DT

**validation**

Accuracy: 0.6742 (+/- 0.0101)

Recall: 0.6741 (+/- 0.0101)

Precision: 0.6745 (+/- 0.0102)

**test**

precision: 66.38%,

recall: 70.41%

accuracy: 67.58%

# 6. ADA

**validation**

Accuracy: 0.6742 (+/- 0.0101)

Recall: 0.7076 (+/- 0.0086)

Precision: 0.7077 (+/- 0.0086)

**test**

precision: 71.62%,

recall: 70.64%

accuracy: 71.51%