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)

6. community effects(whether the submission was large before in popular subreddits, whether the submission was very large (>200) in the same subreddit before)

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: 0.6687 (+/- 0.0121)

Recall: 0.6686 (+/- 0.0121)

Precision: 0.6688 (+/- 0.0121)

**test**

precision: 66.66%,

recall: 69.04%

accuracy: 67.47%

Theta :

[-0.3366103 , -0.02542098, 0.04494 , 0.00407469, 0.02336471,

0.01814905, -0.01906309, 0.37406267, 0.08326419, 0.12162135,

0.14154005, 0.09865107, 0.09033634, 0.05236045, -0.08474114,

-0.1314046 , -0.18346188, -0.14551801, 0.00487096, -0.03288222,

-0.05171157, -0.04012235, -0.05949376, -0.03799053, 0.0283868 ,

-0.31000271, -0.17097606, 0.07547683]

# 5. DT

**validation**

Accuracy: 0.6742 (+/- 0.0101)

Recall: 0.6740 (+/- 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: 0.7113 (+/- 0.0119)

Precision: 0.7114 (+/- 0.0119)

**test**

precision: 71.82%,

recall: 70.60%

accuracy: 71.58%(even a little worse)