HW3: Short report

Takuya Ando

Below are the results which compare the performance of 7 models for each metric.

For constructing models, I fixed threshold to 0.2. The reason for this low level of threshold is that, for this case, the project would be at risk if they cannot gather enough money in specific period. Considering the importance of education, we have to decrease such failure cases as least as possible. Therefore, we should pay attention to recall at most in this case, in order to avoid false negative case.

Regarding accuracy, SVM recorded highest, whereas logistic regression, bagging and boosting showed relatively row accuracy. However, looking at recall, SVM recorded lowest, only less than 50%. On the other hand, logistic regression, bagging and boosting performed much better for recall. Also, these models shows higher performance than SVM for f1 score, which is kind of aggregation of precision and recall.For about precision, all models had similar figures, around 30%.

Overall, logistic regression, bagging and boosting would be the best model to predict the project with high risk of failing in getting fully funded. As discussed above, recall should be most important metric in this case, and these models revealed best for this metric, as well as theyrecorded relatively high performace for other metrics such as f1, ap and auc. Actually, I used logistic regression for bagging here, so logistic regression or boosting would be the best model for this prediction.

In [1]:

```
%run -i revmlpipeline.py
```

In [2]:

```
%run -i reportdata.py
```