Kaggle實作範例

2018/2/21(三)

Kaggle ensembling guide

model correlation & weighing (e.g. by voting)

```
1111111100 = 80% accuracy
1111111100 = 80% accuracy
1011111100 = 70% accuracy.
```

```
1111111100 = 80% accuracy
```

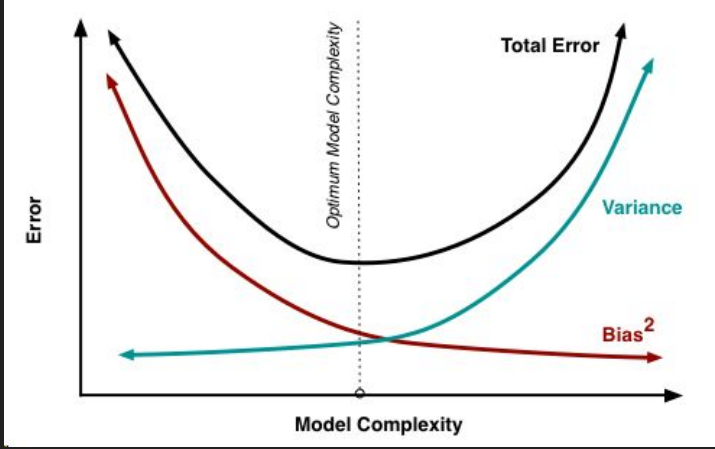
```
1111111100 = 80% accuracy
0111011101 = 70% accuracy
1000101111 = 60% accuracy
```

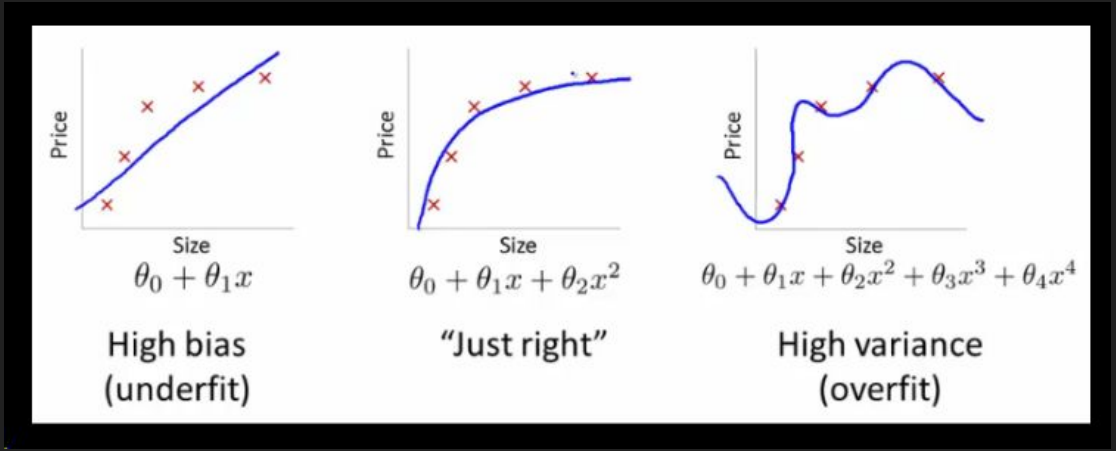
```
1111111101 = 90% accuracy
```

https://mlwave.com/kaggle-ensembling-guide/

Kaggle ensembling guide

Averaging predictions often reduces overfit. You ideally want a smooth separation between classes, and a single model's predictions can be a little rough around the edges.

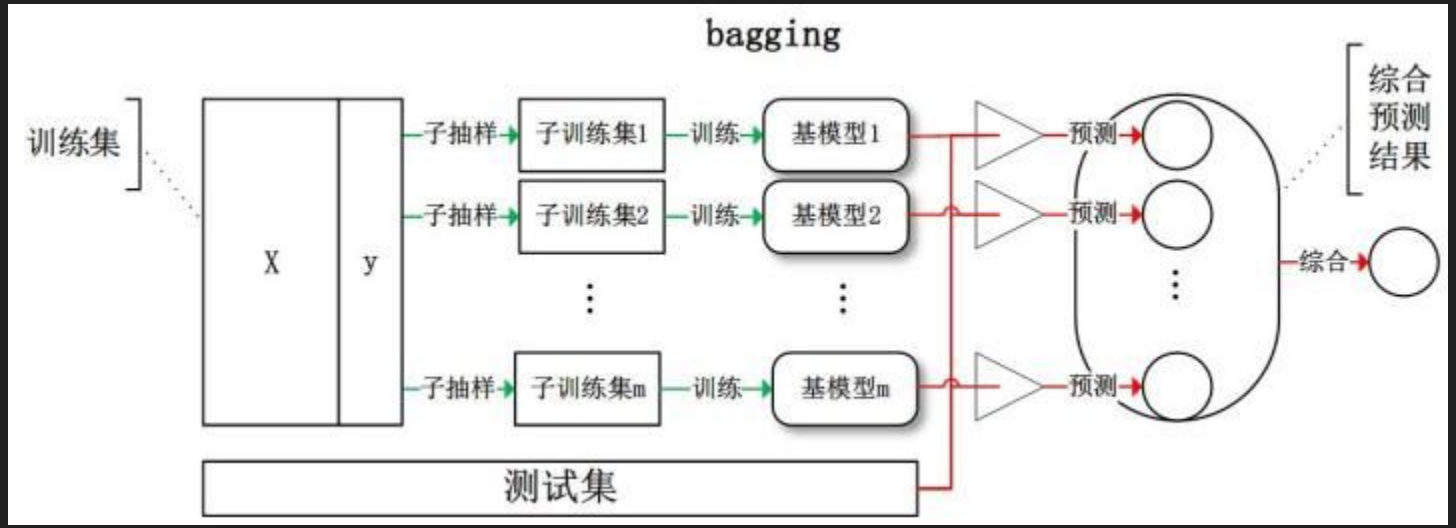




https://kevinbinz.com/tag/overfitting/

Kaggle ensembling guide

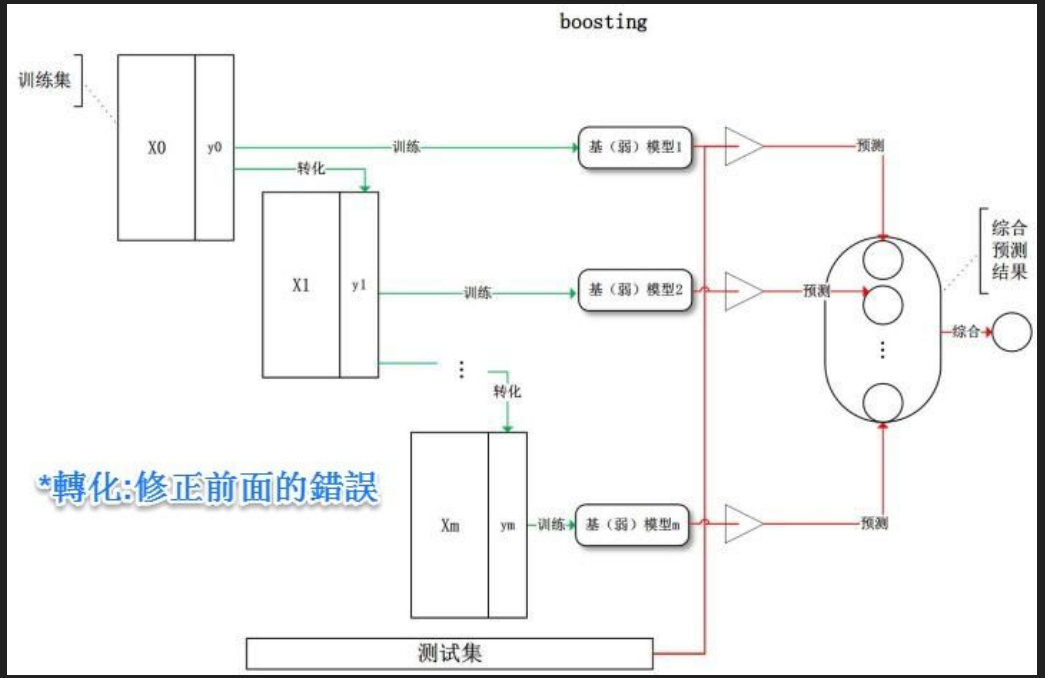
bagging



https://www.zhihu.com/question/29036379

Kaggle ensembling guide

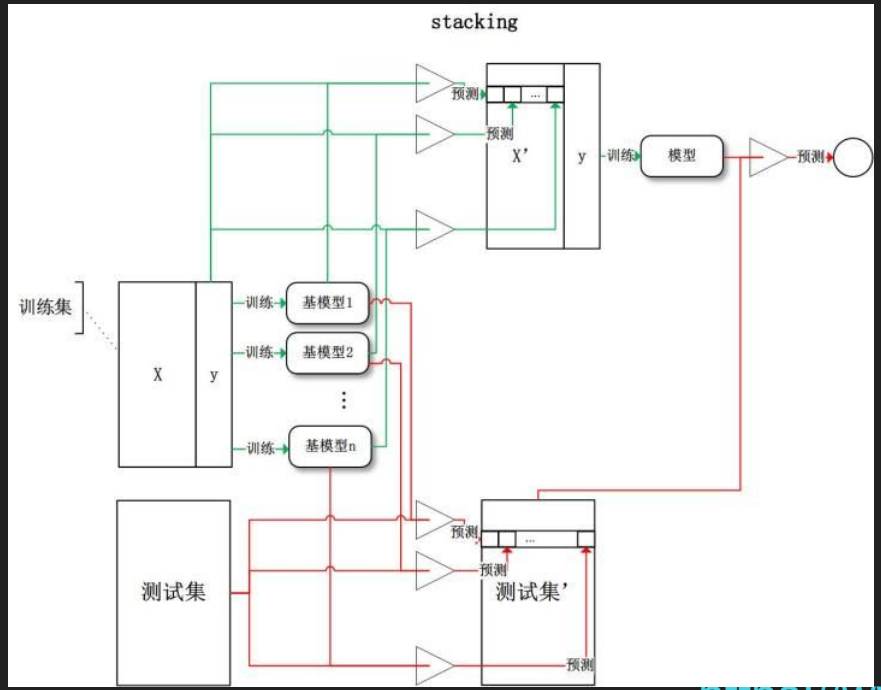
boosting



https://www.zhihu.com/question/29036379

Kaggle ensembling guide

stacking



https://www.zhihu.com/question/29036379

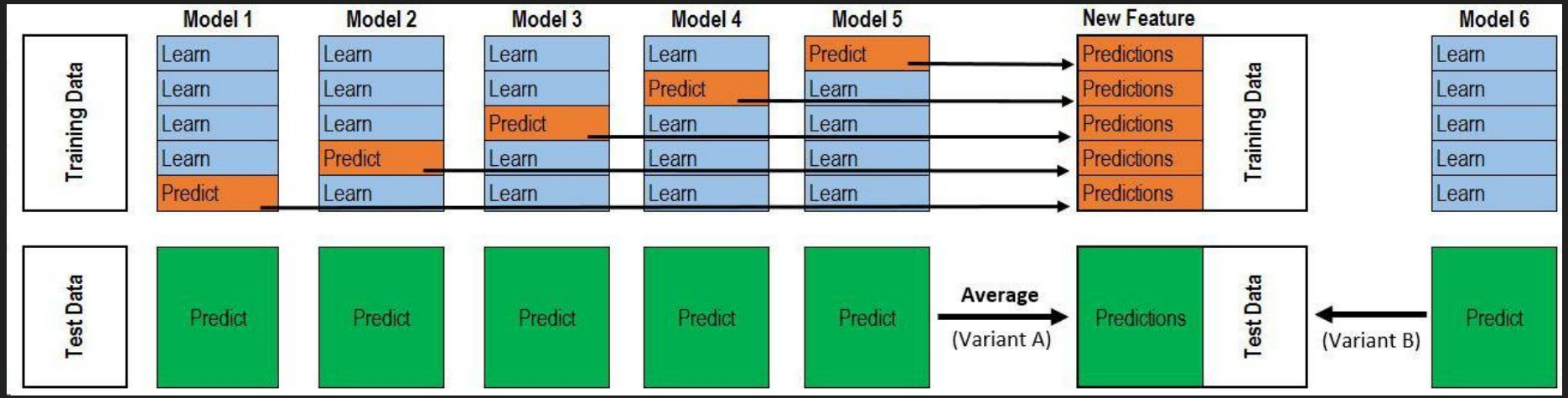
Classification library

1.scikit-learn classification library

2.ensemble library: mlxtend (http://rasbt.github.io/mlxtend/)

github:https://github.com/stuser/temp/blob/master/kaggle_intro/kaggle_intro_iris.ipynb

n-fold stacking



github:https://github.com/stuser/temp/blob/master/kaggle_intro/kaggle_intro_iris3.ipynb