

# GROUP 8

*Ying Zhu*

*11/1/2016*

## Group member:

Yiwei Sun

Xuechun Sun

Jiani Tian

Shujin Cao

Ying Zhu

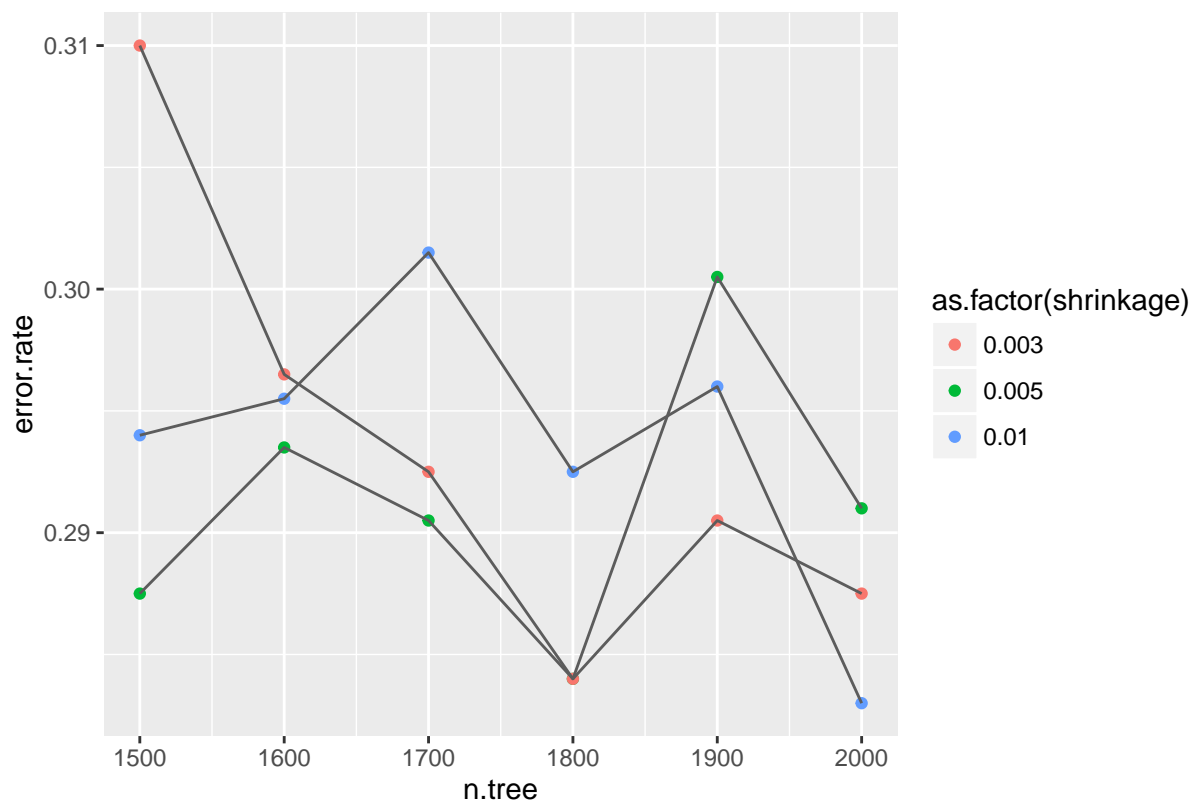
## Baseline model

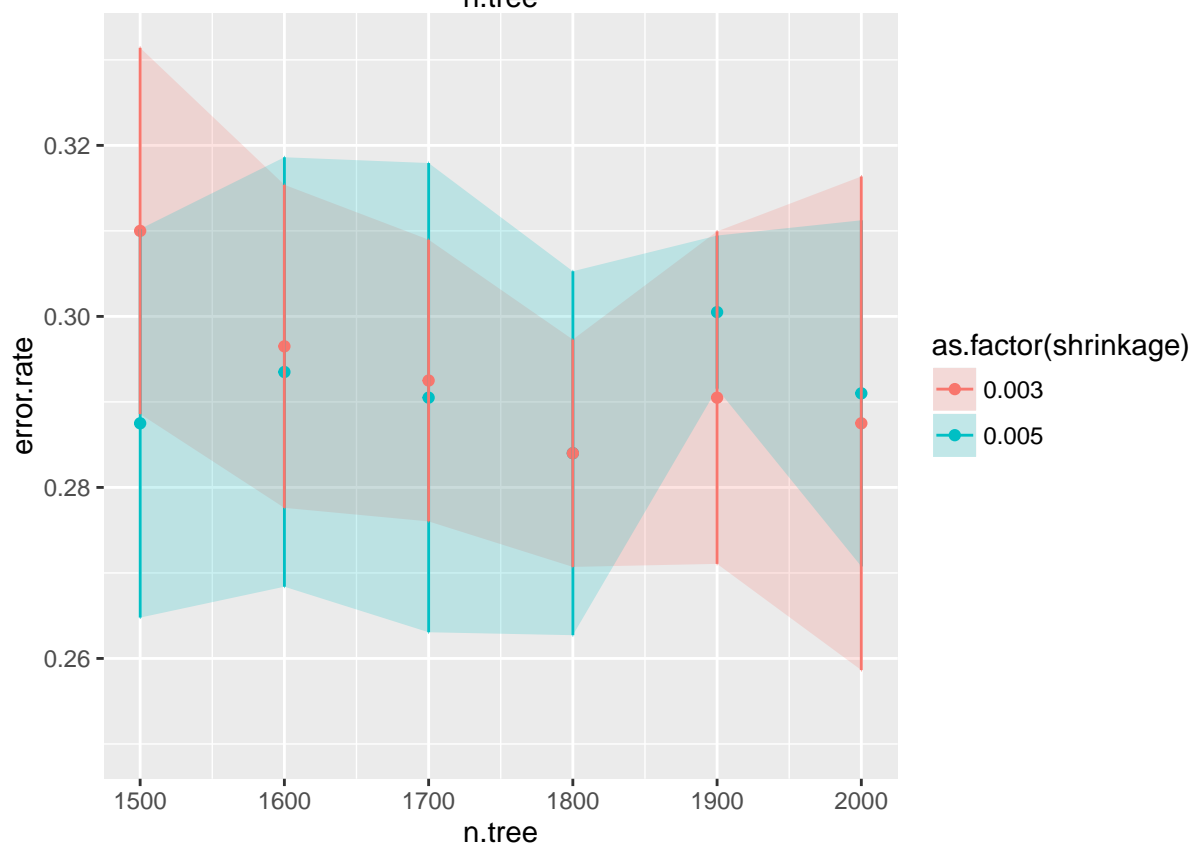
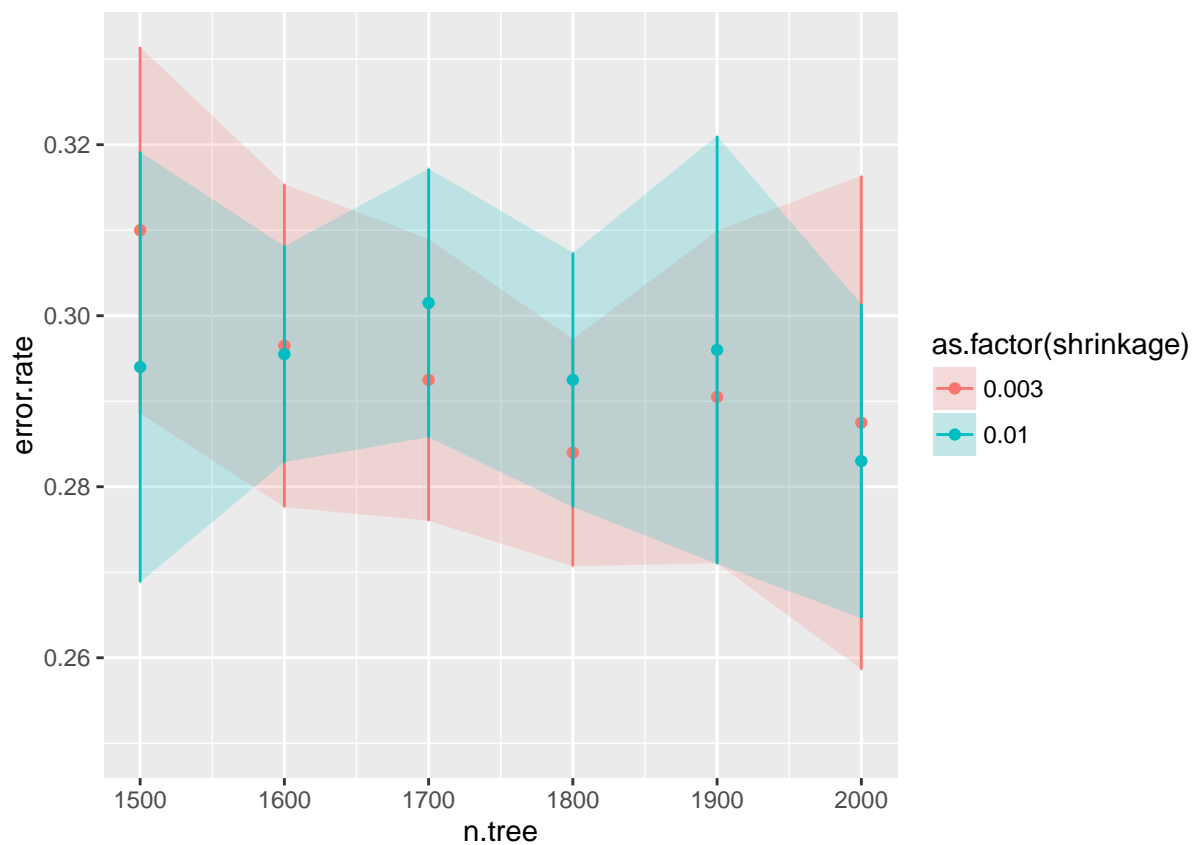
n.tree=1800, shrinkage=0.003

Basic status	Values
Error rate	0.2840
Estimate time	40 mins

When finding the best model of baseline, we have a lot of try.

## Combine of three





## n.tree error.rate sd shrinkage

## 1	1500	0.2875	0.02592000	0.005
## 2	1600	0.2935	0.02864655	0.005
## 3	1700	0.2905	0.03130000	0.005
## 4	1800	0.2840	0.02427705	0.005
## 5	1900	0.3005	0.01021641	0.005
## 6	2000	0.2910	0.02308950	0.005
## 7	1500	0.2940	0.02870100	0.010
## 8	1600	0.2955	0.01440000	0.010
## 9	1700	0.3015	0.01790600	0.010
## 10	1800	0.2925	0.01695582	0.010
## 11	1900	0.2960	0.02853726	0.010
## 12	2000	0.2830	0.02094914	0.010
## 13	1500	0.3100	0.02443103	0.003
## 14	1600	0.2965	0.02154936	0.003
## 15	1700	0.2925	0.01879162	0.003
## 16	1800	0.2840	0.01516575	0.003
## 17	1900	0.2905	0.02217826	0.003
## 18	2000	0.2875	0.03292985	0.003

Based on the picture, we choose n.tree=1800, shrinkage=0.003 to fit our baseline model.

Although n.tree=2000, shrinkage=0.003 has a lower error rate but combine the sd we try to use the another one

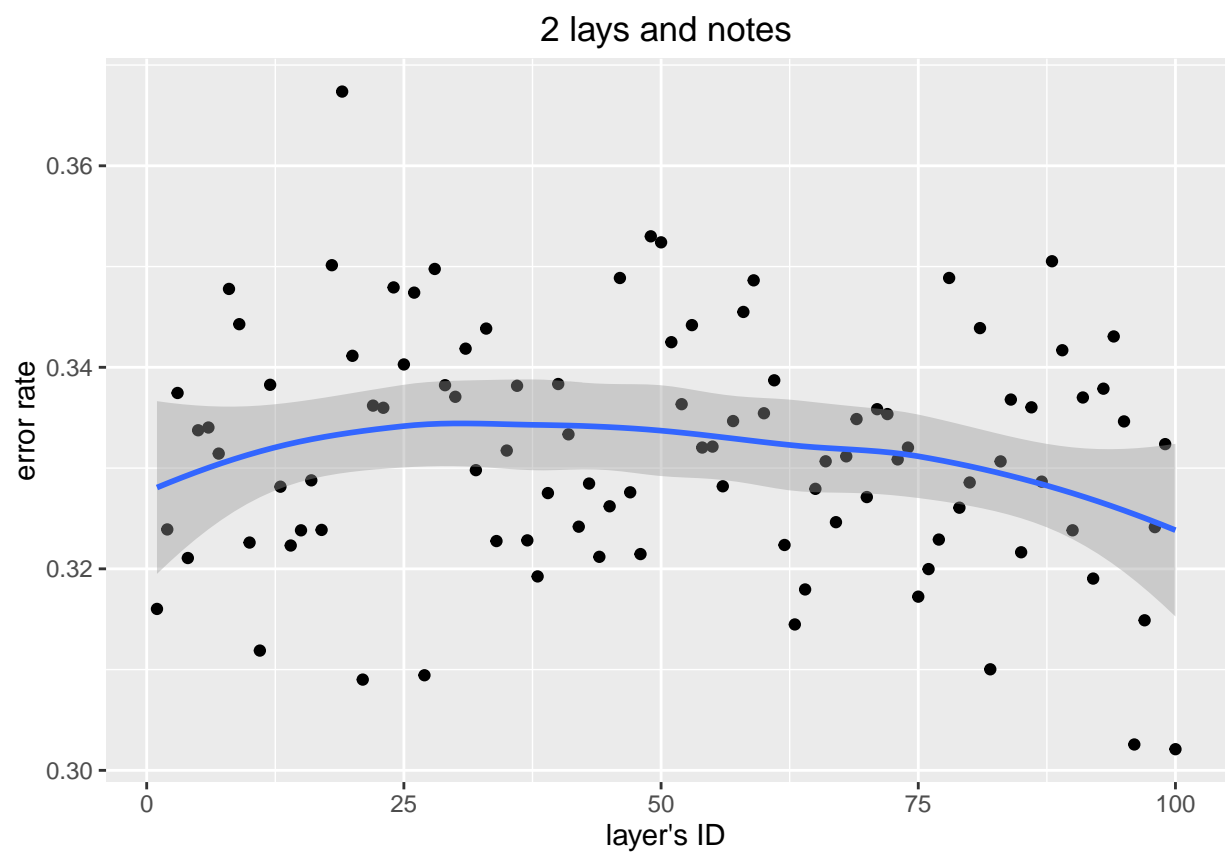
## Improve Models

### Neural Network

Basic status	Values
Error rate	0.3340
Estimate tiem	25 mins

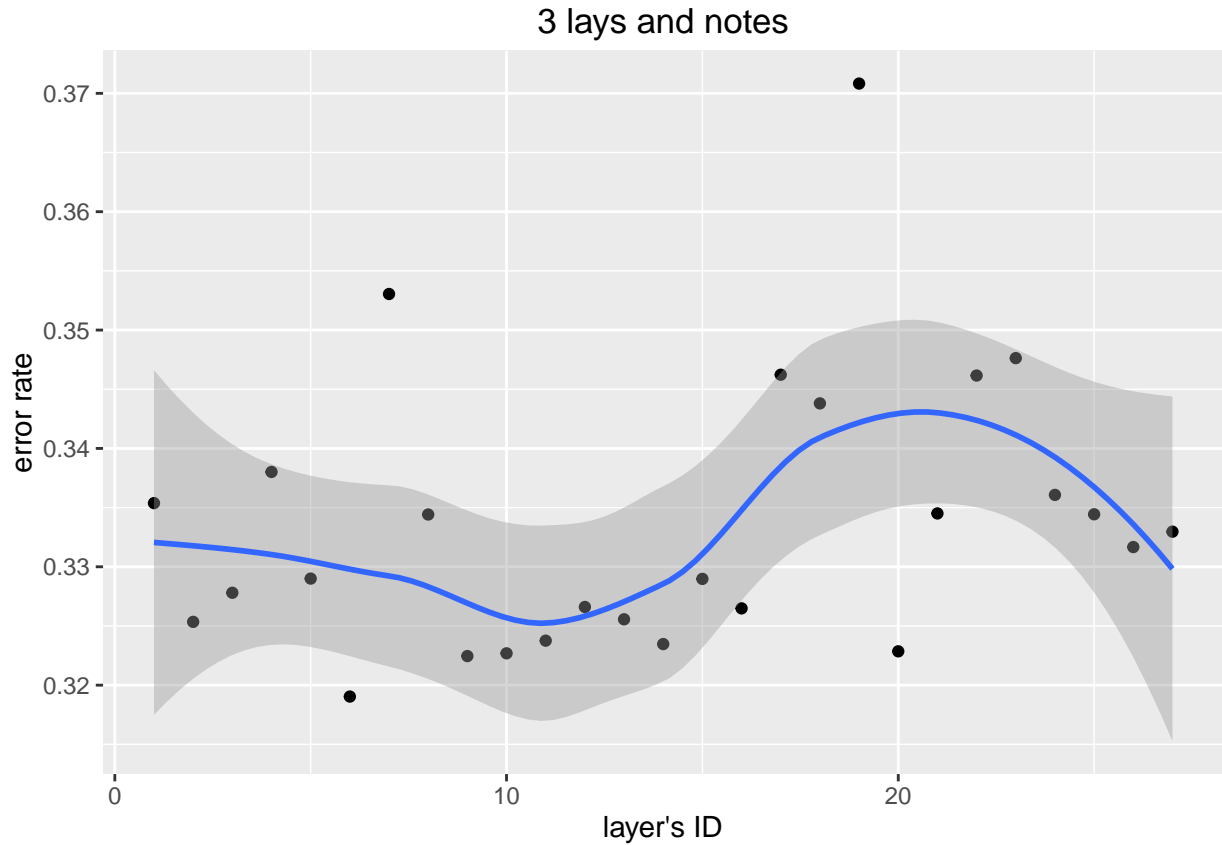
```
## [1] "mean error"      "0.331709083355"
```

```
## [1] "minimal error" "0.3021002189"
```



```
## [1] "mean error"      "0.333674166540741"
```

```
## [1] "minimal error" "0.3190361672"
```



#### Random Forest

Basic status	Values
Error rate	0.2930
Estimate tiem	20 mins

n tree is the times that random forest train, Mtry sometimes needs to base on our data scale. we have 5000 variables in train data, therefore we first tried the square root of 5000, which is around 70

By this picture, we finally choose random forest when mtry=70 as our advanced model.

#### Methodology: Advance model by new method to extract features

Focus on: TEXTURE, AND COLOR;

Basic status	Values
Error rate	0.1295
Estimate tiem	277 s

Therefore, we choose the method that combining the rgb, orb, and sift features, then use pca to reduce the dimension as our new features.

for ORB (Oriented FAST and Rotated BRIEF):

ORB is basically a fusion of FAST keypoint detector and BRIEF descriptor with many modifications to

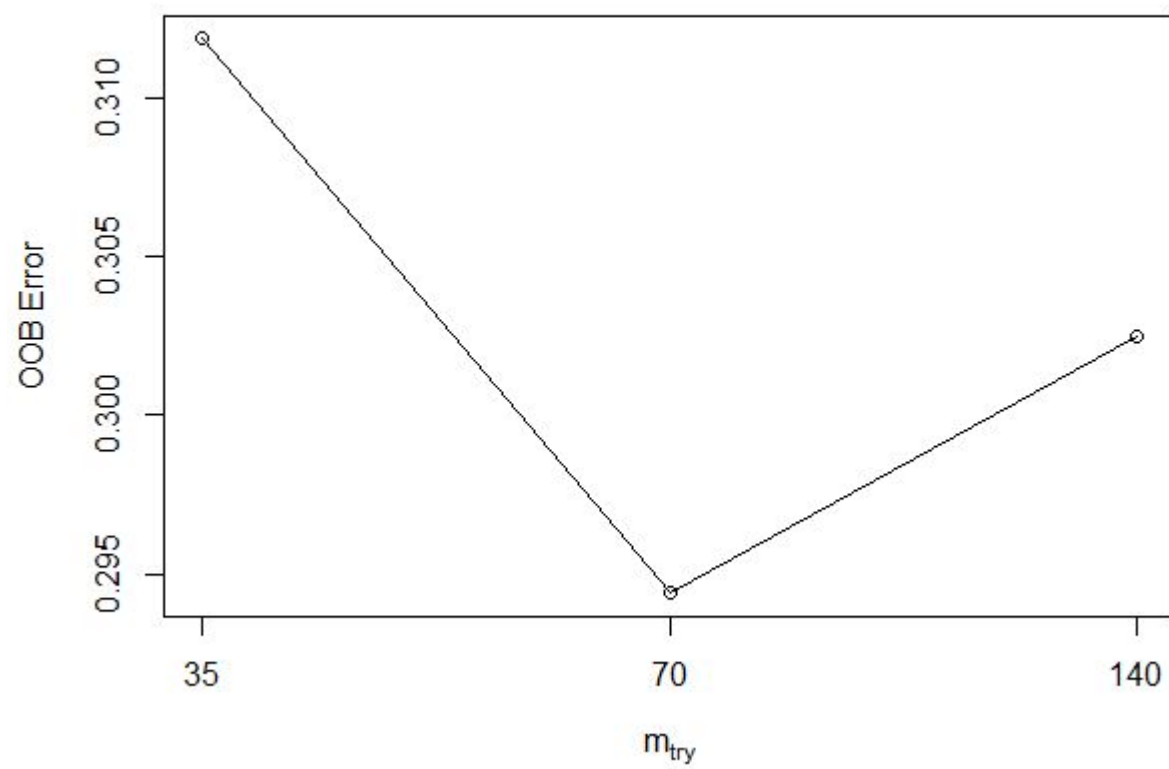


Figure 1:



Figure 2:  
7



Figure 3:



enhance the performance. First it use FAST to find keypoints, then apply Harris corner measure to find top N points among them. It also use pyramid to produce multiscale-features.

first we use Median Filter to get rid of noise of each image

then we use ORB Method to extract edge and corner point feature

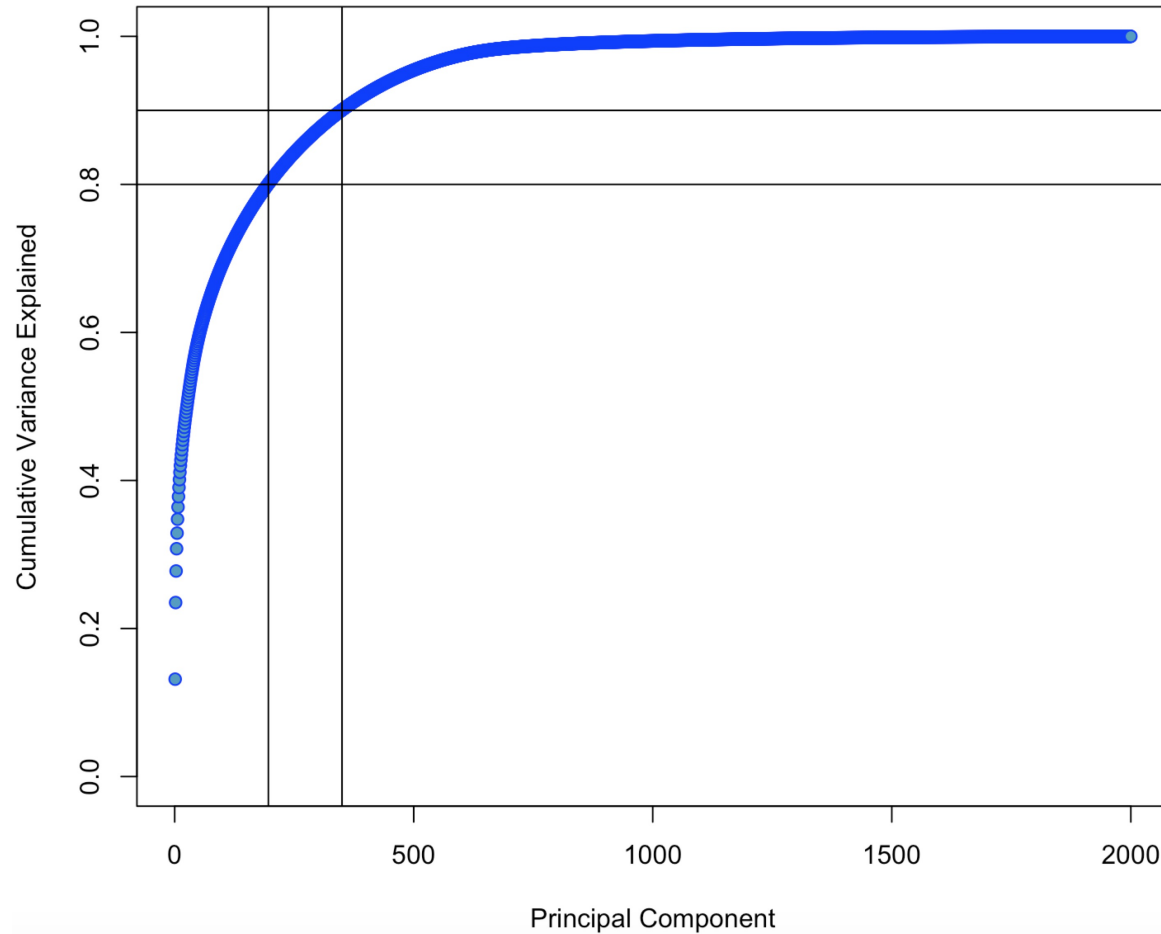


Figure 4:

After pca, we choose 196 variables to as the new features. Those 194 variables can explain the more than 80% vairiance.

By mofidy this we reduced our error rate to 12.95%, sd is 0.01848