Machine Learning Report

1. kNN
2. adaboost
3. Random Forest

Using the random forest way, we need to train a decision tree first. We post the result of a single decision tree below. We use 500 pictures to train and ‘test-data.txt’ to get the result.

|  |  |  |
| --- | --- | --- |
| FeatureNum | Accuracy rate | Training time/s |
| 8 | 0.3796394485683987 | 1.220587985477323 |
| 16 | 0.4305408271474019 | 3.421173229226042 |
| 24 | 0.4835630965005302 | 4.80559430448011 |
| 32 | 0.513255567338282 | 6.9898500861036155 |
| 48 | 0.49946977730646874 | 11.104106148221035 |
| 56 | 0.5143160127253447 | 13.94556450896016 |
| 64 | 0.5185577942735949 | 17.2981764367305 |
| 96 | 0.5376458112407211 | 29.18401044858183 |
| 144 | 0.5312831389183457 | 49.493924550410156 |
| 192 | 0.5726405090137858 | 76.28177388826771 |

We can also plot it.



Here is the result when we use 50 trees.

|  |  |  |
| --- | --- | --- |
| FeatureNum | Accuracy rate | Training time/s |
| 16 | 0.5546129374337222 | 162.1433158650543 |
| 32 | 0.6002120890774125 | 373.5613685701355 |
| 96 | 0.6383881230116649 | 1716.581987196077 |
| 192 | 0.7020148462354189 | 4432.005874951358 |

We can find that random forest provide a quite good result when we use more features. But it also cost a lot of time to train.

1. best

For the best result, we use a random forest contains 100 trees and each tree uses 192 features. In our test dataset, we achieved a 0.7243 accuracy.