Q3.2

= 1;%10; % Number of trees

= 4; % trees depth

= 5;%3; % Number of split functions to try

= 'IG'; % Currently support 'information gain' only

= 3;

find k-means optimal parameters

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **param.num** | 100 | 200 | **300** | 400 | 500 | 100 | 200 | 300 | 400 | 500 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG | IG | IG |
| **param.split\_func** | 1 | 1 | **1** | 1 | 1 | 4 | 4 | 4 | 4 | 4 |
| Accuracy(10 average) | 0.5940 | 0.6167 | 0.6327 | 0.580 | 0.650 | 0.5840 | 0.6380 | 0.6220 | 0.6507 | 0.6047 |
| ticktok | 1.3722 | 2.7417 | 4.1335 | 5.5182 | 7.0601 | 2.0763 | 4.2476 | 6.3647 | 8.3169 | 10.5151 |

Average accuracy of using 4th split function is better than using 1st

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| param.num | 300 | 300 | ~~300~~ | ~~300~~ | ~~300~~ | 300 | 300 | 300 | 300 |  |
| **param.depth** | 6 | 10 | ~~14~~ | ~~18~~ | ~~22~~ | 10 | 10 | 10 | 10 |  |
| param.splitNum | 5 | 5 | ~~5~~ | ~~5~~ | ~~5~~ | 20 | 50 | 70 | 100 |  |
| param.split | IG | IG | ~~IG~~ | ~~IG~~ | ~~IG~~ | IG | IG | IG | IG |  |
| param.split\_func | 1 | 1 | ~~1~~ | ~~1~~ | ~~1~~ | 1 | 1 | 1 | 1 |  |
| accuracy | 0.6593 | 0.6873 |  |  |  |  |  |  |  |  |
| ticktok | 7.3385 | 18.6867 |  |  |  |  |  |  |  |  |
| param.num | 300 | 300 | 400 | 400 | 300 | 300 | 300 | 400 | 400 | 150 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 5 | 50 | 100 | 150 | 100 | 100 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG | IG | IG |
| param.split\_func | 1 | 4 | 4 | 1 | 2 | 1 | 4 | 4 | 1 | 2 |
| accuracy |  |  | 0.52 |  |  |  |  |  |  |  |
| param.num | 300 | 300 | 400 | 400 | 300 | 300 | 300 | 400 | 400 | 150 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 5 | 50 | 100 | 150 | 100 | 100 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG | IG | IG |
| param.split\_func | 1 | 4 | 4 | 1 | 2 | 1 | 4 | 4 | 1 | 2 |
| accuracy |  |  | 0.52 |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| After find the optimal setting here vary vocabulary size |  |  |  |  |  |  |  |  |  |  |
| No of clusters | 128 | 512 | 1024 |  |  |  |  |  |  |  |

conf = confusionmat(data\_test(:,end), predictions);

Q3.3

present the optimal setting for RF codebook and RF classifier

Calculate the time consuming of the optimal setting