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 | 350 | **400** | 500 | 100 | 200 | 300 | 400 | 500 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG | IG | IG | IG |
| **param.split\_func** | 1 | 1 | 1 | 1 | **1** | 1 | 4 | 4 | 4 | 4 | 4 |
| Accuracy | 0.5940 | 0.6503 | 0.6107 | 0.6263 | 0.6797 | 0.6503 | 0.5570 | 0.6647 | 0.6020 | 0.6937 | 0.6233 |
| (best in 20) | 0.6 | 0.7067 | 0.6400 | 0.6800 | 0.7133 | 0.6733 | 0.6133 | 0.7067 | 0.6333 | 0.7467 | 0.6667 |
| ticktok | 1.3722 | 1.9215 | 2.9050 | 3.5629 | 3.8724 | 5.1314 | 1.7353 | 2.7257 | 4.1619 | 5.5968 | 6.8707 |

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| param.num | 400 | 400 | 400 | 400 | 400 | 400 | 400 | 400 |
| **param.depth** | 6 | 8 | **9** | 10 | 9 | 9 | 9 | 9 |
| **param.splitNum** | 5 | 5 | 5 | 5 | 20 | 30 | 50 | 10 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG |
| **param.split\_func** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Accuracy(mean) | 0.6073 | 0.6483 | 0.7127 | 0.6780 | 0.7003 | 0.6597 | 0.6790 | 0.6830 |
| (best in 20) | 0.6467 | 0.6733 | 0.7667 | 0.7133 | 0.7267 | 0.7000 | 0.7067 | 0.7067 |
| Tic toc | 3.8798 | 10.7982 | 14.1452 | 21.0341 | 25.7252 | 33.6883 | 50.9313 | 18.4548 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| param.num | 300 | 300 | 400 | 400 | 300 | 300 | 400 | 400 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 50 | 100 | 150 | 100 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG |
| param.split\_func | 1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 |
| accuracy |  |  | 0.52 |  |  |  |  |  |
| param.num | 300 | 300 | 400 | 400 | 300 | 300 | 400 | 400 |
| param.depth | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| param.splitNum | 5 | 5 | 5 | 5 | 50 | 100 | 150 | 100 |
| param.split | IG | IG | IG | IG | IG | IG | IG | IG |
| param.split\_func | 1 | 4 | 4 | 1 | 1 | 4 | 4 | 1 |
| accuracy |  |  | 0.52 |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| After find the optimal setting here vary vocabulary size | 0.6613  0.6933  13.6633 | 0.7127  0.7667  14.1452 | 0.5843  0.6067  15.8760 | 0.6080  0.6533  17.7364 |  |  |  |  |  |  |
| No of clusters | 128 | 256 | 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