Lab 1 Report

1. Output

Prior:

prior 1: 0.04259472890229834
prior 2: 0.05155736977549028
prior 3: 0.05075871860857219
prior 4: 0.05208980388676901
prior 5: 0.051024935664211554
prior 6: 0.052533498979501284
prior 7: 0.051646108794036735
prior 8: 0.052533498979501284
prior 9: 0.052888455053687104
prior 10: 0.0527109770165942
prior 11: 0.05306593309078002
prior 12: 0.0527109770165942
prior 13: 0.05244475996095483
prior 14: 0.0527109770165942
prior 15: 0.052622237998047744
prior 16: 0.05315467210932647
prior 17: 0.04836276510781791
prior 18: 0.05004880646020055
prior 19: 0.04117490460555506
prior 20: 0.033365870973467035

MLE for train data:

Overall:
mle for train data 0.9913923152009939
Each class:
class 1 accuracy: 1.0
class 2 accuracy: 0.9793459552495697
class 3 accuracy: 0.993006993006993
class 4 accuracy: 0.9880749574105622
class 5 accuracy: 0.9895652173913043
class 6 accuracy: 0.9847972972972973
class 7 accuracy: 0.993127147766323
class 8 accuracy: 0.9915540540540541
class 9 accuracy: 0.9949664429530202
class 10 accuracy: 0.9949494949494949
class 11 accuracy: 0.9899665551839465
class 12 accuracy: 1.0
class 13 accuracy: 0.9898477157360406
class 14 accuracy: 0.9966329966329966
class 15 accuracy: 0.9966273187183811

class 16 accuracy: 0.986644407345576
class 17 accuracy: 0.9963302752293578
class 18 accuracy: 0.9911347517730497
class 19 accuracy: 0.9892241379310345
class 20 accuracy: 0.9787234042553191
Confusion matrix:
480 0 0 0 1 0 0 0 1 0 0 0 0 1 0 2 0 0 1 2
0 569 3 2 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 2 568 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 4 0 580 1 0 0 1 0 0 1 0 1 0 1 1 0 0 0 0
0 1 0 2 569 0 0 0 1 0 0 0 2 0 0 0 0 0 0 0
0 1 1 1 0 583 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 3 0 2 0 1 578 1 1 1 0 0 2 0 0 0 0 0 1 0
0 0 0 0 0 0 1 587 0 0 2 0 0 0 1 0 1 0 0 0
0 1 0 0 1 0 1 2 593 0 0 0 0 0 0 1 0 1 0 0
0 0 0 0 0 0 0 1 0 591 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 2 592 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 0 0 0 0 1 594 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 585 1 0 0 1 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0 1 592 0 0 0 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 591 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 591 0 2 0 4
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 543 0 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 559 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 459 1
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 0 1 368

BE for train data:
Overall:
be for train data 0.9502174105954388
Each class:
class 1 accuracy: 0.98125
class 2 accuracy: 0.9242685025817556
class 3 accuracy: 0.8986013986013986
class 4 accuracy: 0.9369676320272572
class 5 accuracy: 0.96
class 6 accuracy: 0.9341216216216216
class 7 accuracy: 0.8316151202749141
class 8 accuracy: 0.9611486486486487
class 9 accuracy: 0.9748322147651006
class 10 accuracy: 0.9764309764309764
class 11 accuracy: 0.9782608695652174
class 12 accuracy: 0.9814814814814815
class 13 accuracy: 0.937394247038917

class 14 accuracy: 0.9764309764309764
class 15 accuracy: 0.9814502529510961
class 16 accuracy: 0.9816360601001669
class 17 accuracy: 0.9889908256880734
class 18 accuracy: 0.9663120567375887
class 19 accuracy: 0.9633620689655172
class 20 accuracy: 0.8351063829787234
Confusion matrix:
471 0 1 0 1 1 1 1 1 0 1 0 0 0 1 0 0 0 2 18
0 537 10 10 4 16 2 0 0 3 0 2 4 1 2 3 0 2 2 1
0 7 514 4 2 9 0 0 0 0 1 0 1 0 0 0 0 0 0 0
0 16 23 550 5 4 27 2 1 1 2 0 14 0 1 1 0 0 0 0
0 1 0 3 552 2 7 1 1 0 0 0 3 0 0 1 0 0 0 0
0 8 15 4 1 553 1 2 0 1 0 0 0 1 2 0 0 0 1 1
0 3 2 6 1 0 484 5 5 1 0 0 3 0 0 0 0 0 0 0
0 1 0 0 0 0 19 569 1 2 2 0 1 1 1 0 1 1 1 0
0 1 0 0 2 1 1 1 581 0 0 0 0 1 0 0 0 0 0 1
0 0 0 0 0 1 3 1 0 580 0 0 0 0 0 0 0 1 1 0
0 0 0 0 0 0 3 0 0 3 585 0 0 0 0 0 0 0 0 0
0 2 3 2 2 1 8 1 0 0 1 583 3 1 0 0 0 1 3 0
0 1 1 3 0 0 12 1 0 1 1 0 554 2 1 0 1 0 0 0
0 0 1 0 2 0 2 1 2 1 0 1 2 580 1 1 0 2 1 0
0 3 0 1 1 2 1 1 0 0 0 0 2 1 582 0 0 0 0 1
5 1 2 1 1 0 2 1 2 0 1 0 1 4 1 588 2 7 0 23
0 0 0 1 0 1 5 2 2 1 0 2 2 2 0 2 539 0 4 12
1 0 0 1 0 0 1 0 0 0 1 0 0 0 0 2 0 545 2 3
1 0 0 1 1 1 3 2 0 0 3 6 1 0 1 1 2 5 447 2
2 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 314

MLE for test data:
Overall:
mle for test data 0.7593604263824117
Each class:
class 1 accuracy: 0.7641509433962265
class 2 accuracy: 0.6812339331619537
class 3 accuracy: 0.4629156010230179
class 4 accuracy: 0.6709183673469388
class 5 accuracy: 0.6292428198433421
class 6 accuracy: 0.7538461538461538
class 7 accuracy: 0.5968586387434555
class 8 accuracy: 0.8405063291139241
class 9 accuracy: 0.8942065491183879
class 10 accuracy: 0.853904282115869
class 11 accuracy: 0.949874686716792

class 12 accuracy: 0.9164556962025316
class 13 accuracy: 0.6183206106870229
class 14 accuracy: 0.8142493638676844
class 15 accuracy: 0.8724489795918368
class 16 accuracy: 0.8969849246231156
class 17 accuracy: 0.8104395604395604
class 18 accuracy: 0.8936170212765957
class 19 accuracy: 0.6161290322580645
class 20 accuracy: 0.5298804780876494
Confusion matrix:
243 0 1 0 1 0 0 0 0 0 1 0 3 4 1 13 1 7 12 38
0 265 36 10 21 47 11 3 1 0 0 3 21 6 8 2 0 1 1 1
0 9 181 32 11 11 7 1 0 0 1 0 3 0 0 0 0 0 0 0
0 13 49 263 36 4 32 1 0 1 0 1 15 3 1 0 0 1 0 0
0 9 17 20 241 2 19 0 0 1 0 3 5 0 0 0 0 0 0 0
1 30 26 6 2 294 2 0 0 0 0 1 4 0 4 2 0 0 1 0
0 4 4 12 8 1 228 8 2 3 2 1 8 2 0 0 1 1 0 0
0 1 1 2 1 1 16 332 24 0 0 0 11 5 1 0 2 2 1 0
1 1 2 0 1 1 8 21 355 2 2 0 7 3 1 0 1 0 0 1
0 0 0 0 0 0 0 0 0 339 3 0 0 0 0 0 0 1 0 1
0 2 0 1 1 0 2 0 0 14 379 0 0 0 1 0 1 0 0 0
3 17 26 8 8 9 4 2 1 2 1 362 35 2 4 0 4 1 8 2
0 7 4 30 23 3 17 10 3 1 0 2 243 7 4 0 0 0 0 0
3 10 16 2 19 7 11 1 2 8 0 2 15 320 6 3 8 0 7 6
5 17 10 5 6 8 11 5 2 8 0 3 17 10 342 1 2 1 7 9
27 2 5 1 0 0 1 1 0 4 1 1 1 9 1 357 3 8 2 34
1 1 2 0 2 1 5 0 3 3 4 8 4 8 3 0 295 3 54 15
6 0 1 0 0 0 3 0 2 2 0 3 0 6 3 3 6 336 16 4
5 1 9 0 2 1 3 9 2 8 4 4 1 7 11 2 28 13 191 7
23 0 1 0 0 0 2 1 0 1 1 1 0 1 1 15 12 1 10 133

BE for test data:
Overall:
be for test data 0.7905396402398401
Each class:
class 1 accuracy: 0.779874213836478
class 2 accuracy: 0.7583547557840618
class 3 accuracy: 0.5242966751918159
class 4 accuracy: 0.7780612244897959
class 5 accuracy: 0.7336814621409922
class 6 accuracy: 0.7717948717948718
class 7 accuracy: 0.6832460732984293
class 8 accuracy: 0.8886075949367088
class 9 accuracy: 0.9042821158690176

class 10 accuracy: 0.9017632241813602
class 11 accuracy: 0.9523809523809523
class 12 accuracy: 0.9088607594936708
class 13 accuracy: 0.6692111959287532
class 14 accuracy: 0.8396946564885496
class 15 accuracy: 0.8724489795918368
class 16 accuracy: 0.9346733668341709
class 17 accuracy: 0.9093406593406593
class 18 accuracy: 0.8377659574468085
class 19 accuracy: 0.5838709677419355
class 20 accuracy: 0.38247011952191234
Confusion matrix:
248 4 2 0 0 0 0 0 0 2 2 0 3 9 2 9 1 16 7 52
0 295 34 9 10 44 8 2 2 2 0 4 18 8 11 2 0 2 2 4
0 7 205 15 9 11 2 0 0 0 1 1 1 1 0 0 0 0 0 0
0 13 60 305 31 9 43 1 0 1 0 1 23 2 0 1 0 0 0 0
0 10 13 20 281 2 15 0 0 1 0 2 10 0 0 1 0 0 0 0
1 21 32 2 1 301 0 1 0 2 0 1 2 0 0 1 0 0 0 0
0 1 0 5 3 1 261 7 2 3 0 1 1 0 0 0 1 0 0 0
0 2 3 7 9 0 26 351 23 4 1 0 14 5 0 0 2 3 1 0
1 1 2 0 0 1 3 11 359 1 1 1 4 1 0 0 1 1 0 1
0 0 3 0 1 1 1 1 2 358 6 0 0 0 0 0 1 1 1 1
1 0 1 1 0 0 0 0 0 11 380 0 0 0 1 0 1 1 0 0
1 13 13 3 4 10 2 0 0 2 1 359 39 2 1 0 4 3 4 0
2 6 2 23 16 0 9 4 0 3 1 3 263 3 5 1 1 0 0 0
4 2 2 0 6 3 1 0 1 2 2 1 6 330 3 1 3 0 2 3
3 8 5 1 5 3 2 2 0 0 0 1 4 3 342 1 1 1 7 5
31 4 5 0 0 0 3 1 1 1 0 1 3 12 3 372 2 5 1 57
4 0 0 0 4 2 2 5 5 1 2 11 1 6 2 2 331 7 95 19
7 0 0 0 0 0 2 1 0 0 0 0 0 5 1 2 1 315 5 4
5 2 9 1 3 2 2 8 2 3 2 8 0 6 20 2 11 20 181 9
10 0 0 0 0 0 0 0 0 0 0 0 1 0 1 3 3 1 4 96

2. Discussion

a. The difference between P_{MLE} and P_{BE}?

For some classes and some words, P_{MLE} can be zero because no such word in that class.

b. Compare the results of the train data and test data

The results of the train data are better than the test data both for MLE and BE, because the classifier is obtained from the train data.

c. Compare the results of the MLE and BE

The results of BE are better than MLE on the test data, but the MLE are better than the BE on the train data. This is because in the test data, some P_{MLE} can be zero, which influence

the accuracy of the classier.