Assignment 10 Austin Frownfelter

# Problem 1

## Part a

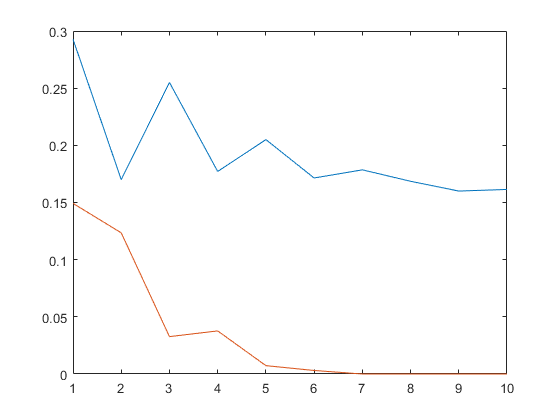
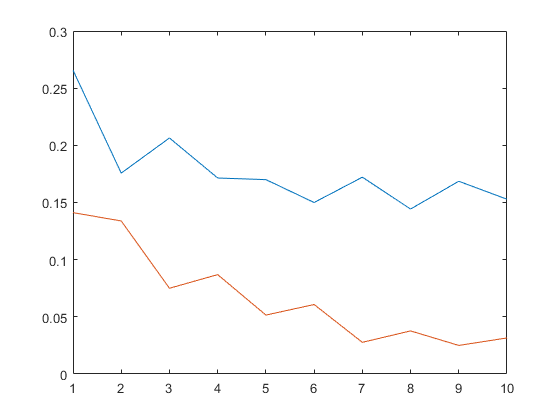
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Test | 0.2007 | 0.1464 | 0.1557 | 0.1471 | 0.1507 | 0.1386 | 0.1493 | 0.1314 | 0.1479 | 0.1321 |
| Train | 0.1135 | 0.0892 | 0.0585 | 0.0573 | 0.0373 | 0.0315 | 0.0281 | 0.0285 | 0.0204 | 0.0188 |
| Boost | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Test | 0.2143 | 0.1421 | 0.1936 | 0.1393 | 0.1593 | 0.1471 | 0.1514 | 0.1371 | 0.1543 | 0.1486 |
| Train | 0.0973 | 0.0950 | 0.0069 | 0.0173 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

# 

Orange is training error, blue is test error; Left is bagging, right is boosting. We can see that the error decreases as we add more models. For boosting, the training error becomes 0, and the training error hovers around the same error as the bagging model.

## Part b

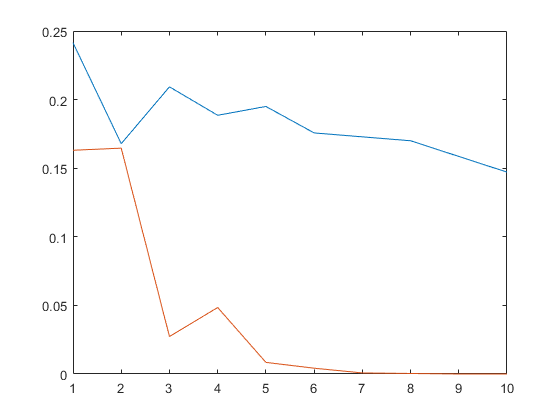
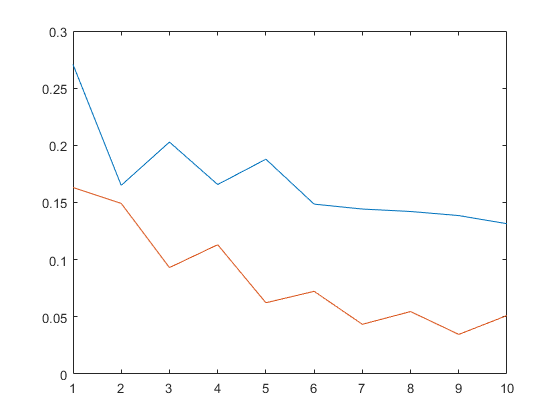
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Test | 0.2657 | 0.1757 | 0.2064 | 0.1714 | 0.1700 | 0.1500 | 0.1721 | 0.1443 | 0.1686 | 0.1529 |
| Train | 0.1412 | 0.1338 | 0.0750 | 0.0869 | 0.0515 | 0.0608 | 0.0277 | 0.0377 | 0.0250 | 0.0315 |
| Boost | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Test | 0.2929 | 0.1700 | 0.2550 | 0.1771 | 0.2050 | 0.1714 | 0.1786 | 0.1686 | 0.1600 | 0.1614 |
| Train | 0.1492 | 0.1235 | 0.0327 | 0.0377 | 0.0073 | 0.0031 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |



Orange is training error, blue is test error; Left is bagging, right is boosting. Like before, the error is decreasing for both as we add more models, and boosting’s training error becomes zero with test errors staying near bagging’s test error.

## Part c

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Bag | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Test | 0.2707 | 0.1650 | 0.2029 | 0.1657 | 0.1879 | 0.1486 | 0.1443 | 0.1421 | 0.1386 |
| Train | 0.1631 | 0.1492 | 0.0931 | 0.1131 | 0.0623 | 0.0723 | 0.0435 | 0.0546 | 0.0346 |
| Boost | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Test | 0.2414 | 0.1679 | 0.2093 | 0.1886 | 0.1950 | 0.1757 | 0.1729 | 0.1700 | 0.1586 |
| Train | 0.1631 | 0.1646 | 0.0273 | 0.0485 | 0.0085 | 0.0042 | 0.0008 | 0.0004 | 0.0000 |



Orange is training error, blue is test error; Left is bagging, right is boosting. Like before, the error is decreasing for both as we add more models, and boosting’s training error becomes zero. Boosting’s test error was not quite as good as bagging’s test error. It is interesting that even a basic model (this 1-decision DT) can perform quite well with a bagging/boosting model, in this case it performed better!

# Problem 2

## Part a

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Score | 0.3192 | 0.2140 | 0.1910 | 0.1892 | 0.1693 | 0.1673 | 0.1650 | 0.1402 | 0.1255 | 0.1212 |
| Dim | 48 | 25 | 21 | 70 | 65 | 40 | 29 | 19 | 57 | 20 |
| Score | 0.0995 | 0.0950 | 0.0858 | 0.0846 | 0.0607 | 0.0579 | 0.0527 | 0.0462 | 0.0461 | 0.0422 |
| Dim | 24 | 30 | 12 | 47 | 61 | 10 | 34 | 27 | 39 | 41 |

## Part b

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Score | 0.7340 | 0.7133 | 0.6887 | 0.6837 | 0.6833 | 0.6730 | 0.6707 | 0.6695 | 0.6661 | 0.6620 |
| Dim | 25 | 48 | 40 | 29 | 21 | 67 | 70 | 11 | 47 | 65 |
| Score | 0.6459 | 0.6432 | 0.6412 | 0.6383 | 0.6315 | 0.6270 | 0.6208 | 0.6174 | 0.6168 | 0.6090 |
| Dim | 12 | 24 | 39 | 6 | 19 | 57 | 20 | 34 | 5 | 14 |

There are 15 shared dimensions between the two results. They aren’t in the same order, but they are pretty close. I would expect this to be the case, since the scores are used to pick the best dimensions, which should be fairly consistent across different measures.