EXOS Performance Evaluation

Egawati Panjei

2022-05-17

Contents

Case

In this report, we generate 30 datasets for Case 1 (all data attributes have strong correlation) and Case 4 (all data attributes have weak correlation). Dataset i for each case is generated using the same seed random number. Hence, they have the same number of attributes in each group (stream). Each attribute j in Case 1 and Case 2 has similar mean and range. However, the mean and range of each attribute in each case is different. Cases 1 and 2 also have the same number of inlier and outlier data points. The difference is in their data attribute correlation.

Case 1

```
df_case1 = read.csv('dbpca/pickles/performance/small_cases/Case1/aggregate_3_1K_Case1_30.csv')
df_case1
```

```
##    experiment precision    recall  f1_score running_time
## 1           1 0.5305556 0.7166667 0.5755556     11.16485
## 2           2 0.5944444 0.6416667 0.5883333     11.22272
## 3           3 0.5805556 0.6444444 0.5798413     11.08926
## 4           4 0.5583333 0.6166667 0.5726190     11.42335
## 5           5 0.6027778 0.6277778 0.5866667     11.15206
## 6           6 0.4805556 0.6638889 0.5407143     11.22820
## 7           7 0.5166667 0.5944444 0.5296825     11.23469
## 8           8 0.5088889 0.5294444 0.5066931     11.15765
## 9           9 0.5527778 0.5750000 0.5343651     11.18522
## 10         10 0.5388889 0.5527778 0.5222222     11.24516
## 11         11 0.5000000 0.5000000 0.4777778     11.10427
## 12         12 0.5444444 0.6361111 0.5644444     11.20386
## 13         13 0.5166667 0.6194444 0.5385714     11.16366
## 14         14 0.5111111 0.6611111 0.5566667     11.06926
## 15         15 0.4583333 0.5083333 0.4738889     11.13464
## 16         16 0.6700000 0.6361111 0.6312698     11.22130
## 17         17 0.6750000 0.6694444 0.6432540     11.15420
## 18         18 0.5972222 0.6222222 0.5859524     11.18812
## 19         19 0.5472222 0.6361111 0.5527778     11.16732
## 20         20 0.6266667 0.6211111 0.5809524     11.13932
## 21         21 0.5722222 0.5750000 0.5419048     11.32427
## 22         22 0.6444444 0.6500000 0.6134921     11.18158
## 23         23 0.6066667 0.6777778 0.6004762     11.17796
## 24         24 0.5966667 0.6555556 0.5860317     11.19333
## 25         25 0.4972222 0.6805556 0.5607143     11.13376
## 26         26 0.5216667 0.6488889 0.5513757     11.09964
## 27         27 0.4861111 0.4916667 0.4794444     11.11297
## 28         28 0.5777778 0.7277778 0.6144444     11.32180
## 29         29 0.5111111 0.4944444 0.4722222     11.16073
## 30         30 0.6277778 0.6083333 0.5996825     11.45709
```

Case 4

```
df_case4 = read.csv('dbpca/pickles/performance/small_cases/Case4/aggregate_3_1K_Case4_30.csv')
df_case4
```

```
##    experiment precision    recall  f1_score running_time
## 1           1 0.5611111 0.7555556 0.6088889     11.10616
## 2           2 0.5194444 0.5750000 0.5229365     11.21000
## 3           3 0.5472222 0.6250000 0.5496825     11.16100
## 4           4 0.5666667 0.6583333 0.5853175     11.24222
## 5           5 0.6250000 0.6277778 0.5892063     11.13956
## 6           6 0.4972222 0.6527778 0.5405556     11.12045
## 7           7 0.5283333 0.6200000 0.5429365     11.29755
## 8           8 0.5000000 0.5777778 0.5169048     11.16045
## 9           9 0.6000000 0.5666667 0.5517460     11.19042
## 10         10 0.6333333 0.6527778 0.6076190     11.26226
## 11         11 0.4277778 0.4055556 0.4022222     11.15099
## 12         12 0.5216667 0.6000000 0.5381746     11.18038
## 13         13 0.4694444 0.5861111 0.5027778     11.19985
## 14         14 0.5111111 0.6611111 0.5601587     11.07977
```

```
## 15         15 0.5111111 0.5333333 0.5111111     11.15410
## 16         16 0.5944444 0.6166667 0.5690476     11.24763
## 17         17 0.6722222 0.6888889 0.6517460     11.15730
## 18         18 0.6055556 0.5722222 0.5522222     11.19341
## 19         19 0.5694444 0.6694444 0.5784921     11.14410
## 20         20 0.6250000 0.6083333 0.5765873     11.14221
## 21         21 0.5744444 0.6155556 0.5597884     11.29272
## 22         22 0.5916667 0.6222222 0.5736508     11.17476
## 23         23 0.6183333 0.7127778 0.6263492     11.22256
## 24         24 0.6222222 0.6777778 0.6112698     11.08572
## 25         25 0.5188889 0.7211111 0.5872222     11.12132
## 26         26 0.5477778 0.7111111 0.5830952     11.21196
## 27         27 0.4972222 0.5027778 0.4905556     11.13760
## 28         28 0.5722222 0.7111111 0.5966667     11.14866
## 29         29 0.5666667 0.5694444 0.5352381     11.12403
## 30         30 0.6472222 0.6055556 0.6035714     11.17501
```

Precision

Variance Test for Precision

```
var.test(df_case1$precision, df_case4$precision)
```

```
##
##  F test to compare two variances
##
## data:  df_case1$precision and df_case4$precision
## F = 0.98496, num df = 29, denom df = 29, p-value = 0.9677
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4688073 2.0694015
## sample estimates:
## ratio of variances
##          0.9849622
```

The var test shows that the true ratio of precision variances of Case 1 and Case 4 is equal to 1 (accept the null hypothesis).

t Test for Precision

```
t.test(df_case1$precision, df_case4$precision, alternative = "greater", var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  df_case1$precision and df_case4$precision
## t = -0.20301, df = 58, p-value = 0.5801
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  -0.027701       Inf
## sample estimates:
## mean of x mean of y
## 0.5584259 0.5614259
```

The t test shows that there is no difference between the precision mean of Case 1 and Case 4. We accept the null hypothesis.

Recall

Variance Test for recall

```
var.test(df_case1$recall, df_case4$recall)
```

```
##
##  F test to compare two variances
##
## data:  df_case1$recall and df_case4$recall
## F = 0.76749, num df = 29, denom df = 29, p-value = 0.4806
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3652992 1.6124979
## sample estimates:
## ratio of variances
##          0.7674922
```

The var test shows that the true ratio of recall variances of Case 1 and Case 4 is equal to 1 (accept the null hypothesis).

t Test for recall

```
t.test(df_case1$recall, df_case4$recall, alternative = "greater", var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  df_case1$recall and df_case4$recall
## t = -0.42003, df = 58, p-value = 0.662
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.03651697         Inf
## sample estimates:
## mean of x mean of y
## 0.6160926 0.6234259
```

The t test shows that there is no difference between the recall mean of Case 1 and Case 4. We accept the null hypothesis.

F1 score

Variance Test for F1 score

```
var.test(df_case1$f1_score, df_case4$f1_score)
```

```
##
##  F test to compare two variances
##
## data:  df_case1$f1_score and df_case4$f1_score
## F = 0.8971, num df = 29, denom df = 29, p-value = 0.772
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4269881 1.8848040
## sample estimates:
## ratio of variances
##          0.8971003
```

The var test shows that the true ratio of f1 score variances of Case 1 and Case 4 is equal to 1 (accept the null hypothesis).

t Test for F1 score

```
t.test(df_case1$f1_score, df_case4$f1_score, alternative = "greater", var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  df_case1$f1_score and df_case4$f1_score
## t = -0.17492, df = 58, p-value = 0.5691
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.02241589         Inf
## sample estimates:
## mean of x mean of y
## 0.5587346 0.5608580
```

The t test shows that there is no difference between the F1 score mean of Case 1 and Case 4. We accept the null hypothesis since p-value is not small enough.