

10727122 廖奕銘

10727132 張任宏

藉由客觀數據
找出影響紅白酒品質之因素

載入DATASET

```
1 rm(list=ls())
2 library(readr)
3 library(arules)
4 library(sigmoid)
5 library(plyr)
6
7 setwd("/Users/brianliao/Documents/資料探勘導論/final")
8 # please change this path with our file dir.
9 winequality_red <- read_delim("winequality-red.csv",
10                               ";", escape_double = FALSE, trim_ws = TRUE)
11 winequality_white <- read_delim("winequality-white.csv",
12                                 ";", escape_double = FALSE, trim_ws = TRUE)
13
14 View(winequality_red)
15 View(winequality_white)
```

初始化及載入LIBRARY

```
1 rm(list=ls())
2 library(readr)
3 library(arules)
4 library(sigmoid)
5 library(plyr)
6
7
8
9
10
11
12
13
14
15
```


設置WORK DIRECTORY及讀入DATASET

```
1  
2  
3  
4  
5  
6  
7 setwd("/Users/brianliao/Documents/資料探勘導論/final")  
8 # please change this path with our file dir.  
9 winequality_red <- read_delim("winequality-red.csv",  
10                               ";", escape_double = FALSE, trim_ws = TRUE)  
11 winequality_white <- read_delim("winequality-white.csv",  
12                                ";", escape_double = FALSE, trim_ws = TRUE)  
13  
14  
15
```

載入DATASET

查看讀入結果

```
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14 View(winequality_red)  
15 View(winequality_white)
```


WINEQUALITY_RED

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
1	7.4	0.700	0.00	1.90	0.076	11	34	0.9978	3.51	0.56	9.4	5
2	7.8	0.880	0.00	2.60	0.098	25	67	0.9968	3.20	0.68	9.8	5
3	7.8	0.760	0.04	2.30	0.092	15	54	0.9970	3.26	0.65	9.8	5
4	11.2	0.280	0.56	1.90	0.075	17	60	0.9980	3.16	0.58	9.8	6
5	7.4	0.700	0.00	1.90	0.076	11	34	0.9978	3.51	0.56	9.4	5
6	7.4	0.660	0.00	1.80	0.075	13	40	0.9978	3.51	0.56	9.4	5
7	7.9	0.600	0.06	1.60	0.069	15	59	0.9964	3.30	0.46	9.4	5
8	7.3	0.650	0.00	1.20	0.065	15	21	0.9946	3.39	0.47	10.0	7
9	7.8	0.580	0.02	2.00	0.073	9	18	0.9968	3.36	0.57	9.5	7
10	7.5	0.500	0.36	6.10	0.071	17	102	0.9978	3.35	0.80	10.5	5
11	6.7	0.580	0.08	1.80	0.097	15	65	0.9959	3.28	0.54	9.2	5
12	7.5	0.500	0.36	6.10	0.071	17	102	0.9978	3.35	0.80	10.5	5
13	5.6	0.615	0.00	1.60	0.089	16	59	0.9943	3.58	0.52	9.9	5
14	7.8	0.610	0.29	1.60	0.114	9	29	0.9974	3.26	1.56	9.1	5
15	8.9	0.620	0.18	3.80	0.176	52	145	0.9986	3.16	0.88	9.2	5
16	8.9	0.620	0.19	3.90	0.170	51	148	0.9986	3.17	0.93	9.2	5

WINEQUALITY_WHITE

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
1	7.0	0.270	0.36	20.70	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
2	6.3	0.300	0.34	1.60	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
3	8.1	0.280	0.40	6.90	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
4	7.2	0.230	0.32	8.50	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
5	7.2	0.230	0.32	8.50	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
6	8.1	0.280	0.40	6.90	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
7	6.2	0.320	0.16	7.00	0.045	30.0	136.0	0.9949	3.18	0.47	9.6	6
8	7.0	0.270	0.36	20.70	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
9	6.3	0.300	0.34	1.60	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
10	8.1	0.220	0.43	1.50	0.044	28.0	129.0	0.9938	3.22	0.45	11.0	6
11	8.1	0.270	0.41	1.45	0.033	11.0	63.0	0.9908	2.99	0.56	12.0	5
12	8.6	0.230	0.40	4.20	0.035	17.0	109.0	0.9947	3.14	0.53	9.7	5
13	7.9	0.180	0.37	1.20	0.040	16.0	75.0	0.9920	3.18	0.63	10.8	5
14	6.6	0.160	0.40	1.50	0.044	48.0	143.0	0.9912	3.54	0.52	12.4	7
15	8.3	0.420	0.62	19.25	0.040	41.0	172.0	1.0002	2.98	0.67	9.7	5
16	6.6	0.170	0.38	1.50	0.032	28.0	112.0	0.9914	3.25	0.55	11.4	7

PREPROCESSING

```
18 whiteCopy <- winequality_white
19 for ( col in colnames(whiteCopy) ) {
20   thisMedian <- median(whiteCopy[[col]])
21   tmp <- as.character( ( whiteCopy[[col]]-thisMedian ) / ( max(whiteCopy[[col]])-thisMedian ) )
22   print( thisMedian )
23   # whiteCopy[[col]] <- tmp
24   process <- c()
25   for ( index in tmp ) {
26     if ( index >= 0 )
27       process <- append( process, "1" )
28     else
29       process <- append( process, "-1" )
30   }
31   whiteCopy[[col]] <- process
32 }
```


將**WINEQUALITY_WHITE**放到**WHITECOPY**，之後以**COL**遍歷**COLNAME**

```
18 whiteCopy <- winequality_white
19 for ( col in colnames(whiteCopy) ) {
20
21
22
23
24
25
26
27
28
29
30
31
32
```

取得各 **ATTRIBUTE** 的中位數，再以各欄位減去中位數後除最大值與中位數的差

```
18
19 for ( col in colnames(whiteCopy) ) {
20   thisMedian <- median(whiteCopy[[col]])
21   tmp <- as.character( ( whiteCopy[[col]]-thisMedian ) / ( max(whiteCopy[[col]])-thisMedian ) )
22   print( thisMedian )
23
24
25
26
27
28
29
30
31
32
```


將正值改成 1，負值改為 -1

```
18
19
20
21
22
23
24 process <- c()
25 for ( index in tmp ) {
26     if ( index >= 0 )
27         process <- append( process, "1" )
28     else
29         process <- append( process, "-1" )
30
31
32
```

APRIORI

以 APRIORI 找 FREQUENT，以 MIN SUMP = 0.33 找 ALCOHOL, QUALITY

```
36 # using apriori to find the frequent set
37 freq <- apriori( whiteCopy, parameter=list(supp=0.33, target="frequent",minlen=2))
38 freq=sort(freq,decreasing=T,by="support")
39 out <- cbind(labels = labels(freq), quality(freq))
40 result1<-out[str_detect(out$labels, "quality"), ]
41 result1<-result1[str_detect(result1$labels, "alcohol"), ]
42 nrow(out)
43 out[c(1:30),]
```


RESULT

WHITE WINE FREQUENT SET (ALCOHOL, QUALITY)

	labels	support	count
254	{alcohol=1,quality=1}	0.4189465	2052
1792	{density=-1,alcohol=1,quality=1}	0.3446305	1688
1167	{chlorides=-1,alcohol=1,quality=1}	0.3066558	1502
1293	{total sulfur dioxide=-1,alcohol=1,quality=1}	0.2878726	1410
1737	{residual sugar=-1,alcohol=1,quality=1}	0.2837893	1390
9032	{residual sugar=-1,density=-1,alcohol=1,quality=1}	0.2739894	1342
6773	{chlorides=-1,density=-1,alcohol=1,quality=1}	0.2672519	1309
7355	{total sulfur dioxide=-1,density=-1,alcohol=1,quality=1}	0.2558187	1253
6	{alcohol=-1,quality=-1}	0.2478563	1214
133	{alcohol=-1,quality=1}	0.2462229	1206
1596	{free sulfur dioxide=-1,alcohol=1,quality=1}	0.2394855	1173
1977	{pH=1,alcohol=1,quality=1}	0.2390772	1171
6353	{chlorides=-1,total sulfur dioxide=-1,alcohol=1,quality=1}	0.2286648	1120
1021	{fixed acidity=-1,alcohol=1,quality=1}	0.2166190	1061
2011	{sulphates=1,alcohol=1,quality=1}	0.2166190	1061
2012	{volatile acidity=1,alcohol=1,quality=1}	0.2166190	1061
6728	{residual sugar=-1,chlorides=-1,alcohol=1,quality=1}	0.2109024	1033
1672	{citric acid=-1,alcohol=1,quality=1}	0.2106982	1032
7311	{residual sugar=-1,total sulfur dioxide=-1,alcohol=1,quality=1}	0.2090649	1024
23950	{chlorides=-1,total sulfur dioxide=-1,density=-1,alcohol=1,quality=1}	0.2088608	1023

RED WINE FREQUENT SET (ALCOHOL, QUALITY)

	labels	support	count
228	{alcohol=1,quality=1}	0.3702314	592
10	{alcohol=-1,quality=-1}	0.3333333	533
1768	{density=-1,alcohol=1,quality=1}	0.2595372	415
1937	{sulphates=1,alcohol=1,quality=1}	0.2545341	407
1276	{volatile acidity=-1,alcohol=1,quality=1}	0.2451532	392
1919	{citric acid=1,alcohol=1,quality=1}	0.2301438	368
415	{volatile acidity=1,alcohol=-1,quality=-1}	0.2295184	367
837	{chlorides=-1,alcohol=1,quality=1}	0.2232645	357
310	{sulphates=-1,alcohol=-1,quality=-1}	0.2213884	354
1580	{total sulfur dioxide=-1,alcohol=1,quality=1}	0.2188868	350
412	{total sulfur dioxide=1,alcohol=-1,quality=-1}	0.2163852	346
1940	{residual sugar=1,alcohol=1,quality=1}	0.2132583	341
418	{chlorides=1,alcohol=-1,quality=-1}	0.2113821	338
410	{density=1,alcohol=-1,quality=-1}	0.2076298	332
1934	{fixed acidity=1,alcohol=1,quality=1}	0.2038774	326
7458	{volatile acidity=-1,citric acid=1,alcohol=1,quality=1}	0.2032520	325
1925	{pH=1,alcohol=1,quality=1}	0.2007505	321
372	{citric acid=-1,alcohol=-1,quality=-1}	0.2001251	320
358	{pH=-1,alcohol=-1,quality=-1}	0.1901188	304
411	{free sulfur dioxide=1,alcohol=-1,quality=-1}	0.1863665	298
1648	{free sulfur dioxide=-1,alcohol=1,quality=1}	0.1863665	298
1868	{free sulfur dioxide=1,alcohol=1,quality=1}	0.1838649	294

RESULT

RED WINE

FREQUENT SET

	labels	support	count
{free sulfur dioxide=1,total sulfur dioxide=1}	0.4071295	651	
free sulfur dioxide=-1,total sulfur dioxide=-1}	0.4015009	642	
{volatile acidity=1,citric acid=-1}	0.3877423	620	
{fixed acidity=1,pH=-1}	0.3864916	618	
{fixed acidity=-1,pH=1}	0.3808630	609	
{volatile acidity=-1,citric acid=1}	0.3808630	609	
{fixed acidity=1,citric acid=1}	0.3796123	607	
{alcohol=1,quality=1}	0.3702314	592	
{fixed acidity=1,density=1}	0.3608505	577	
{fixed acidity=-1,citric acid=-1}	0.3602251	576	
{citric acid=-1,pH=1}	0.3502189	560	
{sulphates=1,quality=1}	0.3502189	560	
{fixed acidity=-1,density=-1}	0.3439650	550	
{citric acid=1,pH=-1}	0.3433396	549	
{density=-1,alcohol=1}	0.3352095	536	
{density=1,alcohol=-1}	0.3339587	534	
{alcohol=-1,quality=-1}	0.3333333	533	
{citric acid=1,sulphates=1}	0.3333333	533	
{chlorides=1,density=1}	0.3308318	529	

RESULT

RED WINE

FREQUENT SET

	labels	support	count
{alcohol=1,quality=1}	0.4189465	2052	
{residual sugar=1,density=1}	0.4187423	2051	
{residual sugar=-1,density=-1}	0.4167007	2041	
{density=-1,alcohol=1}	0.4064924	1991	
{density=1,alcohol=-1}	0.4007758	1963	
{density=-1,quality=1}	0.3926092	1923	
{chlorides=-1,quality=1}	0.3768885	1846	
{free sulfur dioxide=1,total sulfur dioxide=1}	0.3677011	1801	
{total sulfur dioxide=-1,quality=1}	0.3670886	1798	
{chlorides=1,alcohol=-1}	0.3656595	1791	
{total sulfur dioxide=1,density=1}	0.3613720	1770	
{chlorides=-1,alcohol=1}	0.3603512	1765	
{sulphates=1,quality=1}	0.3576970	1752	
{free sulfur dioxide=-1,total sulfur dioxide=-1}	0.3568804	1748	
{pH=1,quality=1}	0.3568804	1748	
{volatile acidity=-1,quality=1}	0.3552470	1740	
{total sulfur dioxide=-1,density=-1}	0.3552470	1740	
{residual sugar=-1,quality=1}	0.3542262	1735	
{chlorides=1,density=1}	0.3527971	1728	
{density=-1,alcohol=1,quality=1}	0.3446305	1688	
{total sulfur dioxide=-1,alcohol=1}	0.3436096	1683	
{total sulfur dioxide=1,alcohol=-1}	0.3436096	1683	
{residual sugar=-1,alcohol=1}	0.3436096	1683	
{residual sugar=1,total sulfur dioxide=1}	0.3419763	1675	
{chlorides=-1,density=-1}	0.3413638	1672	
{residual sugar=1,alcohol=-1}	0.3395263	1663	
{fixed acidity=1,quality=1}	0.3380972	1656	
{chlorides=1,total sulfur dioxide=1}	0.3364639	1648	
{fixed acidity=1,pH=-1}	0.3344222	1638	
{citric acid=1,quality=1}	0.3344222	1638	

THANKS FOR LISTENING