Introduction

The data examines the absence or presence of 36 clinical criteria in a sample of 464 patients. In the data, 0 indicates the absence of a given criteria whilst 1 indicates its' presence. There are a total of 16704 instances in the data (464 patients X 36 criteria). A small number of these (44) were coded with the value 'NA', and have been converted into 0 for ease of analysis.

The 36 clinical criteria are used to categorise the patients into three categories, namely, "Nociceptive" (256 patients or 55% of those sampled), "Peripheral Neuropathic" (102 patients or 22% of those sampled) and "Central Neuropathic" (106 patients or 23% of those sampled).

The prevalence of the clinical criteria ranges from a minimum of 9% for X32 (positive findings of allodynia) – meaning that 9% of those sampled suffer from this criteria, through to a maximum of 77% for X5 (intermittent sharp or constant dull ache). There are a significant number of criteria which affect more than 70% of the sample (as in **table 1** below), along with several others which are prevalent in less than 20% of the patients (**table 2**). Both extremes, however, are marked by a number of 'borderline' cases which lie within 2% of the chosen boundaries.

Patients in the sample can suffer from zero to any number of the clinical criteria. The minimum number of criteria present in the sample is 5 (this applies to 1 patient) and the maximum number of criteria present is 25 (which also applies to 1 patient). The distribution of the criteria amongst the patients sampled is shown in **figure 1** below. It can be seen that 52 of the patients sampled have 9 of the criteria, and that the median number of criteria is 12.

Code	Frequency (%)	Frequency (# Patients)
X2	72	334
X5	77	356
X11	76	353
X26	75	349

Table 1 – Clinical criteria with frequency >70%

Criteria Code	Frequency (%)	Frequency (# Patients)
X1	15	70
X10	16	73
X28	19	90
X31	18	85
X32	9	42

Table 2 – Clinical criteria with frequency <20%

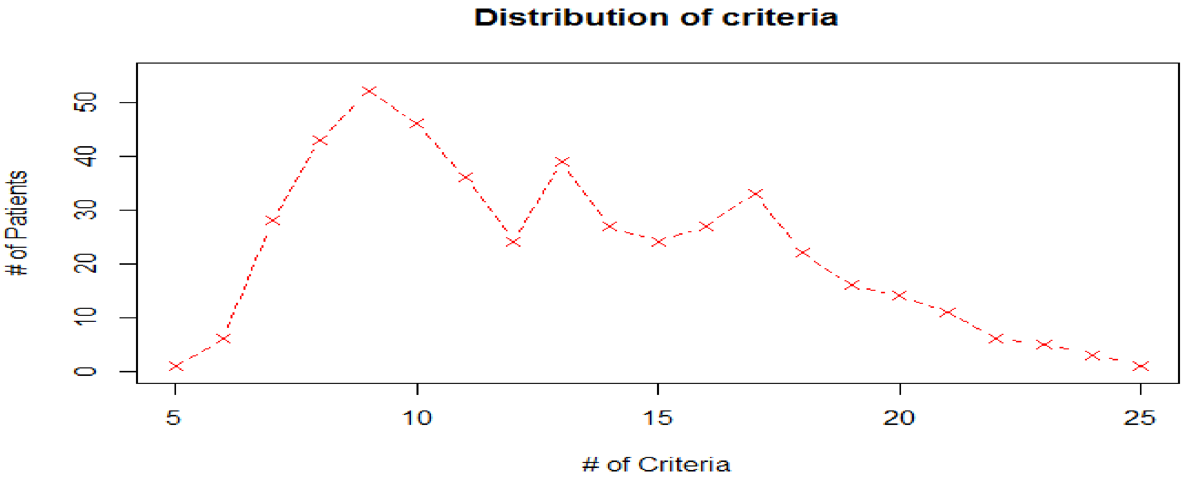


Figure 1 – Distribution of criteria amongst patients

Question 1 – Patterns in the presence / absence clinical criteria for lower back pain

To identify patterns within the 36 clinical criteria, the *a priori* algorithm was used to complete an association rule analysis of the data. The minimum support threshold (showing the co-occurrence of two criteria) was set to 9% (or 0.09) as the least frequent criterion (x32) occurs in 9.05% of those surveyed. A threshold greater than this, would result in this rule not being taken into account. A number of confidence thresholds were tried out. It was noticed that the choice of confidence threshold had a significant effect on the number of rules returned (as shown in **table 3** below). Eventually a confidence threshold of 90% (or 0.9) was selected, meaning that the consequent (rhs) of any rule has probability of at least 0.9 given the antecedent (lhs). The number of rules found, using these settings came to a (manageable) 29.

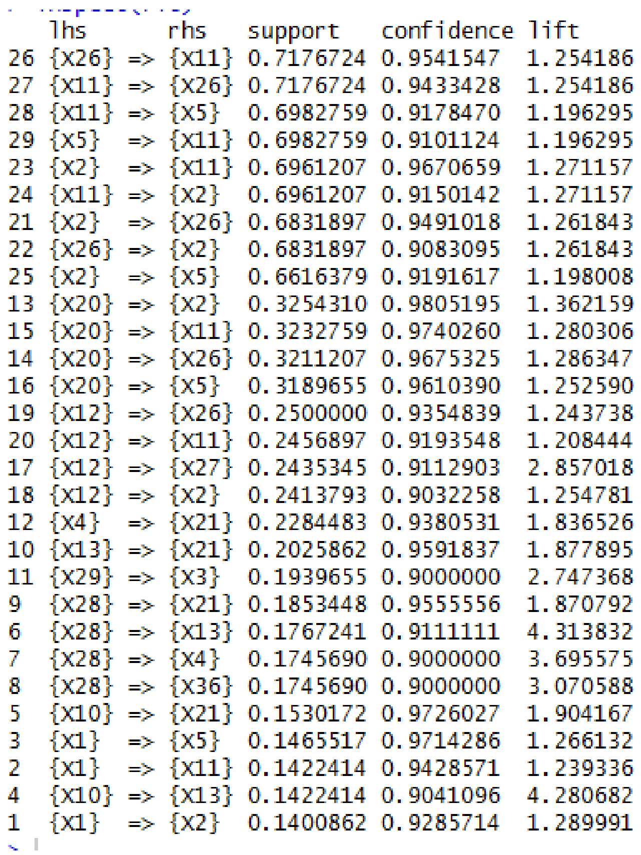
Confidence Threshold	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
# of rules (len=2)	401	291	188	112	29
# of rules (len=3)	3697	3178	2530	1647	708

Table 3 – Relationship between choice of confidence level and number of rules found (*support = 0.09)*

The following twenty nine rules were found (**table 4**) and are sorted using the support measure of quality on the left and the lift measure of quality on the right.

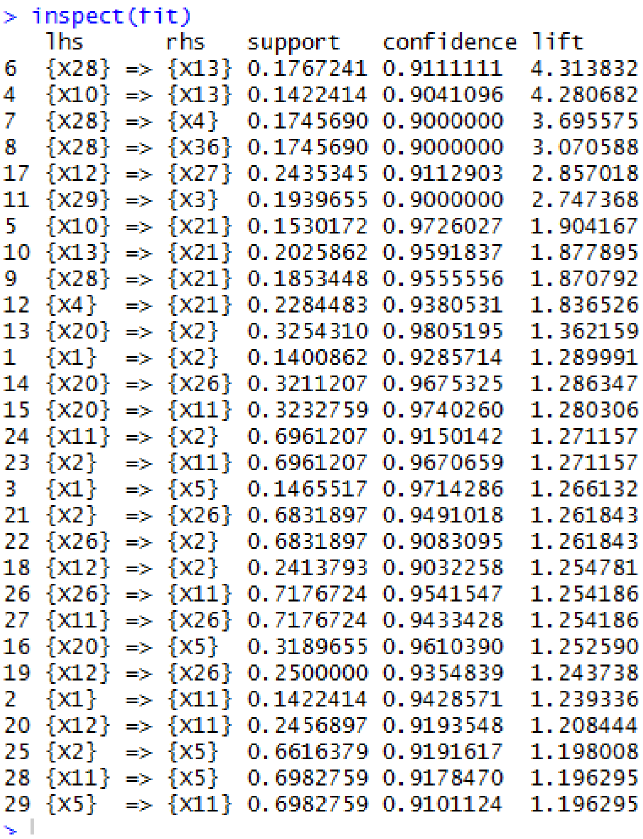
Table 4 – The twenty nine association rules in the lower back pain data

Sorted by support (high -> low)



```
   lhs      rhs    support   confidence lift
26 {x26} => {x11} 0.7176724 0.9541547  1.254186
27 {x11} => {x26} 0.7176724 0.9433428  1.254186
28 {x11} => {x5}  0.6982759 0.9178470  1.196295
29 {x5}  => {x11} 0.6982759 0.9101124  1.196295
23 {x2}  => {x11} 0.6961207 0.9670659  1.271157
24 {x11} => {x2}  0.6961207 0.9150142  1.271157
21 {x2}  => {x26} 0.6831897 0.9491018  1.261843
22 {x26} => {x2}  0.6831897 0.9083095  1.261843
25 {x2}  => {x5}  0.6616379 0.9191617  1.198008
13 {x20} => {x2}  0.3254310 0.9805195  1.362159
15 {x20} => {x11} 0.3232759 0.9740260  1.280306
14 {x20} => {x26} 0.3211207 0.9675325  1.286347
16 {x20} => {x5}  0.3189655 0.9610390  1.252590
19 {x12} => {x26} 0.2500000 0.9354839  1.243738
20 {x12} => {x11} 0.2456897 0.9193548  1.208444
17 {x12} => {x27} 0.2435345 0.9112903  2.857018
18 {x12} => {x2}  0.2413793 0.9032258  1.254781
12 {x4}  => {x21} 0.2284483 0.9380531  1.836526
10 {x13} => {x21} 0.2025862 0.9591837  1.877895
11 {x29} => {x3}  0.1939655 0.9000000  2.747368
9  {x28} => {x21} 0.1853448 0.9555556  1.870792
6  {x28} => {x13} 0.1767241 0.9111111  4.313832
7  {x28} => {x4}  0.1745690 0.9000000  3.695575
8  {x28} => {x36} 0.1745690 0.9000000  3.070588
5  {x10} => {x21} 0.1530172 0.9726027  1.904167
3  {x1}  => {x5}  0.1465517 0.9714286  1.266132
2  {x1}  => {x11} 0.1422414 0.9428571  1.239336
4  {x10} => {x13} 0.1422414 0.9041096  4.280682
1  {x1}  => {x2}  0.1400862 0.9285714  1.289991
>
```

Sorted by lift (high -> low)



```
> inspect(fit)
   lhs      rhs    support   confidence lift
6  {x28} => {x13} 0.1767241 0.9111111  4.313832
4  {x10} => {x13} 0.1422414 0.9041096  4.280682
7  {x28} => {x4}  0.1745690 0.9000000  3.695575
8  {x28} => {x36} 0.1745690 0.9000000  3.070588
17 {x12} => {x27} 0.2435345 0.9112903  2.857018
11 {x29} => {x3}  0.1939655 0.9000000  2.747368
5  {x10} => {x21} 0.1530172 0.9726027  1.904167
10 {x13} => {x21} 0.2025862 0.9591837  1.877895
9  {x28} => {x21} 0.1853448 0.9555556  1.870792
12 {x4}  => {x21} 0.2284483 0.9380531  1.836526
13 {x20} => {x2}  0.3254310 0.9805195  1.362159
1  {x1}  => {x2}  0.1400862 0.9285714  1.289991
14 {x20} => {x26} 0.3211207 0.9675325  1.286347
15 {x20} => {x11} 0.3232759 0.9740260  1.280306
24 {x11} => {x2}  0.6961207 0.9150142  1.271157
23 {x2}  => {x11} 0.6961207 0.9670659  1.271157
3  {x1}  => {x5}  0.1465517 0.9714286  1.266132
21 {x2}  => {x26} 0.6831897 0.9491018  1.261843
22 {x26} => {x2}  0.6831897 0.9083095  1.261843
18 {x12} => {x2}  0.2413793 0.9032258  1.254781
26 {x26} => {x11} 0.7176724 0.9541547  1.254186
27 {x11} => {x26} 0.7176724 0.9433428  1.254186
16 {x20} => {x5}  0.3189655 0.9610390  1.252590
19 {x12} => {x26} 0.2500000 0.9354839  1.243738
2  {x1}  => {x11} 0.1422414 0.9428571  1.239336
20 {x12} => {x11} 0.2456897 0.9193548  1.208444
25 {x2}  => {x5}  0.6616379 0.9191617  1.198008
28 {x11} => {x5}  0.6982759 0.9178470  1.196295
29 {x5}  => {x11} 0.6982759 0.9101124  1.196295
>
```

The rule with the highest support is {Consistent, proportionate pain reproduction on mechanical testing} -> {Mechanical nature to aggs + eases} (x26->x11). The support of this rule tells us that 71.7% of the respondents have this combination of clinical criteria. The confidence value tells us that amongst all respondents diagnosed with x26, 95.4% also have x11. The lift value tells us that the two problems co-occur 1.25 times more than if these two criteria were independent. The relationship between x26 and x11 is equally strong in both directions (in that those with x11 also tend to have x26 to an equal extent to the other way around) – and there is only a small difference in confidence levels between x26->x11 and x11->x26. A similar bi-directional relationship is also evident between x11 and x5, x2 and x11, and x2 and x26, as summarised in **table 5**. These relationships are characterised by almost equal confidence levels in both directions along with (high) support levels in the approximate range 0.6 to 0.7, and with (modest) lift values in the approximate range 1.2 to 1.3. It is also noticeable that x11 figures in three of the four pairs, whilst x26 appears in two of the four pairs, confirming our initial view in the introduction.

Relationship	Support	Confidence	Lift
x26 -> x11	0.71	0.95	1.25
x11 -> x26	0.71	0.94	1.25
x11 -> x5	0.69	0.91	1.19
x5 -> x11	0.69	0.91	1.19
x2 -> x11	0.69	0.96	1.27
x11 -> x2	0.69	0.91	1.27
x2 -> x26	0.63	0.94	1.26
x26 -> x2	0.63	0.90	1.26

Table 5 – Clinical criteria with strong bi-directional relationships

The rule with the highest lift is {Disproportionate, non-mechanical pattern of pain provocation on mechanical testing} -> {Disproportionate, non-mechanical pattern to aggs + eases} (x28->x13), telling us that these criteria co-occur 4.31 more times than if the two were independent. A similar relationship is also evident between {Widespread, non-anatomical distribution} -> {Disproportionate, non-mechanical pattern to aggs + eases} (x10->x13), with the criteria co-occurring 4.28 more times than if the two were independent.

A number of criteria figure prominently as antecedents (on the lhs), when sorted by support. This is particularly the case with x12, x20 and x28. Similarly, x11 and x21 occur more frequently than the others as consequents (on the rhs), again, when the rules are sorted by support.

All of this can be represented graphically, as in the two plots below.

Figure 2 – Scatter plot of the 29 rules

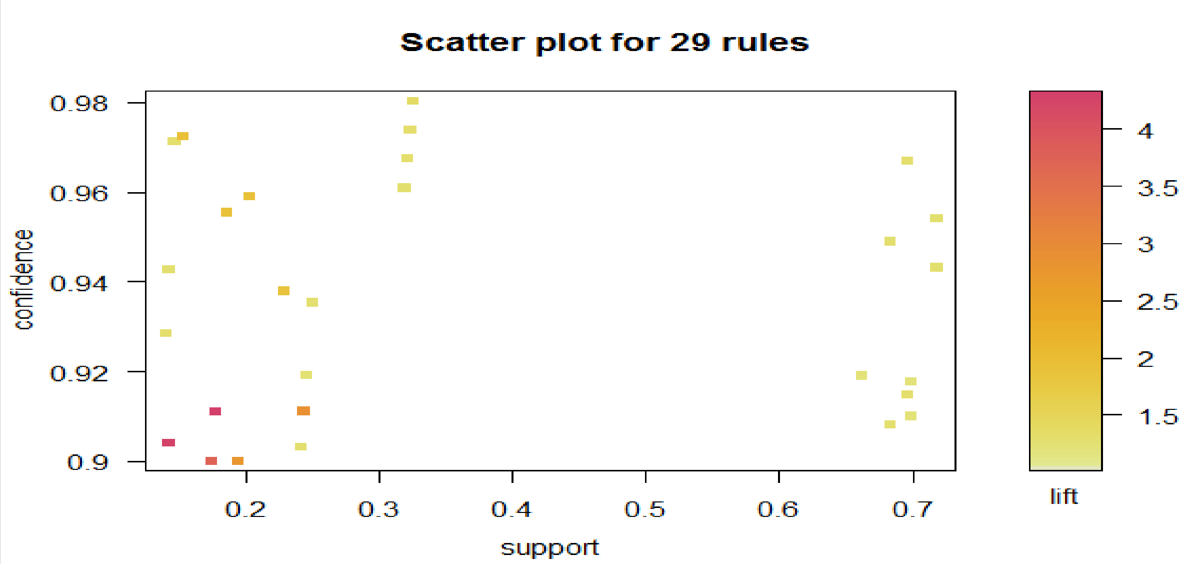
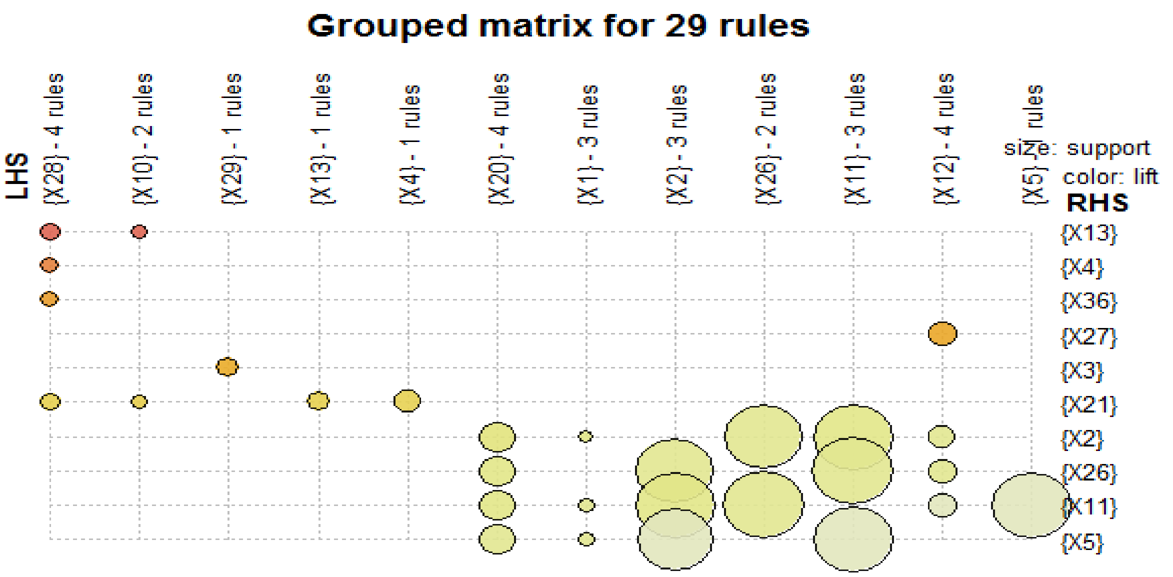


Figure 3 – Grouped Matrix for the 29 rules



The two figures tell complementary stories about the data, and corroborate our initial findings. **Figure 2** examines the relative relationship between confidence, support and lift. It can be noted that five of the rules have high support values (~0.7) combined with high levels of confidence (~0.92), and reasonably high lift (~1.5 – 2). A further group of four rules maintain high levels of support and confidence with higher levels of lift (>2.5) at the top right hand side of the figure. There is a further group of four rules with lower levels of support (0.32) combined with high levels of confidence and high levels of lift (>0.35). The majority of the rules (15), however, are clustered with levels of support between 0.09 and 0.32, confidence levels between 0.9 and 0.98, and lift levels over a wide range from 1 to 4. The importance of both x20 and x28 as antecedents is noticeable in **figure 3**. Each of these criteria are on the lhs for 4 rules. The rules with x28 are characterised by low levels of support and high levels of lift, whilst those for x20 have higher levels of support and lower levels of lift. The high frequency of rules citing x5, x11 and x21 as consequents is also evident. The four bi-directional pairings, discussed above, can also be traced.

In summary, there are strong bi-directional relationships between four pairs of clinical criteria, shaped mainly around x11 and x26. There are also a number of criteria which figure prominently as antecedents (x20, x28) or consequents (x11 again, x21).

Question 2 – Patient groups with similar presence / absence clinical criteria

A naïve, eyeball, analysis of the data proved to be somewhat difficult, and it was not easy to see how the three assignment categories had been derived. Instead a number of k-mean clustering analyses were carried out. The dataset was converted to matrix format and the 'assigned.labels' column was temporarily removed. Initially, two clusters were defined. This was based on 20 random starts to reduce convergence to a local minimum. It resulted in one cluster of 103 patients and another of 361. There were large disparities within the cluster means for each of the 36 clinical criteria. The between sum of squares / total sum of squares came to 27.3%.

By plotting the within cluster sum of squares values against K, there is a significant fall in the former between 2 and 3 clusters and a smaller drop between 3 and 4 clusters - suggesting that 3 is an appropriate number of clusters. The between sum of squares / total sum of squares figure, increases from 27.3% for 2 clusters to 39.6% for 3 clusters, and 43% for 4 clusters.

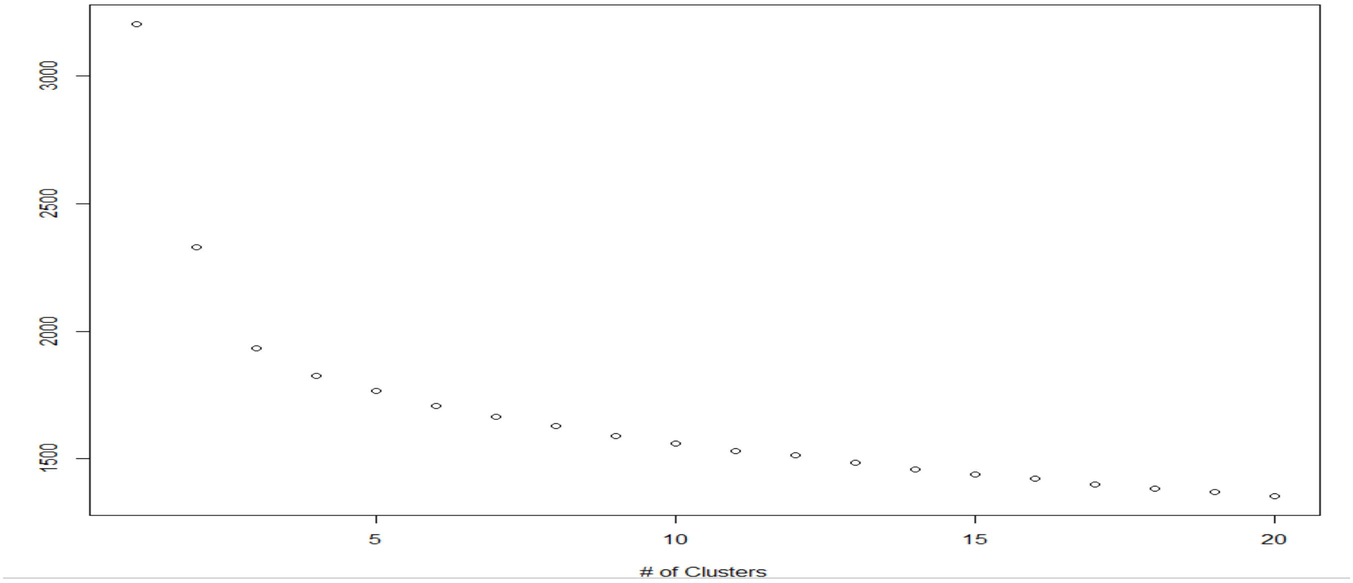
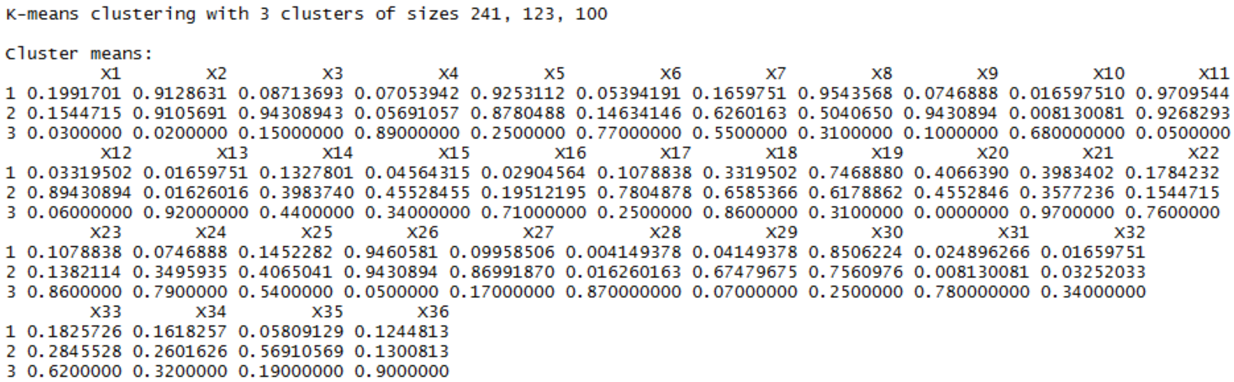


Figure 4 – Within cluster sum of squares for each value of K up to 20

With 3 clusters, a number of trends become evident. The first, and largest, cluster of 241 patients is shaped by the large means (>0.9) of clinical criteria such as X2, X5, X8, X11 and X26. What is interesting is that many of these also have approximately similar means in cluster 2 (123 patients) – this is especially the case with X2, X11 and X26. Criteria X3 and X9 also have high means in this second cluster. The third cluster is the smallest (with 100 patients) and appears to be more distinct (ie less overlap of criteria) than the other two. Clinical criteria with mean > 0.9 include X4, X13, X21, X28 and X36.



```
K-means clustering with 3 clusters of sizes 241, 123, 100

Cluster means:
         x1        x2         x3         x4        x5         x6        x7        x8        x9         x10       x11
1 0.1991701 0.9128631 0.08713693 0.07053942 0.9253112 0.05394191 0.1659751 0.9543568 0.0746888 0.016597510 0.9709544
2 0.1544715 0.9105691 0.94308943 0.05691057 0.8780488 0.14634146 0.6260163 0.5040650 0.9430894 0.008130081 0.9268293
3 0.0300000 0.0200000 0.15000000 0.89000000 0.2500000 0.77000000 0.5500000 0.3100000 0.1000000 0.680000000 0.0500000
         x12        x13       x14        x15        x16       x17       x18       x19       x20       x21       x22
1 0.03319502 0.01659751 0.1327801 0.04564315 0.02904564 0.1078838 0.3319502 0.7468880 0.4066390 0.3983402 0.1784232
2 0.89430894 0.01626016 0.3983740 0.45528455 0.19512195 0.7804878 0.6585366 0.6178862 0.4552846 0.3577236 0.1544715
3 0.06000000 0.92000000 0.4400000 0.34000000 0.71000000 0.2500000 0.8600000 0.3100000 0.0000000 0.9700000 0.7600000
        x23       x24       x25       x26        x27         x28        x29       x30         x31        x32
1 0.1078838 0.0746888 0.1452282 0.9460581 0.09958506 0.004149378 0.04149378 0.8506224 0.024896266 0.01659751
2 0.1382114 0.3495935 0.4065041 0.9430894 0.86991870 0.016260163 0.67479675 0.7560976 0.008130081 0.03252033
3 0.8600000 0.7900000 0.5400000 0.0500000 0.17000000 0.870000000 0.07000000 0.2500000 0.780000000 0.34000000
        x33       x34        x35       x36
1 0.1825726 0.1618257 0.05809129 0.1244813
2 0.2845528 0.2601626 0.56910569 0.1300813
3 0.6200000 0.3200000 0.19000000 0.9000000
```

Figure 5 – Cluster means for each of the 36 Clinical Criteria (k=3, nstarts=20)

The silhouette plot of the data, based on the squared Euclidean distance; shows that the largest cluster, (number 1), has a barely average silhouette (>0.5). Cluster 2 (123 patients) is the most interesting as the average silhouette of 0.20 is significantly below the average. There is a long tail where observations are similar to observations from another cluster. Moreover, four of the observations have negative silhouette, indicating that they should really be in another cluster. All of this indicates weaker clustering than in the previous case. Cluster 3, the smallest with 100 patients, has a slightly below average silhouette of 0.45, and one observation with negative silhouette.

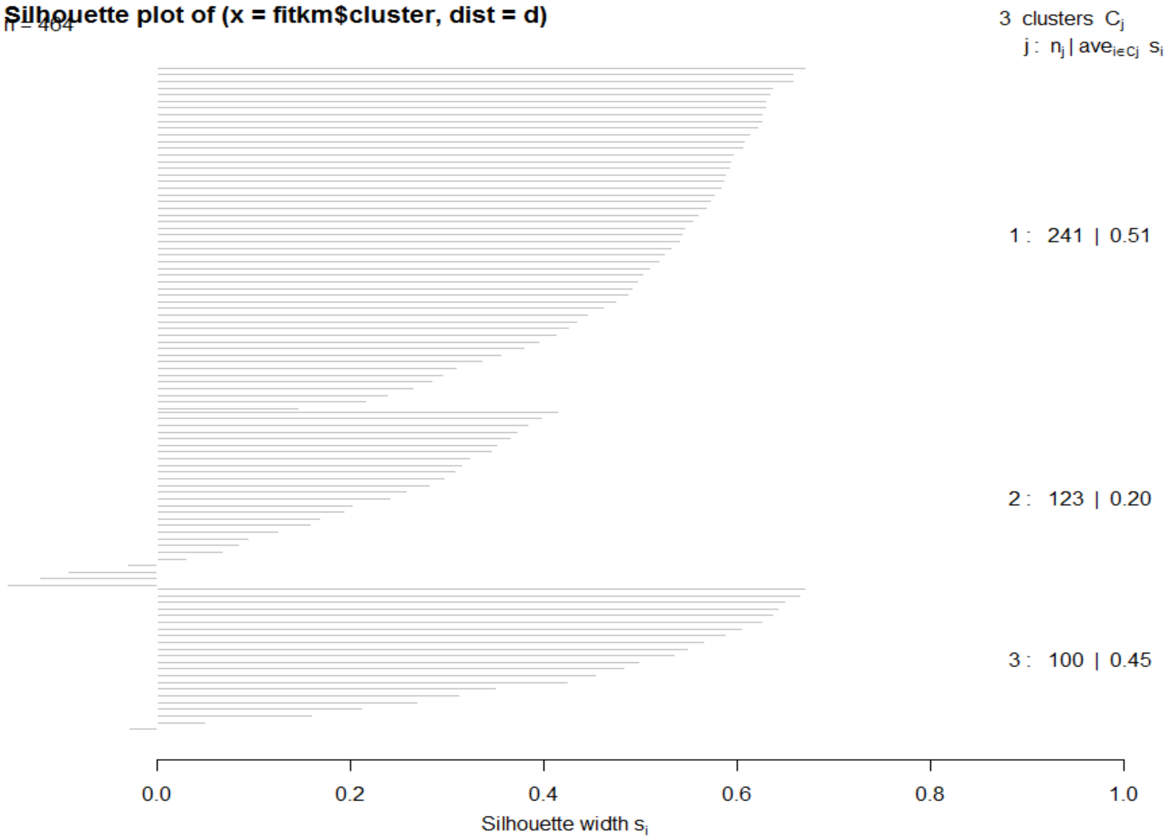
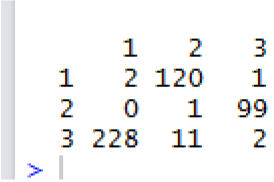


Figure 6 – Silhouette Plot of the K-means solution (k=3, nstarts=20)

The data was also analysed using the k-medoids algorithm with binary distance, and k=3. The results of the two approaches are shown in **figure 7**. From this we get a similar breakdown of clusters. It is noticeable, however, that cluster one is slightly smaller than before (243 observations to 230) and that cluster two has increased slightly (123 observations to 132). Finally, there is a Rand Index of 0.947 suggesting a high degree of agreement between the two algorithms. The Rand Index, corrected for agreement by chance, ('crand' in the e1071 package in r) is slightly lower at 0.888.



```
      1   2   3
  1   2 120   1
  2   0   1  99
  3 228  11   2
> 
```

Figure 7 – Silhouette Plot of the K-means solution (k=3, nstarts=20)

Therefore, there is broad agreement between the approximate cluster sizes, whether observed empirically, through k-means clustering, or through k-medoids clustering. Initial examination showed three patient clusters of 256, 102 and 106 respectively. The clustering algorithm estimated the cluster sizes at 241, 123 and 100 respectively, whilst the k-medoids algorithm placed sized them as 230, 132 and 102. Across the 3 approaches, the biggest variations lie with the first and second clusters, suggesting that a number of borderline observations could belong to either cluster one or cluster two, depending on how the analysis is carried out. This is corroborated by the findings of the silhouette plot, and in particular by the weak profile of cluster two within it.

Question 3 – Use of presence / absence clinical criteria to accurately predict clinical pain types

In the data, the presence and / or absence of the criteria are used to predict the 3 clinical pain types. There are several ways in which this can be done, and each of the methods in question are likely to offer slightly different results. The aim, therefore, is to find broad agreement across a number of different classification methods. One elementary approach, is to use a ***classification tree*** to examine the data and to assess how closely this fits the original derivation of the three pain types. When we do this we notice that the tree uses classifiers X13, X12 and X9 in order to organise the data and predict the pain types, with the data ultimately split into four nodes (**figure 8**). The predictive model then analyses which of the tree pain types each observation is likely to be assigned to, and then compares its' performance with the original data (**figure 9**). We can note that the predicted observations match the actual ones about 90% of the time, with the predicted sizes of the 3 pain types being slightly different than in the original data. A number of observations are also classified under different pain types. Although useful, we need to be careful with this approach as the same data is used both to set up and test the model. Its' predictive accuracy may be over-estimated, in other words.

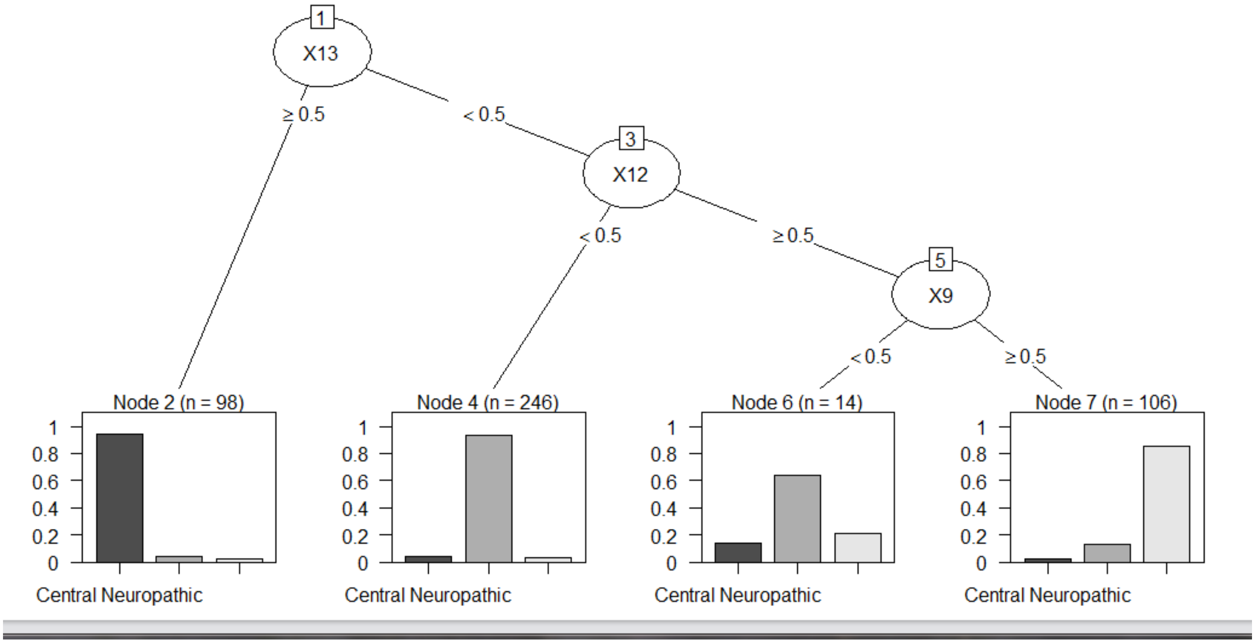
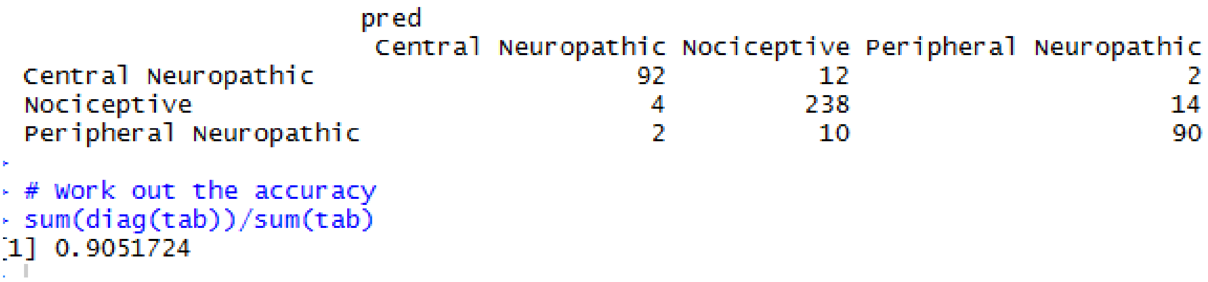


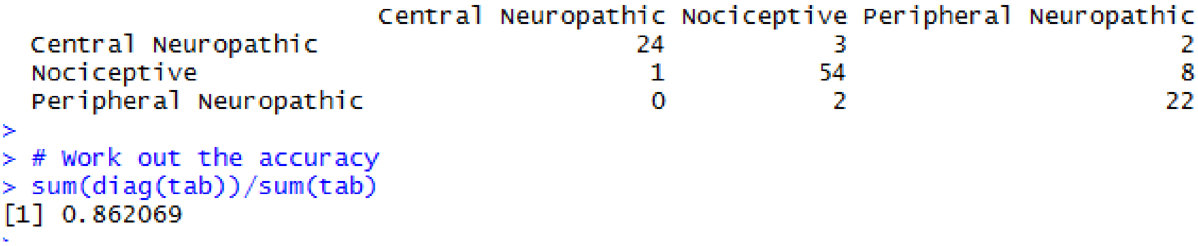
Figure 8 – Initial classification tree of the dataset



```
                        pred
                         Central Neuropathic Nociceptive Peripheral Neuropathic
 Central Neuropathic                      92          12                      2
 Nociceptive                               4         238                     14
 Peripheral Neuropathic                    2          10                     90
>
> # work out the accuracy
> sum(diag(tab))/sum(tab)
[1] 0.9051724
```

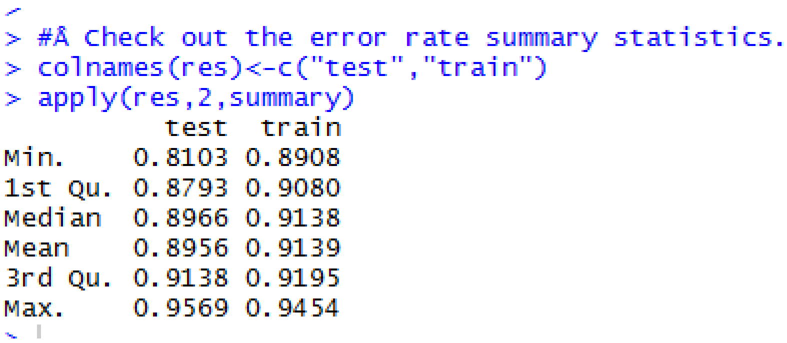
Figure 9 – Predicted vs Actual observations of the initial classification tree

An alternative approach is to again use a classification tree, but this time to ***segment the data into two categories*** (**figure 10**). We can use one of the categories (say, 75% of the data) to set up the model, before then using the remaining 25% of the data to test it. This provides a more realistic test environment for the model. When we do this, it is noticeable that the agreement between the actual and predicted classification values of the model falls to around 86%. Whilst there are few differences between the actual and test versions in classifying 'Central Neuropathic' or 'Nociceptive' pain types; the model struggles somewhat with the 'Peripheral Neuropathic' category – something which concurs with our earlier analysis. If we repeat the process using 100 simulated values (**figure 11**), the gap in predictive accuracy between the test and training data falls to about 2% (89% vs 91%).



```
                            Central Neuropathic Nociceptive Peripheral Neuropathic
  Central Neuropathic                        24           3                      2
  Nociceptive                                 1          54                      8
  Peripheral Neuropathic                      0           2                     22
>
> # Work out the accuracy
> sum(diag(tab))/sum(tab)
[1] 0.862069
```

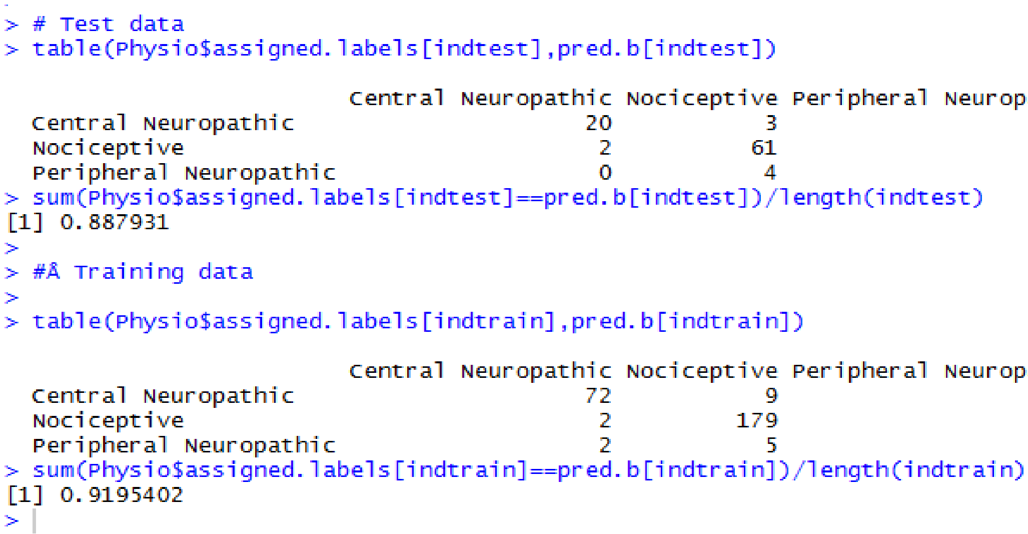
Figure 10 – Test observations with for a 25% segment of the data.



```
> #Â Check out the error rate summary statistics.
> colnames(res)<-c("test","train")
> apply(res,2,summary)
          test  train
Min.    0.8103 0.8908
1st Qu. 0.8793 0.9080
Median  0.8966 0.9138
Mean    0.8956 0.9139
3rd Qu. 0.9138 0.9195
Max.    0.9569 0.9454
```

Figure 11 – Test vs training data with 100 simulated values

A final, more robust, means of assessing the predictive accuracy of the model is offered by both ***bagging*** and ***Random Forests***. The former is a short name for bootstrap aggregating. It works by taking a bootstrap sample of the data, and then fitting a classifier (in this case a classification tree) to that sample data. The classifier is then used on the non-bootstrap sample in order to test its' accuracy (to get an 'out of bag' prediction). This process is then repeated multiple times, with the findings from each of the iterations aggregated to produce a final answer. In other words, it uses multiple classification trees (instead of just one) in order to derive a more diverse, but also more accurate solution. As before, there is a 75%/25% split between training and test data. The results (**figure 12**) concur with the earlier examinations, with a slightly lower accuracy rate for the test data (88% vs 91%). Again, the model's reduced accuracy in classifying the 'Peripheral Neuropathic' group (relative to the other two) is evident.



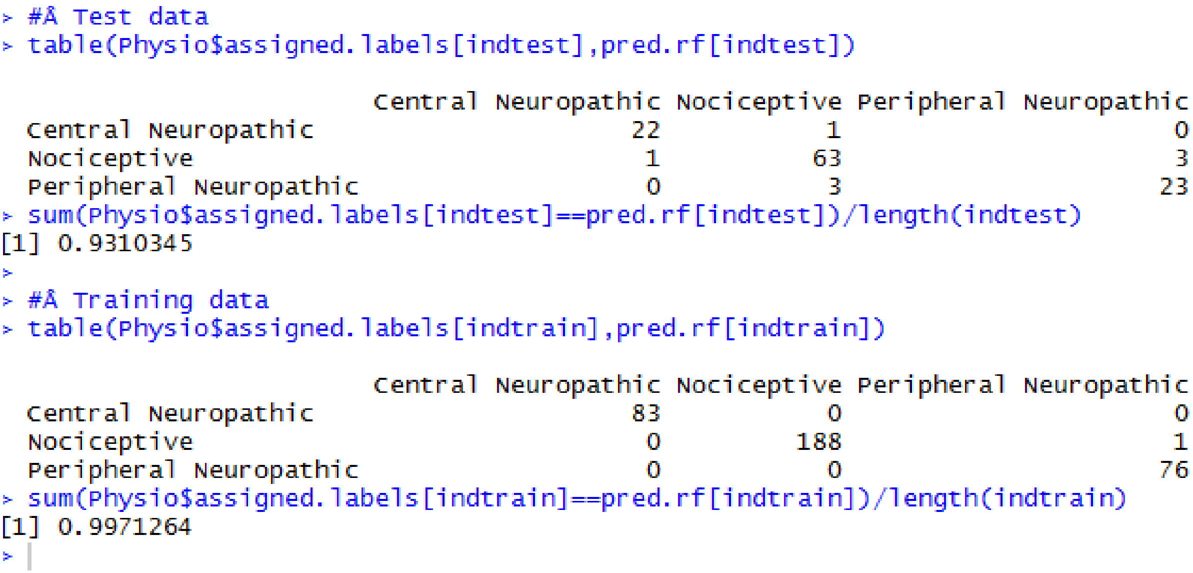
```
> # Test data
> table(Physio$assigned.labels[indtest],pred.b[indtest])

                          Central Neuropathic Nociceptive Peripheral Neuropathic
  Central Neuropathic                      20           3                      0
  Nociceptive                               2          61                      4
  Peripheral Neuropathic                    0           4                     22
> sum(Physio$assigned.labels[indtest]==pred.b[indtest])/length(indtest)
[1] 0.887931
>
> #Â Training data
>
> table(Physio$assigned.labels[indtrain],pred.b[indtrain])

                          Central Neuropathic Nociceptive Peripheral Neuropathic
  Central Neuropathic                      72           9                      2
  Nociceptive                               2         179                      8
  Peripheral Neuropathic                    2           5                     69
> sum(Physio$assigned.labels[indtrain]==pred.b[indtrain])/length(indtrain)
[1] 0.9195402
> |
```

Figure 12 – Test vs training data (Bagging)

Random Forest also uses bootstrapping and multiple classification trees. This time, however, a different random subset of the variables is used at each split of the tree. This makes the approach even more diverse than bagging. When applied to the dataset, the accuracy of this approach is slightly higher (93% for test data / 94% for training data) than before (**figure 13**). It is noticeable that the error rate for the 'Peripheral Neuropathic' group is considerably lower than before, whilst the training classification for the other two groups reaches 100%



```
> #Â Test data
> table(Physio$assigned.labels[indtest],pred.rf[indtest])
                        
                         Central Neuropathic Nociceptive Peripheral Neuropathic
  Central Neuropathic                     22           1                      0
  Nociceptive                              1          63                      3
  Peripheral Neuropathic                   0           3                     23
> sum(Physio$assigned.labels[indtest]==pred.rf[indtest])/length(indtest)
[1] 0.9310345
> 
> #Â Training data
> table(Physio$assigned.labels[indtrain],pred.rf[indtrain])
                        
                         Central Neuropathic Nociceptive Peripheral Neuropathic
  Central Neuropathic                     83           0                      0
  Nociceptive                              0         188                      1
  Peripheral Neuropathic                   0           0                     76
> sum(Physio$assigned.labels[indtrain]==pred.rf[indtrain])/length(indtrain)
[1] 0.9971264
> 
```

Figure 13 – Test vs training data (Random Forest)

In short, each of the methods of assessing classifier accuracy produces slightly varying responses, with results ranging from around 85% (single classification trees) through to 99% (Random Forests). Although this indicates broad agreement, it is interesting how some of the earlier, simpler approaches struggle with classification of the 'Peripheral Neuropathic' group.

Conclusion

This assignment has moved from an initial examination of the dataset, through to an attempt to find relationships between the 36 clinical criteria, towards simple clustering of the data, and then finally to an assessment of the predictive performance of the classifier. At each stage, a number of data mining techniques have been used ranging from association rules analysis, k means clustering, and various methods of assessing classifier accuracy.