

Figure 1

Left Panel – The pruned classification tree. Each branch is created by using the value of a variable which most cleanly distinguishes one activity or set of activities from another. The first split is based on the jerks associated with taking a step and separates sedentary activities from ones involving movement. Each of the subsequent splits are made using other features which are present or strong in one or some groups and are absent or weak in one or other groups. This process is repeated until a single classification is made. The numbers next to the variable name reflect the level of the value at which the split was made.

Right Panel – A graph of the misclassifications made by the original tree model as the complexity of the model increases under cross-validation before pruning. The model will reach a point where it as it becomes so complicated it will become over-fitted to the training data and the misclassification will start to rise. This graph shows that that point had not been reached with 10 leaves. The final leaves add very little improving the model and in fact the distinctions it was making at the lowest level were not seen as helpful, resulting in the model being pruned back to 8 levels. This was done because clarity could be improved without affecting misclassification rates, not because of over-fitting.