Hi John,

That looks good. I think I can reproduce your result in R.

To do this:

1) Load the data into R: (type) cars <- read.table('http://www.win-vector.com/dfiles/car.data.csv',header=TRUE,sep=',')
2) Take a look at the data: (type) View(cars); summary(cars)
3) Build the safety/ration pivot table: (type) with(cars,table(safety,rating))

I thought about this some more, and I also noticed no two-seater gets an acceptable rating. In fact I think we can recover a pretty accurate decision diagram by doing the following:

1) (paste):
```
library(rpart); library(rpart.plot)
model <- rpart(rating ~ buying + maint + doors + persons + lug_boot + safety, data=cars,
control=rpart.control(maxdepth=6))
rpart.plot(model,extra=4)
levels(cars$rating)
```

What this tells us is that there are four ratings in the magazine (in alphabetic order): "acc" "good" "unacc" "vgood" and a procedure for mapping measured car features to ratings is given by the following diagram (which we could print and use. For example: a car that seats 4 people, medium safety, high price and low maintenance would be should be rated unacceptable.   Each node on the diagram is showing in the data we know about what proportion of the examples in that part of the diagram/tree have a given rating.   For example the node ".83 .00 .07 .10" is 83% acceptable, 0% good, 7% unacceptable, and 10% very good. So our procedure for that leaf node would be to say "acceptable" and we would expect to by right about 83% of the time (at least on the original data).

I think we could even circulate the diagram (with some training) to our representatives and they could make the prediction on paper.

Hi John,

Please find attached the sample car data I mentioned. It is a sample from the UCI database
https://archive.ics.uci.edu/ml/datasets/Car+Evaluation and is synthetic data that simulates recovering the automobile
ratings/recommendations of fictional consumer magazine as function of car price, maintenance, efficiency, and safety. I know we wanted to
get a good estimate of what policy is being used to convert these measured features into ratings.

I looked at it and found (as we would hope) that safety is a big factor in the recommendations.

You can also download my prepared file from: http://www.win-vector.com/dfiles/car.data.csv

To do this I:
 1) Loaded the data into Microsoft Excel (File -> Open -> car.data.csv , make sure "All Files" is selected in the Enable tab on the file
dialogue box).
 2) Build a pivot table of safety versus rating:
      Build an extra column called "CarCount" filled with 1s.
      Click on any cell in the spreadsheet
      Set the control panel to "Data" (top panes)
      Click on PivotTable (in the Analysis group on the left).
      In the Pivot Table Builder dialogue box clear out all the stuff Excel selected (by dragging it to nowhere) and then drag "safety" into Row
Labels,
          "rating" into Column Labels, and CarCount into the Sigma Values panel.
 3) Notice how all low-safety cars (this is a row) have unacceptable ratings and only high-safety cars get a "vgood" (very good rating).

I think we can recover the rating rules and relative importance of the car features with a bit more work. This would allow us rate cars on our
own (after we examine a car and assign the feature values).