Practice problems for multinomial regression models with partial answers

1. This problem involves using multinomial regression models for the turkey data set given in Johnson’s textbook. The data set is in the turkey.csv file available on my course website. Below is how I read in the data:

> turkey <- read.csv(file = "C:\\chris\\turkey.csv")

 TYPE HUM FEM TIB TIN MAX MIN MMS BC STL

1 W 151 139 228 139 13.0 9.8 12.3 22 213

2 W 154 146 234 151 14.7 10.7 13.5 23 232

3 W 154 146 234 151 14.7 10.7 13.5 23 232

4 OT 146 133 207 141 19.0 14.4 17.2 28 127

5 OT 152 139 214 145 19.0 14.0 16.8 29 134

6 OT 147 133 198 132 18.1 12.5 16.1 26 134

> tail(turkey)

 TYPE HUM FEM TIB TIN MAX MIN MMS BC STL

95 OT 164 148 242 155 14.4 10.5 13.1 25 226

96 OT 157 144 239 155 13.7 9.6 12.2 24 196

97 OT 154 142 229 149 14.7 10.3 12.8 24 205

98 OT 153 145 233 147 14.1 10.6 12.7 25 226

99 W 155 143 234 152 14.3 9.9 13.0 21 227

100 W 158 142 235 152 13.9 11.2 12.6 21 223

There are three types of turkeys (TYPE) in the sample – big-breasted (BB), wild turkeys (W), and all other types (OT). The purpose is to determine how well the remaining variables will classify the observations into these three types. These remaining variables are bone lengths measured in millimeters with the following definitions:

* HUM – humerus
* FEM – femur
* TIB – tibiotarsus
* TIN – tibiotarsus measured to the nutrient foramen
* MAX – tibiotarsus maximum diameter at the nutrient foramen
* MIN – tibiotarsus minimum diameter at the nutrient foramen

Unfortunately, I cannot find definitions of MMS, BC, and STL. Using the above data set, complete the following.

* 1. Examine the data using the appropriate graphical methods discussed earlier in our course. How well do you think that multinomial regression models will be able to differentiate between the different types of turkeys? Explain.

Below is one of the plots that I created.



With respect to the parallel coordinate plot, there definitely appears to be some separation between the BB and W turkey types for the MAX, MIN, MMS, BC, and STL variables. These same variables also have many OT turkeys that have much different values than the W turkeys.

* 1. Using a multinomial regression model with all variables in the data set, examine how well the model correctly classifies the turkeys.

Cross-validation results:

> summarize.class(original = turkey$TYPE, classify = save.cv$classify)

$class.table

 classify

original BB OT W

 BB 17 3 0

 OT 3 12 11

 W 1 5 48

$prop

 classify

original BB OT W

 BB 0.8500 0.1500 0.0000

 OT 0.1154 0.4615 0.4231

 W 0.0185 0.0926 0.8889

$overall.correct

[1] 0.77

The largest problem is with classifying the OT turkeys as wild.

* 1. Use the appropriate graphical methods to assess where the model correctly and incorrectly classifies the turkeys.

Below is one of my plots: 

We can see that the OT turkeys are often incorrectly classified as being W (small red triangles overlapped by the larger green diamonds).