Multinomial Logistic Regression

Example: Vertebral Column

Our data example uses "six biomechanical attributes derived from the shape and orientation of the pelvis and lumbar spine" for a patient to classify the patient into one of three groups representing different conditions that may be affecting their spine: "DH" for disk hernia, "SL" for Spondylolisthesis, or "NO" for normal (neither of the other two conditions).

The data are available at https://archive.ics.uci.edu/ml/datasets/Vertebral+Column and were discussed in:

Berthonnaud, E., Dimnet, J., Roussouly, P. & Labelle, H. (2005). 'Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters', Journal of Spinal Disorders & Techniques, 18(1):40â€“47.

Reading the data in, preprocessing, train/test split

```
library(readr)
library(purrr)
library(dplyr)
library(ggplot2)
library(gridExtra)
library(rpart)
library(caret)

vertebral_column <- read_table2("http://www.evanlray.com/data/UCIML/vertebral_column/column_3C.dat",
  col_names = FALSE)
names(vertebral_column) <- c(paste0("X_", 1:6), "type")

vertebral_column <- vertebral_column %>%
  mutate(
    type = factor(type)
  )

set.seed(723)

# Train/test split
tt_inds <- caret::createDataPartition(vertebral_column$type, p = 0.7)
train_set <- vertebral_column %>% slice(tt_inds[[1]])
test_set <- vertebral_column %>% slice(-tt_inds[[1]])
```

Fit and test set classification error rate via multinomial logistic regression

Fit model

```
multilogistic_fit <- train(
  type ~ .,
  data = train_set,
  trace = FALSE,
  method = "multinom",
  trControl = trainControl(method = "none")
)
```

Test set performance: looking for low test set error rate, high test set accuracy

```
mean(test_set$type != predict(multilogistic_fit, test_set))
```

```
## [1] 0.1397849
```

```
mean(test_set$type == predict(multilogistic_fit, test_set))
```

```
## [1] 0.8602151
```