Multinomial Logistic Regression

Logistic Regression II

Learning Objectives

1. Understand how the logistic model can be extended to handle multiple classes.

Multi-class Classification Problems

- Multi-class: Categorical response variable represents more than 2 categories.
- Examples
  - Given high school grades, which program will a student select in college: general, academic or vocation?
  - Given the symptoms of a patient in a medical emergency room, what medical condition do they have: stroke, drug overdose or epileptic seizure?
  - Given activity on online store, which product is the customer likely to buy?
- Response variables are nominal, i.e. the order of the categories does not matter.
- Multinomial Logistic Regression: Applying Logistic Regression to predict nominal categorical response variable.

Multinomial Logistic Model

From "Yes or No?" to "Which class?"

- Binary logistic model: Odds represent the ratio of probability of Positive ($p(X)$) vs. probability of Negative ($1 - p(X)$)

$$ln(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X \tag{1}$$

- Multinomial logistic model:
  - Say, our problem has c classes, and call each class k, where $k = 1, 2, ..., c$. Given an observation with input X, probability of this observation to be in class k is $p_k(X) = p(k|X)$.
  - Define a baseline class, say class c. So, given an input of X, probability of this observation to be in class c is $p_c(X)$.
  - For each of the rest of the classes, i.e. $k = 1, 2, ..., c-1$, the odds represent the ratio of probability of class k ($p_k(X)$) vs. probability of class c, ($p_c(X)$)

$$ln(\frac{p_k(X)}{p_c(X)}) = \beta_0^k + \beta_1^k X \tag{2}$$

Multinomial Logistic Model

Multiple and Multinomial!

- Binary logistic model for multiple predictor variables, X_1, X_2, ..., X_p:

$$ln(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \tag{3}$$

- Multinomial logistic model for multiple predictor variables, X_1, X_2, ..., X_p:

$$ln(\frac{p_k(X)}{p_c(X)}) = \beta_0^k + \beta_1^k X_1 + \beta_2^k X_2 + ... + \beta_p^k X_p \tag{4}$$

Multinomial Logistic Model

Example

- Say, our multi-class classification problem has 4 predictor variables, and the categorical response variable has 3 classes. How many parameters (β), will the multinomial logistic model have?

Multinomial Logistic Model

Example

- Say, our multi-class classification problem has 4 predictor variables, and the categorical response variable has 3 classes. How many parameters (β), will the multinomial logistic model have?
- Multinomial logistic model with 4 predictor variables, X_1, X_2, X_3 and X_4:

$$ln(\frac{p_1(X)}{p_3(X)}) = \beta_0^1 + \beta_1^1 X_1 + \beta_2^1 X_2 + \beta_3^1 X_3 + \beta_4^1 X_4 \tag{5}$$

$$ln(\frac{p_2(X)}{p_3(X)}) = \beta_0^2 + \beta_1^2 X_1 + \beta_2^2 X_2 + \beta_3^2 X_3 + \beta_4^2 X_4 \tag{6}$$

Multinomial Logistic Model

Example

- Say, our multi-class classification problem has 4 predictor variables, and the categorical response variable has 3 classes. How many parameters (β), will the multinomial logistic model have?
- Multinomial logistic model with 4 predictor variables, X_1, X_2, X_3 and X_4:

$$ln(\frac{p_1(X)}{p_3(X)}) = \beta_0^1 + \beta_1^1 X_1 + \beta_2^1 X_2 + \beta_3^1 X_3 + \beta_4^1 X_4 \tag{5}$$

$$ln(\frac{p_2(X)}{p_3(X)}) = \beta_0^2 + \beta_1^2 X_1 + \beta_2^2 X_2 + \beta_3^2 X_3 + \beta_4^2 X_4 \tag{6}$$

- Total number of parameters $= (4 + 1) \times (3 - 1) = 10$

Multinomial Logistic Model

Example

- Say, our multi-class classification problem has 4 predictor variables, and the categorical response variable has 3 classes. How many parameters (β), will the multinomial logistic model have?
- Multinomial logistic model with 4 predictor variables, X_1, X_2, X_3 and X_4:

$$ln(\frac{p_1(X)}{p_3(X)}) = \beta_0^1 + \beta_1^1 X_1 + \beta_2^1 X_2 + \beta_3^1 X_3 + \beta_4^1 X_4 \tag{5}$$

$$ln(\frac{p_2(X)}{p_3(X)}) = \beta_0^2 + \beta_1^2 X_1 + \beta_2^2 X_2 + \beta_3^2 X_3 + \beta_4^2 X_4 \tag{6}$$

- Total number of parameters $= (4+1) \times (3-1) = 10$
- To find the log odds for the pair of class 1 and 2, we can derive the equation:

$$ln(\frac{p_1(X)}{p_2(X)}) = (\beta_0^1 - \beta_0^2) + (\beta_1^1 - \beta_1^2)X_1 + (\beta_2^1 - \beta_2^2)X_2 + (\beta_3^1 - \beta_3^2)X_3 + (\beta_4^1 - \beta_4^2)X_4 \tag{7}$$

Multiclass Classification Example

Dataset

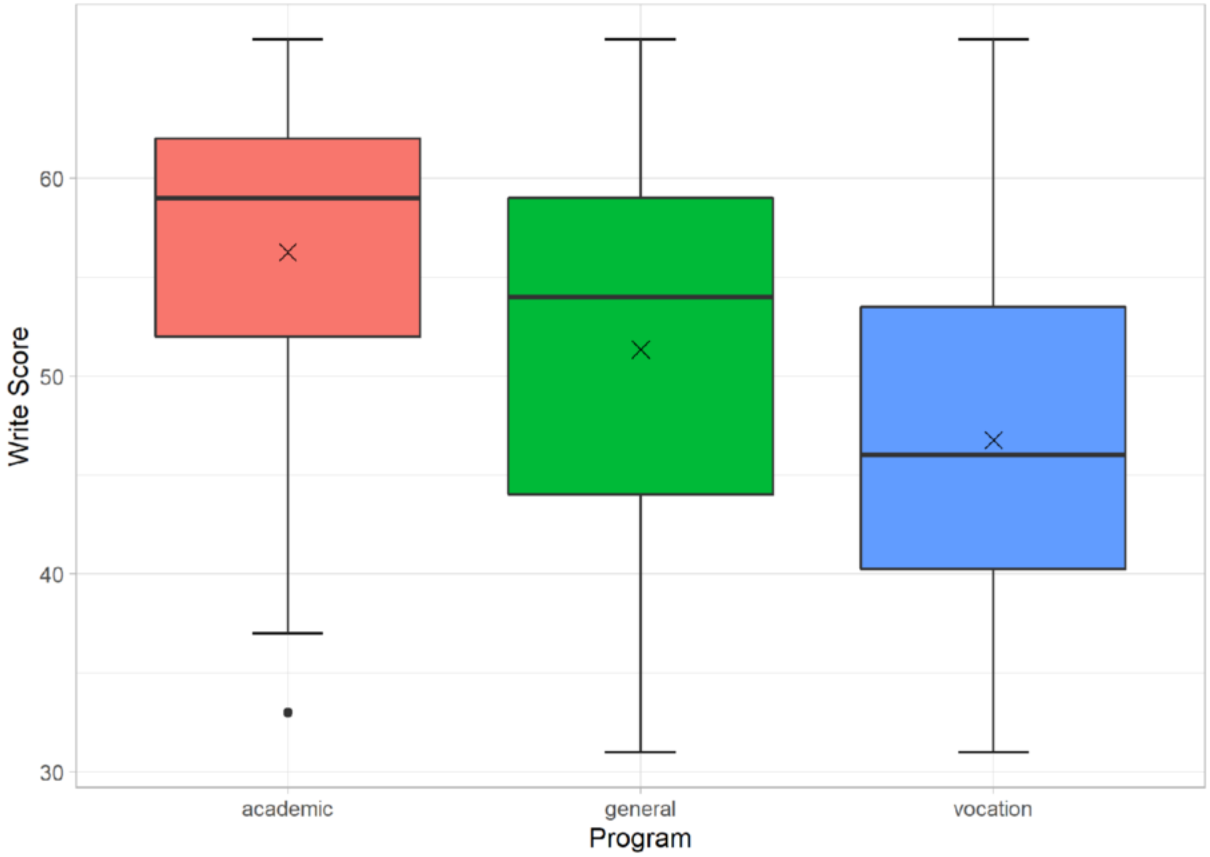
- High School & Beyond dataset contains demographic information and standardised test scores of high school students.

```
prog_data <- read.csv("../data/hsbdemo.csv")
str(prog_data)
```

```
'data.frame': 200 obs. of  13 variables:
$ id     : int  45 108 15 67 153 51 164 133 2 53 ...
$ schtyp : chr  "public" "public" "public" "public" ...
$ prog   : chr  "vocation" "general" "vocation" "vocation" ...
$ write  : int  35 33 39 37 31 36 36 31 41 37 ...
...
```

- Predictor variables: Write score (`write`).
- Response variable: Program choice with 3 possible classes - academic, general, vocational.

Predictor Variable: Write Score



- Students selecting academic programs have the highest write scores, and those selecting vocation have the lowest write scores.
- Performing an ANOVA test to compare the means gives a p-value less than 0.05. Writing score (`write`) is a significant predictor for `program`.

Predictor Variable: Write Score

Code

```
ggplot(prog_data, aes(x=prog, y=write)) +
  stat_boxplot(geom = 'errorbar', width = 0.2) +
  geom_boxplot(aes(fill=prog)) +
  stat_summary(fun="mean", geom="point", shape=4, size=3) +
  theme_light() + theme(legend.position = "none") +
  labs(x="Program",y="Write Score")
one.way <- aov(write ~ prog, data = prog_data)
summary(one.way)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)
prog          2   3176  1587.8   21.27 4.31e-09 ***
Residuals   197  14703    74.6
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Fitting Multinomial Logistic Regression Model

- Response variable, `prog` has 3 classes: (1) academic (2) general (3) vocation.
- Using prog = general as the baseline, multinomial logistic regression equations are:

$$ln(\frac{p_1(X)}{p_2(X)}) = \beta_0^1 + \beta_1^1 \times \text{write} \tag{8}$$

$$ln(\frac{p_3(X)}{p_2(X)}) = \beta_0^2 + \beta_1^2 \times \text{write} \tag{9}$$

Fitting Multinomial Logistic Regression Model

Code

```
library(VGAM)
ml <- vglm(prog ~ write, family = multinomial(refLevel = "general"),
           data = prog_data)
summary(ml)
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.71249    1.13280  -2.395  0.01664 *
(Intercept):2  2.64650    1.12744   2.347  0.01891 *
write:1        0.06601    0.02100   3.143  0.00167 **
write:2       -0.05180    0.02251  -2.301  0.02140 *
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
```

Fitting Multinomial Logistic Regression Model

```
levels(factor(prog_data$prog))
coef(ml, matrix=TRUE)
```

```
[1] "academic" "general"  "vocation"

            log(mu[,1]/mu[,2]) log(mu[,3]/mu[,2])
(Intercept)        -2.71248969         2.64650420
write               0.06600789        -0.05180105
```

- Here is our model:

$$ln(\frac{p_1(X)}{p_2(X)}) = -2.71 + 0.066 \times \text{write} \tag{10}$$

$$ln(\frac{p_3(X)}{p_2(X)}) = 2.65 - 0.052 \times \text{write} \tag{11}$$

Interpreting Coefficients

- Here is our model for response, (1) academic (2) general (3) vocation:

$$ln(\frac{p_1(X)}{p_2(X)}) = -2.71 + 0.066 \times \text{write} \tag{12}$$

$$ln(\frac{p_3(X)}{p_2(X)}) = 2.65 - 0.052 \times \text{write} \tag{13}$$

- As the write score increases by 1 unit, the log odds of selecting academic versus general, increases by 0.066. This means the odds are multiplied by $e^{0.066} = 1.068$.
- As the write score increases by 1 unit, the log odds of selecting vocation versus general, decreases by 0.052. This means the odds are multiplied by $e^{-0.052} = 0.949$.

Estimated Probabilities

Code

- The probabilities $p_1(X)$, $p_2(X)$ and $p_3(X)$ add up to 1.

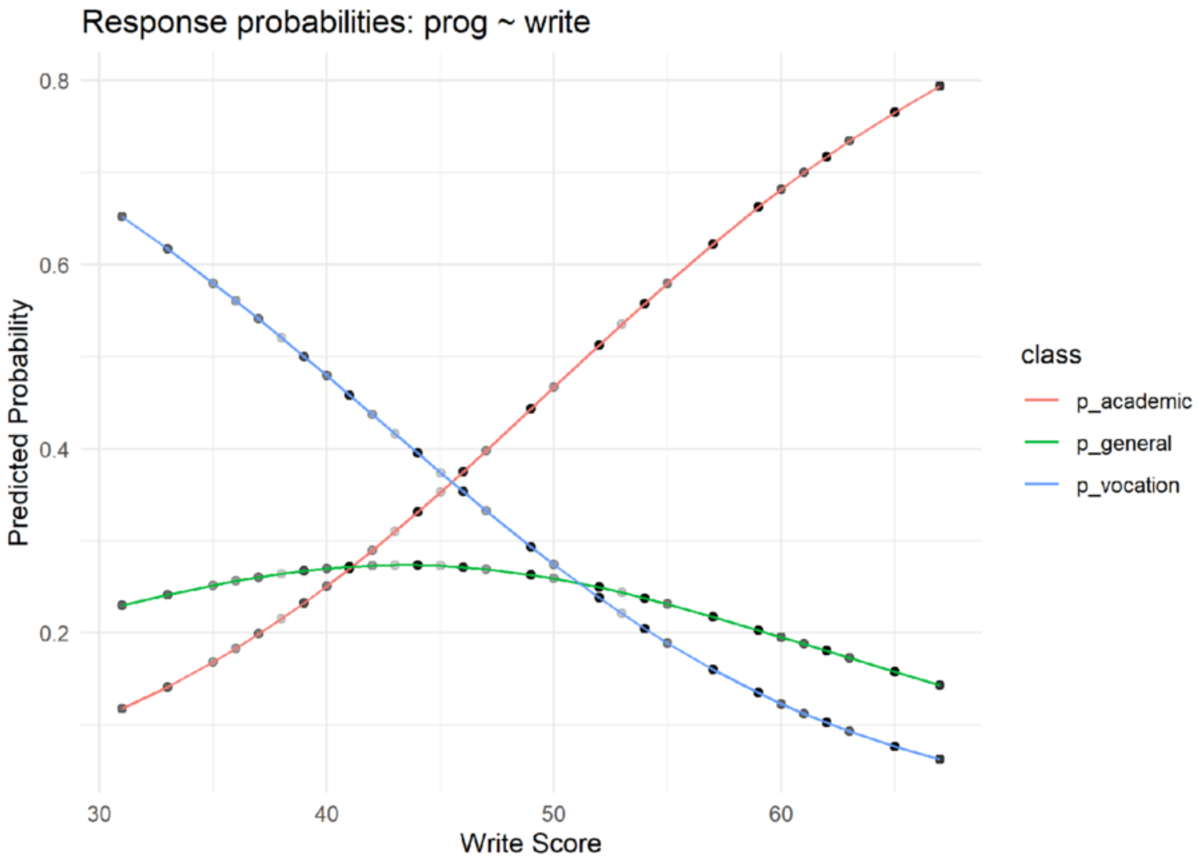
```
probs <- fitted(ml)
pd1 <- prog_data %>% select(write, prog) %>%
         mutate(p_academic = probs[,1], p_general = probs[,2],
                p_vocation = probs[,3])
head(pd1)
```

```
  write     prog p_academic p_general p_vocation
1    35 vocation  0.1684678 0.2518814  0.5796508
2    33  general  0.1416226 0.2416269  0.6167505
3    39 vocation  0.2327754 0.2672691  0.4999555
4    37 vocation  0.1988595 0.2605503  0.5405902
5    31 vocation  0.1182193 0.2301628  0.6516179
6    36  general  0.1832184 0.2564374  0.5603441
```

Correct Prediction

Wrong Prediction

Visualising the Model



- Model predicts higher probabilities for `academic` as write score increases.
- Model predicts lower probabilities for `vocation` as write score increases.
- For any value of write score, select the class with the highest probability as the final prediction.
- Also, for any value of write score, the estimated probabilities add up to 1.

Visualising the Model

Code

```
pd2 <- pivot_longer(pd1, cols = starts_with("p_"),
                    names_to = "class", values_to = "pred_prob")
ggplot(pd2, aes(x=write, y=pred_prob)) +  geom_point(alpha = 0.2) +
  geom_line(aes(color=class)) + theme_minimal() +
  labs(x="Write Score", y="Predicted Probability",
  title="Response probabilities: prog ~ write")
head(pd2)
```

```
     write prog      class        pred_prob
     <int> <chr>     <chr>            <dbl>
1       35 vocation p_academic      0.168
2       35 vocation p_general       0.252
3       35 vocation p_vocation      0.580
```

Evaluating Model

```
preds <- colnames(fitted(ml))[apply(fitted(ml), 1, which.max)]
pd1 <- pd1 %>% mutate(pred_class = preds)
head(pd1)
mean(pd1$prog == pd1$pred_class)
```

```
  write     prog p_academic p_general p_vocation pred_class
1    35 vocation  0.1684678 0.2518814  0.5796508   vocation
2    33  general  0.1416226 0.2416269  0.6167505   vocation
3    39 vocation  0.2327754 0.2672691  0.4999555   vocation
4    37 vocation  0.1988595 0.2605503  0.5405902   vocation
5    31 vocation  0.1182193 0.2301628  0.6516179   vocation
6    36  general  0.1832184 0.2564374  0.5603441   vocation
```

Correct Prediction

Wrong Prediction

```
[1] 0.585
```

References I

Agresti, A. (2018).
An introduction to categorical data analysis.
John Wiley & Sons.