Model Explanation Simplification

Dusty Turner

Building the model

I present below an arbitrary design matrix. It consists of 4 factors, x_1, x_2, y_1, and y_2 and all their interactions.

X =

Frame	Intercept	x_1	x_2	y_1	y_2	$x_1 \times x_2$	$x_1 \times y_1$	$x_1 \times y_2$	$x_2 \times y_1$	$x_2 \times y_2$	$y_1 \times y_2$
Frame 1	1	x_{11}	x_{21}	y_{11}	y_{21}	$x_{11}x_{21}$	$x_{11}y_{11}$	$x_{11}y_{21}$	$x_{21}y_{11}$	$x_{21}y_{21}$	$y_{11}y_{21}$
Frame 2	1	x_{12}	x_{22}	y_{12}	y_{22}	$x_{12}x_{22}$	$x_{12}y_{12}$	$x_{12}y_{22}$	$x_{22}y_{12}$	$x_{22}y_{22}$	$y_{12}y_{22}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Frame N	1	x_{1N}	x_{2N}	y_{1N}	y_{2N}	$x_{1N}x_{2N}$	$x_{1N}y_{1N}$	$x_{1N}y_{2N}$	$x_{2N}y_{1N}$	$x_{2N}y_{2N}$	$y_{1N}y_{2N}$

(1)

We use logistic regression to create the model.

$$\log\left(\frac{p}{1-p}\right) = X\beta \tag{2}$$

We can expand this to the following:

$$\log\left(\frac{p}{1-p}\right) = \beta_0+\beta_1 x1+\beta_2 x2+\beta_3 y1+\beta_4 y2+\beta_5(x1\times x2)+\beta_6(x1\times y1)+\beta_7(x1\times y2)+\beta_8(x2\times y1)+\beta_9(x2\times y2)+\beta_{10} \tag{3}$$

Where:

1) $\log\left(\frac{p}{1-p}\right)$ is the logit function, the natural logarithm of the odds of the outcome.

2) β_0 is the intercept of the model.

3) $\beta_1, \beta_2, \beta_3, \beta_4$ are the coefficients for the main effects of the predictors $x1, x2, y1$, and $y2$ respectively.

4) $\beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ are the coefficients for the interaction effects between the predictors.

5) $x1, x2, y1, y2$ are the independent variables or predictors.

6) $(x1 \times x2), (x1 \times y1), (x1 \times y2), (x2 \times y1), (x2 \times y2), (y1 \times y2)$ are the interaction terms between the predictors.

7) ϵ is the error term, representing the variation in the outcome not explained by the model.

Using the model to make predictions

Later we want to use this model to predict something new. Lets use the new data below.

X =

Frame	Intercept	$x5$	$x6$	$y5$	$y6$	$x5 \times x6$	$x5 \times y5$	$x5 \times y6$	$x6 \times y5$	$x6 \times y6$	$y5 \times y6$
Frame 1	1	x_{51}	x_{61}	y_{51}	y_{61}	$x_{51}x_{61}$	$x_{51}y_{51}$	$x_{51}y_{61}$	$x_{61}y_{51}$	$x_{61}y_{61}$	$y_{51}y_{61}$

(4)

1) $\log \widehat{\left(\frac{p}{1-p}\right)} = \beta_0 + \beta_1 x_{51} + \beta_2 x_{61} + \beta_3 y_{51} + \beta_4 y_{61} + \beta_5(x_{51} \times x_{61}) + \beta_6(x_{51} \times y_{51}) + \beta_7(x_{51} \times y_{61}) + \beta_8(x_{61} \times y_{51}) + \beta_9(x_{61} \times y_{61}) + \beta_{10}(y_{51} \times y_{61})$

Lets go a step farther an predict another row:

2) $\log \widehat{\left(\frac{p}{1-p}\right)} = \beta_0 + \beta_1 x_{51} + \beta_2 x_{71} + \beta_3 y_{51} + \beta_4 y_{71} + \beta_5(x_{51} \times x_{71}) + \beta_6(x_{51} \times y_{51}) + \beta_7(x_{51} \times y_{71}) + \beta_8(x_{71} \times y_{51}) + \beta_9(x_{71} \times y_{71}) + \beta_{10}(y_{51} \times y_{71})$

We know that the results for equation 1 and equation 2 will be different.