What is Multinomial Logistic Regression?

Bookmark this page

There are several extensions to standard logistic regression when the response variable, Y, has more than two categories. The two most common are:

Ordinal logistic regression: used when the categories have a specific hierarchy (like class year: Freshman, Sophomore, Junior, Senior; or a 7-point rating scale, from Strongly Disagree to Strongly Agree).

Nominal logistic regression: used when the categories have no inherent order (like eye color: blue, green, brown, hazel, etc).

For example, we could attempt to predict a student's concentration

$$y = \begin{cases} 1 & \text{if Computer Science (CS)} \\ 2 & \text{if Statistics} \\ 3 & \text{if Physics} \end{cases}$$

from predictors x_1, number of psets per week, and x_2, time spent in the library.

One vs. Rest (OVR) Logistic Regression

An option for **nominal** (non-ordinal) categorical logistic regression model is called the '**One vs. Rest**' approach.

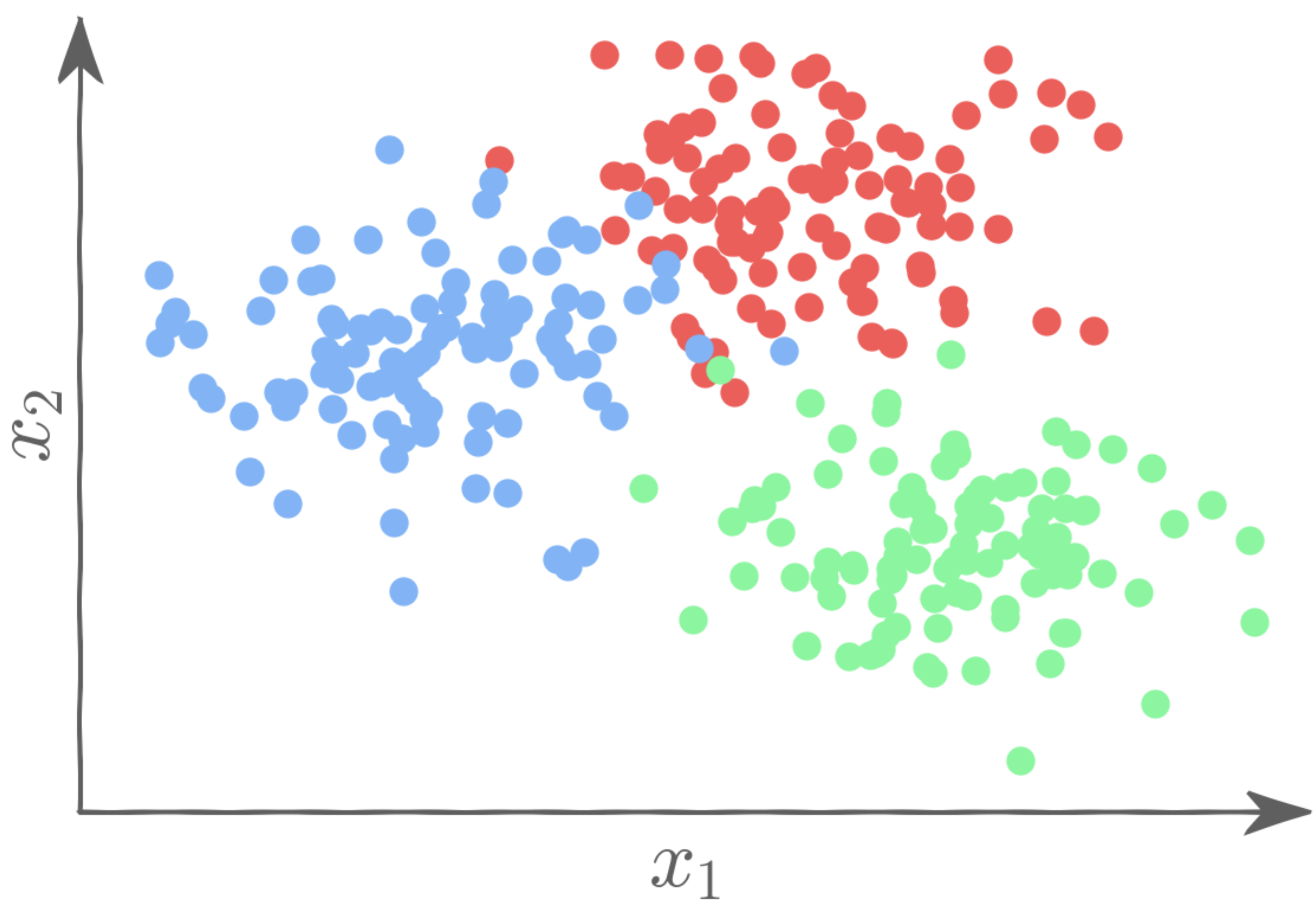
If there are 3 classes, then *3 separate logistic regression models are fit*. Each model is associated with a specific class and predicts the probability of a given observation belonging to that class as opposed to all the others.

So, for the concentration example, we would fit 3 models:

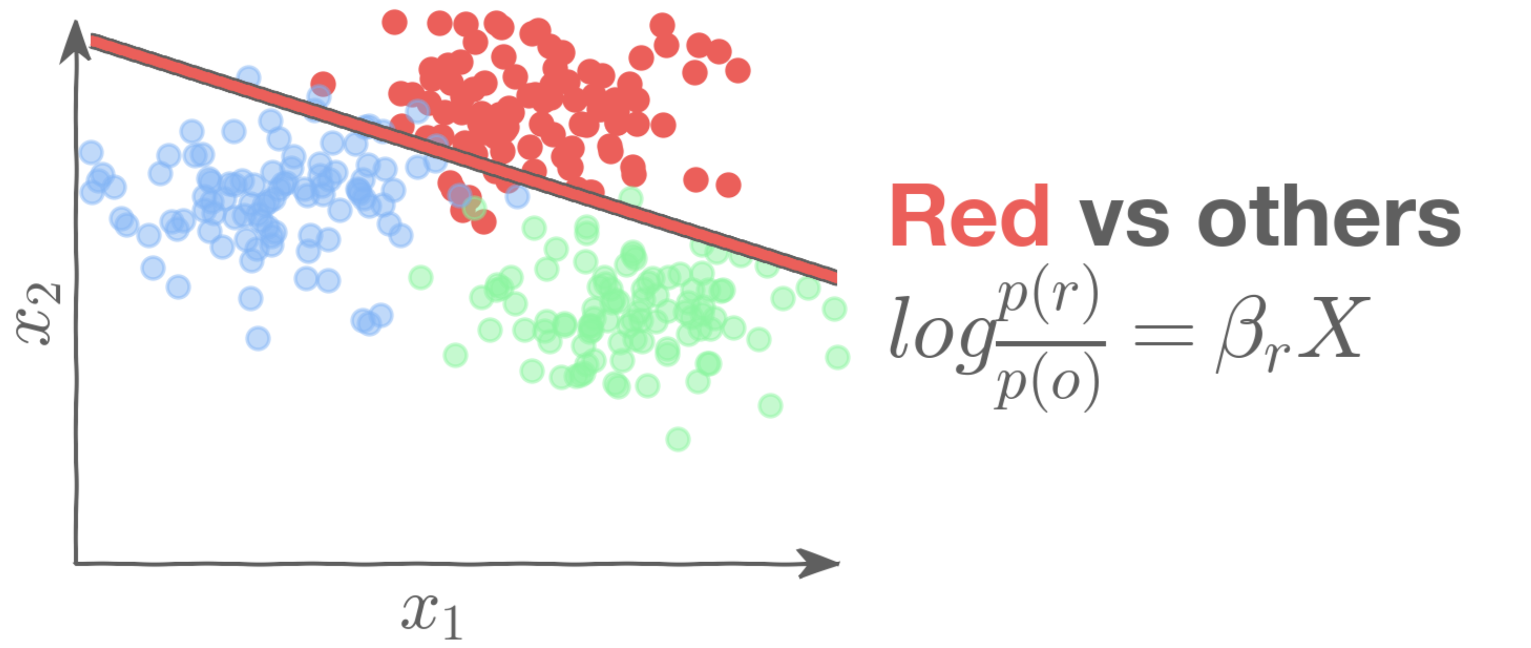
- a first model to predict *CS* from *Stat* and *Physics* combined,
- a second model to predict *Stat* from *CS* and *Physics* combined,
- a third model to predict *Physics* from *CS* and *Stat* combined.

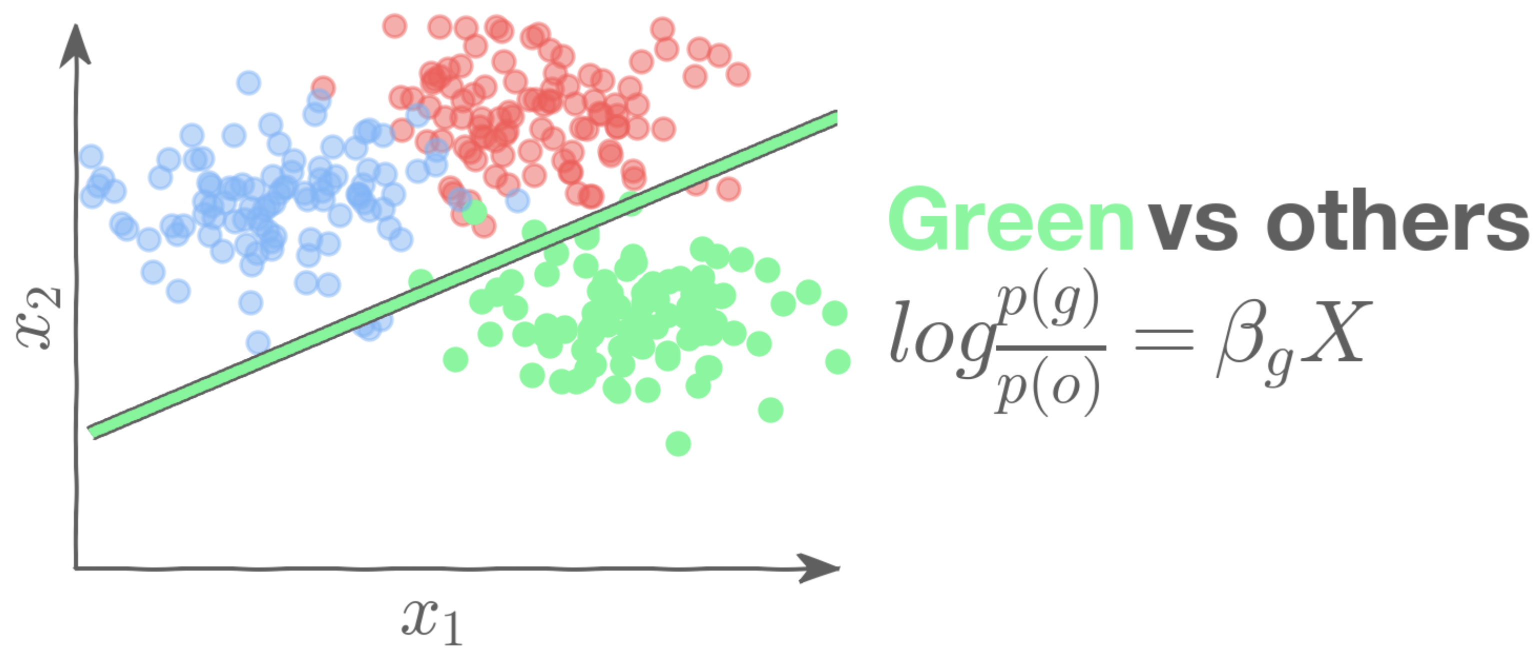
A Visual Example of OVR Logistic Regression

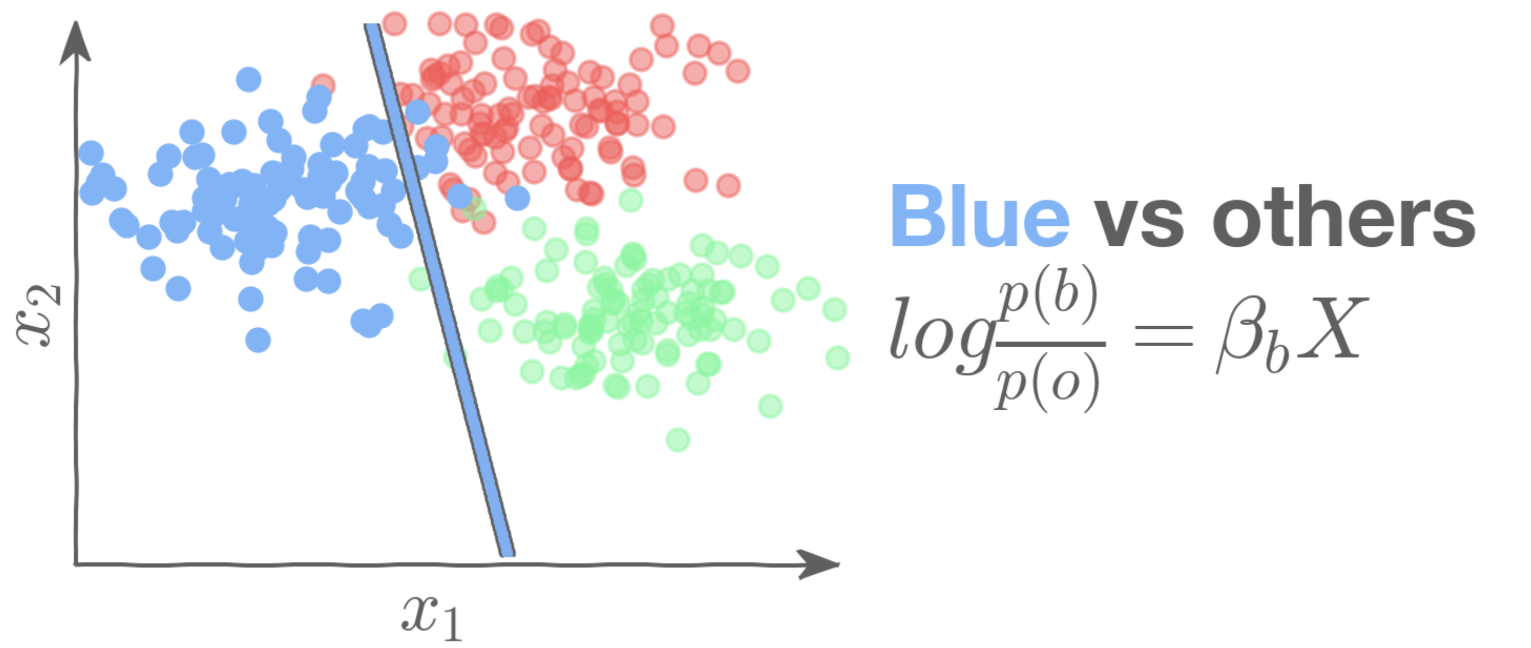
Let's say we want to classify observations as belonging to one of three classes: Red, Blue, or Green.



We can then use three binary logistic regression models:







Sklearn normalizes the output of each of the three models when predicting probabilities:

$$\tilde{P}(b) = \frac{P(b)}{P(g) + P(b) + P(r)}$$

$$\tilde{P}(g) = \frac{P(g)}{P(g) + P(b) + P(r)}$$

$$\tilde{P}(r) = \frac{P(r)}{P(g) + P(b) + P(r)}$$

True Multinomial Logistic Regression

Another option for a multiclass logistic regression model is what we will call the "true" multinomial logistic regression model. Here one of the classes is chosen as the baseline group (think $Y = 0$ group in typical logistic regression), and the other $K - 1$ classes are compared to it. Thus, a sequence of $K - 1$ binary models are built to predict being in class k as opposed to the baseline class.

$$\log\left(\frac{P(Y=k)}{\dots}\right) = \beta_{0,k} + \beta_{1,k} X_1 + \dots + \beta_{p,k} X_p$$

$$\log\left(\frac{}{P\left(Y=K\right)}\right)=\beta_{0,k}+\beta_{1,k}X_1+\ldots+\beta_{p,k}X_p$$

This is mathematically equivalent to using the softmax function:

$$P\left(y=k\right)=\frac{e^{X\beta_k}}{\sum_{k=1}^{K}e^{X\beta_k}}$$

And the cross-entropy as the loss function:

$$L=-\sum_i\sum_k I\left(y_i=k\right)P\left(y_i=k\right)$$

Comparing OVR and true multinomial logistic regression

True Multinomial is slightly more efficient in estimation since there are technically fewer parameters (though sklearn reports extra ones to normalize the calculations to 1) and it is more suitable for inferences/group comparisons.

OVR is often preferred for determining classification: you simply just predict from all 3 separate models (for each individual observation) and choose the highest probability.

They give **VERY similar results** in estimated probabilities and classifications.

> ⚠ **More Than 2 Categories**
>
> When there are more than 2 categories in the response variable, then there is no guarantee that $P\left(Y=k\right)\geq 0.5$ for any one category. So, any classifier based on logistic regression will instead have to select the group with the largest estimated probability.

Discussion Board (External resource)

Haga clic en Aceptar para que su nombre de usuario y dirección de correo electrónico se envíen a una aplicación de terceros.

Aceptar

< Previous Next >

Calculator