Multinomial logistic regression

- ✓ What is Multinomial logistic regression?
- ✓ Where it is applicable?
- ✓ How it is solved / worked?
- ✓ How does it predict the dependent variable?

What is Multinomial logistic regression?

Binary outcome – dichotomous

- The dependent variable is taking values like
  - ✓ Responder / non responder
  - ✓ Loss giving / good profile
  - ✓ Buyer / non buyer
  - ✓ Account holder will make payment / no payment

More than two outcome – polytomous / multiclass / polychotomous logistic / softmax regression / multinomial logit/ maximum entropy (MaxEnt) classifier / conditional maximum entropy model

- At times, the dependent variable has
  - ✓ More than two possible outcome
  - ✓ They are nominal variable : There is no order in the outcome
  - ✓ And we need to use the independent variables to predict the outcome

Example 1 : which stream will be chosen by student

Dependent variable - major

- Science
- Arts
- Commerce

Note there is no order in dependent variable here

Independent Variable

- Grade
- Mathematical aptitude
- IQ
- Parents profile

Example 2 : which ice cream will be chosen by kids

Dependent variable – ice cream type

- Vanilla
- Strawberry
- Chocolate

Independent Variable

- Age
- Gender
- etc

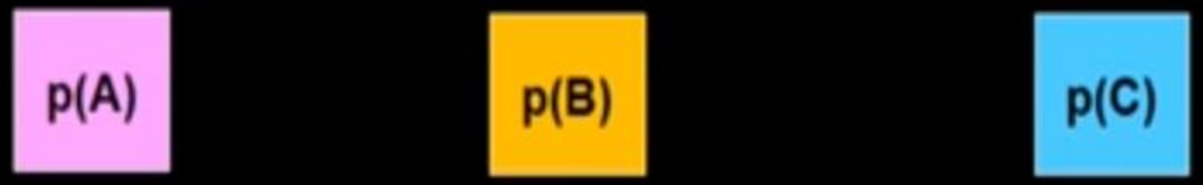
How it is solved?

- Simple approach – k models for k classes
- As a set of independent binary regressions

Convert multinomial to many binomial

- Let's say – there are three classes of nominal outcome A, B and C
- Steps 1
  - ✓ Develop three model separately. Class A vs Rest, Class B vs Rest ..

  - ✓ Develop equation for three probabilities



  - ✓ Assign any record to the class, based on the input variables, which has highest probability
  - ✓ Like if p(A) > p(B) and p(A) > p(C) then outcome = class A

Slightly advance solution approach – simultaneous models

- ✓ K-1 models for k classes
- ✓ As a set of independent binary regressions
- ✓ Note it is just one of the many methods

Let's see it in context of logistic regression

- In case of two classes 1 vs 0 (or A vs B) – we used to develop one logistic model, right?
- Log(p/ (1-p))= a+b1*x1+b2*x2....
- If p > =0.5 then class 1 (or class A), else otherwise
- Can we extend this for multi class?
- I means – for k classes, can't we develop just k-1 models?
- Answer is → yes
- Let me explain you how

Let's see it in context of logistic regression

- Say if there are three classes of the dependent variable A, B and C
- Now let's choose C as the reference class then
- Develop first model for log(p(A) / p(C)) = intercept_1 + b1*x1+ ......
- Let's call RHS_A = intercept_1 + b1*x1+ .....
- Then p(A) / p (C) = exp(RHS_A)
- Or p(A) = p(C) * exp(RHS_A)
- Similarly second model for log(p(B) / p(C)) = intercept_2 + b2*x1+ ......
- Let's call RHS_B = intercept _2 + b2*x1+ .....
- Then p(B) / p (C) = exp(RHS_B)
- Or p(B) = p(C) * exp(RHS_B)
- Please note p(A) + p(B) + p(C) =1
- p(C) * exp(RHS_A) + p(C) * exp(RHS_B) + p(C) = 1
- $p(C) = \dfrac{1}{(1 + \exp(RHS\_A) + \exp(RHS\_B))}$

Let's see it in context of logistic regression

- As p(A) = p(C) * exp(RHS_A)
- And p(B) = p(C) * exp(RHS_B)

- Hence $p(A) = \dfrac{\exp(RHS\_A)}{(1 + \exp(RHS\_A) + \exp(RHS\_B))}$

And $p(B) = \dfrac{\exp(RHS\_B)}{(1 + \exp(RHS\_A) + \exp(RHS\_B))}$

And for k class scenario $p(R) = \dfrac{\exp(RHS\_R)}{(1 + \exp(RHS\_A) + \exp(RHS\_B) + \ldots \exp(RHS\_K{-}1))}$ ..

Created by : Gopal Prasad Malakar

Some important aspect

- J -1 model equations simultaneously, results in smaller standard errors for the parameter estimates than when fitting them separately.
- Several methods of estimating parameters of these equations … not covering here
- The choice of baseline category has no effect on the parameter estimates for comparing two categories a and b.