82 Logistics Regression actival kill and the last of thought sports Form of regression analysis used for prediction of discrete variables using mix of continuous and discrete -) Used when the research objective is focussed on whether or not an event occurred trather than when it occurred i.e. time course information is not used acity adual rolon ("black") -> Instead of building a predictive model for "Y" Reps response directly, the approach models Log odds (Y); hence the name logistics or logit -) Used as an alternative to Multivariate Discriminant analysis when underlying assumptions for Discriminant analysis are violated Manyly Syrice Agrice Manager all

F

Logistics Regierion used when - Dependent Variable: Categorical Independent Variable: Continuous or Categorical examples Il monton a formation · Customer default on credit card Payment Customer reponse to direct mailer Customer will buy or not at homely so in months at book motor fell and we would be worden Cast morphory - dott 12 12 - March - March - Diller Howards

Types of Logistics Regression -) Used when response variable is binary or dicho tomous or - It has only 2 outcomes eg Good + Bad Yes + Ho! a Contorno propose of chiralters de Ordered legit In regrection to pud dlines transfer -) wed when response variable has more than 2 outcomes I and the outcomes can be ordered in meaningful way eg. High - Medium -Low strongly Agree- Agree - Disagree- Strongly disagree

Multinomial Logit -> used when response variable has more than 2 outcomes - and the outcomes cannot be ordered in any manner manner e.g. Type choice of Bread Travel Thineary 1 1 William & Sales (all and the Company) Legular Engrerian is Colombian and PCDV) = F(IX) (Probability) But, perbability values can take values believer a cold of

Thois a standard statistical term that denotes probability of success to probability of failure allow the outcome conditions of It probability of successis 0.75. then odds ratio = 0.75 = 3 (P) ie there is 3:1 chance of success the charge on I wood of the state of the sta Logistics Regression PCDV) = f(IN) (P: Paubability) But, Probability values can take values between o and 1

log 0 => -00 log => 0 odds ratio (P/1-P) and then take log The back of the X of goods have - - can take values of 0 to co 1-Partico Collaboration Collaboration The land of the North charge in x will flog(P) con bake values of - as to as and the state of the control of the state of Prediction values can take values from - 00 to 00 (1-8) × 001 = x01000 / 01 pp 10 10 of Commell Values of So the Equation! log(P)=Y=f(IX)

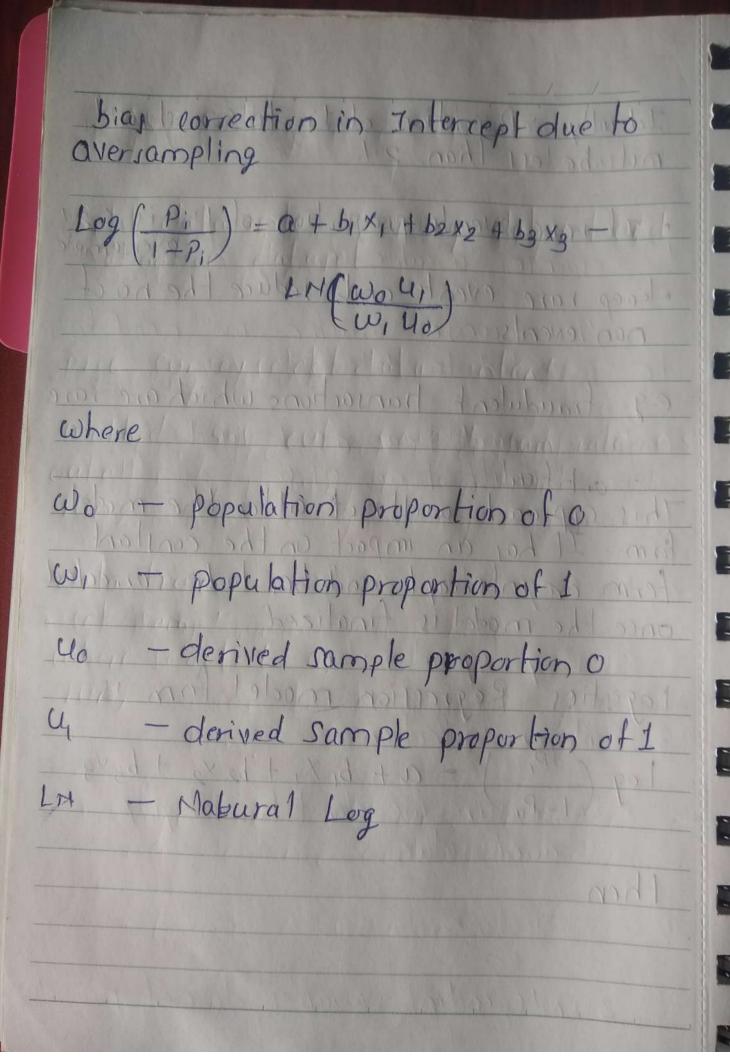
Log Transformation has linear relationship with predictors Aunit change in x will lead to fixed do change in log y Interms of Y =) A unit change in x will lead to multiplicative e change in y model to be out what soldate at the Pediction Mohine our latte votured from the 10 change in y approx = 100 x (eB-1) (for small values of coefficient) marien il cintoppe de parte

pata Preparation for Logistic Regression includes: antomo bros de la paración son o sono of monthly dealing as the light and 1. Response Variable Coding The reponse variable will need to be converted to \$10 Code Repaid Loan as 1 and Default Loan avon dolar a borry well for medaling to 2. Missing value Treatment cusing logical rules e joined if existing set of predicte 3. Outlier Peteotion ereta or an be crea · to ensure we don't have highly skewed values. Multicolinearity 2 indépendent variables du not provide similar info

5. Vaniable Transformation of variables depending on the research and modelling scope. The repeated worldby will need be 6. Descriptive Statistics of bollows meed to be output to validate if correct data is being wed for modelling Company and we proposed as a second s 7. Predictors Terrord Blood of State proposed · review if existing set of predictors are good enough or more meaningful predictors can be created. they share me appropriately sons strwed values 2 independent variables, do not provide

E

Ideally! Proportion of 1's to 0's should not be less than 21. . If rare events proportion of 1's is <21 non events and reduce the no of e.g. fraudulent transactions which are rare Concepted Lang York & Long Son 1 Superior artifical la This approach will not after the model form. It has an impact on the constant term or intercept and has to be corrected once the model is finalized. commended the property of the construction of Logistics Regression model form is Color to the sample of the sam $\begin{pmatrix} P_i \\ 1-P_i \end{pmatrix} = a + b_1 \times_1 + b_2 \times_2 + b_3 \times_3$ then



Estimation for Logistice The Coefficients for the Logistics equation are estimated using a technique known Maximum Likelihood Estimation (MLE) are everly and processing the processing of the second MLE is a popular method of estimation analogo de la contra del la contra de la contra del la contra de · It does not have any underlying assumptions of distribution · when the underlying distribution of error terms is normal, MLE estimates are similar to OLS estimates · OLS like many other distributions is a special raje of MLE Intution of MLE what values of the unknown parameters make the data we see least surprising &

SAS Code Militaria Maria The coefficient la the Logither Split data into Training dataset and Validation dataset proc survey select data = inters THE IL O Paper as mothed an extinction method = SRS out = SAMPI samprate = 0-3 run; data train; set samply long and where selected #0; run; BUCKERION BELLEVILLE our many other outputtent ua coeral colo all MIR a company dellarge peluce Total Harrison Mickey on a famaday sallandonal force econolice soute the data we rec

Logistics Regression model code proc logistics data = xxx outest = modelx; Model Repayment Statey = Credit Hist Lean Amount Imcomelevel Hangel Hangel Hand · Or ca the year ab'as a kalladed bloom Outest - save the output in a temporary SAS dataset Model -input dependent and independent variables hardaged tomandonal populares. of the many of toy of historias BORGER CONCRETE BUILDING especial difference between many der deligation of the long and religion Who I work by bor A done I " collection to a to the data we

Logistics Regression steps

- · Build a logistics model wing all the variables given
- · Choosing Variables
- · Using the variables short listed from profiling fby wing the significant variables, build multiple logistics regression models

9

- 3:

1

- · Best Model Summary will have
 - · P values to choose most significant
 - · Significant différence between nout deviance and residual deviance
 - « Least AIC value

	glm =) generalised linear model
	Model Validation
	· Plot the Rock curve · Choose a point that has high TRR and low FPR. Using this choose the cut off
	with highert Auc value choose the mode
	· Test the model on at least 3 to 4 tes samples to ensure model performance consistency
*	TPRNOID True Paritive Rate
*	FPR : False Positive Rate
1	Auc: Area Under Curve
3	

I'm squiralised linear model R Code: 1 Pec 6 cm 16 # customers who spend more than average are good customers family = "bianomial" = states that we want logistics regression output Comment of the state of the sta Step (mod, direction = "both") complete to ensure model poplamone #stepwise regression # direction =) forward as well as backward selection # It will give suggestion for attributes 3 9 rue ha wide lune

VE DE COLOR SOLO SOLO type = "response" = will give predictions in terms of predictions development that the post of south # kappa matrix - if 7 0.6 then it is good model Mana of the hort phartigher the the # confusion matrix 1816 Below of the State of the positive: 1" =) what we want predict will give parse that as paitive in a source the contract to the contract # what is gains chart ?