VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Essentials of Data Analytics
- (CSE3506)

Faculty – Lakshmi Pathi Jakkamputi **Sir**

Lab-5

Harshanth k Prakash

19BCE1293

L21-22 **Slot**

Tasks for Week-5: Logistic Regression

Understand the following operations/functions on to perform logistic Regression and perform similar operations on 'Social_Network_Ads' dataset based on given instructions.

AIM

To Understand and Perform logistic Regression and perform similar operations on given 'Social_Network_Ads' dataset.

Algorithm

1. Start
2. Set the directory to required location.
3. Import the given Dataset.
4. Convert the gender and purchased column to factor datas.
5. Use the glm command to perform the Logistic Regression on the given dataset.
6. Get the response using predict command.
7. Calculate the model measurements using confusion matrix.
8. Find the accuracy of the model.
9. If the accuracy value predicted is greater than at least 70%, then we accept the model as best-fit.
10. Stop.

Result

```
> head(mydata)
   User.ID Gender Age EstimatedSalary Purchased
1 15624510   Male  19           19000         0
2 15810944   Male  35           20000         0
3 15668575 Female  26           43000         0
4 15603246 Female  27           57000         0
5 15804002   Male  19           76000         0
6 15728773   Male  27           58000         0
```

```
> res
           1            2            3            4            5            6            7            8            9           10           11           12 
0.0007061408 0.0314610656 0.0063340671 0.0132775722 0.0056085336 0.0191141370 0.0347435349 0.5660038569 0.0048544343 0.1070697733 0.0239596097 0.0087710604 
          13           14           15           16           17           18           19           20           21           22           23           24 
0.0101864281 0.0146172401 0.0055072691 0.0652298859 0.4010242646 0.3018093315 0.3707863178 0.4128227573 0.2720074170 0.5348300047 0.6032030763 0.2111004499 
          25           26           27           28           29           30           31           32           33           34           35           36 
0.3293666637 0.3581494560 0.5453904942 0.3652065493 0.0177980692 0.0115689066 0.0826335502 0.1989265096 0.0007281918 0.0105077533 0.0588611284 0.0402350625 
          37           38           39           40           41           42           43           44           45           46           47           48 
0.0190157172 0.0277847372 0.0180098224 0.0051901789 0.0031226054 0.0428956007 0.4451681151 0.0082105504 0.0436287047 0.0018873909 0.0254142693 0.0119189397 
          49           50           51           52           53           54           55           56           57           58           59           60 
0.3962420708 0.1002620073 0.0026507711 0.0009920222 0.0528048194 0.0252942698 0.0137635645 0.0061075515 0.0037428398 0.0504122622 0.0013852029 0.2815079328 
          61           62           63           64           65           66           67           68           69           70           71           72 
0.0048554166 0.0337262027 0.0071873506 0.3789924212 0.9855453547 0.0094810297 0.0023055162 0.0128033765 0.0050943677 0.0492855993 0.0263326398 0.0022102072 
          73           74           75           76           77           78           79           80           81           82           83           84 
0.0007414962 0.3003168726 0.0146172401 0.4227687079 0.0018524359 0.0013771030 0.0484247671 0.0024654081 0.0812552057 0.1574328794 0.0026652612 0.2790785368 
          85           86           87           88           89           90           91           92           93           94           95           96 
0.0318238415 0.2427755697 0.0085075303 0.0451747631 0.0343284302 0.0883619006 0.0135901388 0.1904060345 0.0031981851 0.0074566223 0.0528048194 0.0528353368 
          97           98           99          100          101          102          103          104          105          106          107          108 
0.0271546890 0.2087657745 0.1830744859 0.0081612203 0.0549511415 0.0249742815 0.1123786832 0.6144587140 0.0005440362 0.0077701322 0.0047399001 0.0568745692 
         109          110          111          112          113          114          115          116          117          118          119          120 
0.0409086829 0.2966082645 0.2779801367 0.1933415873 0.2275774885 0.1574059105 0.6030570690 0.2903124871 0.1942273199 0.1167111264 0.3055547210 0.3580092470 
         121          122          123          124          125          126          127          128          129          130          131          132 
0.1795135902 0.2576617713 0.3608287774 0.0975737355 0.2298093560 0.2109988836 0.4679443921 0.0059259645 0.0088258224 0.0276156173 0.0478728394 0.0293087136 
         133          134          135          136          137          138          139          140          141          142          143          144 
0.1024471304 0.0048241938 0.0156082327 0.0089803141 0.0063302290 0.1913129629 0.0180134187 0.0008785644 0.0077685648 0.0023754645 0.1185920114 0.1093456179 
         145          146          147          148          149          150          151          152          153          154          155          156 
0.0215489272 0.0207731392 0.0527946507 0.1218899869 0.0337394575 0.0066020270 0.0022925123 0.2508297667 0.0883291485 0.0808631956 0.2212703347 0.0074581273 
         157          158          159          160          161          162          163          164          165          166          167          168 
0.6458466132 0.0549617015 0.0055117253 0.4301957335 0.2274702804 0.0374763491 0.0566146516 0.0589061989 0.0794969414 0.0045673580 0.0038110134 0.1298392754 
         169          170          171          172          173          174          175          176          177          178          179          180 
0.4540346614 0.0147917191 0.0138351841 0.4496477008 0.0892872161 0.0407106382 0.1088144223 0.0018093356 0.0585814819 0.0032564705 0.0026663423 0.0147946825 
         181          182          183          184          185          186          187          188          189          190          191          192 
0.0033164867 0.0546679723 0.2815079328 0.0446668148 0.0585702695 0.0630547910 0.0063302290 0.0301910051 0.1776872209 0.0094848494 0.0240925489 0.0006526970 
         193          194          195          196          197          198          199          200          201          202          203          204 
0.0177980692 0.0045120168 0.0710031185 0.0559426459 0.0575570871 0.0011902948 0.0331406920 0.0337593492 0.0609591338 0.8651054421 0.7927023308 0.3821189093 
         205          206          207          208          209          210          211          212          213          214          215          216 
0.9904455827 0.5166597425 0.9932204213 0.9757092828 0.8664340843 0.2532533664 0.8898540716 0.9952213003 0.9386639921 0.1148355157 0.5633433899 0.9953682522 
         217          218          219          220          221          222          223          224          225          226          227          228 
0.8220629685 0.4678937697 0.8341476617 0.9988216889 0.4619254824 0.3015951494 0.8271566631 0.9958701456 0.0908555023 0.1479791227 0.5839121868 0.9965542913 
         229          230          231          232          233          234          235          236          237          238          239          240 
0.3360175342 0.5210814759 0.7041523007 0.1574328794 0.7167084702 0.9085163205 0.5750819964 0.7907836823 0.2903124871 0.2496516390 0.7512211830 0.9932190521 
         241          242          243          244          245          246          247          248          249          250          251          252 
0.9494375932 0.2150206593 0.9065026326 0.9862996131 0.3907589811 0.9902647508 0.0649097266 0.9943444535 0.3017236483 0.2778985259 0.2817547489 0.1434432582 
         253          254          255          256          257          258          259          260          261          262          263          264 
0.9699393522 0.7866161085 0.6611108085 0.9436664151 0.3907589811 0.2903124871 0.9881382560 0.9342438841 0.1566017312 0.7906154359 0.9918764532 0.1340122671 
         265          266          267          268          269          270          271          272          273          274          275          276 
0.9006481985 0.7511451767 0.4407976568 0.2718463965 0.9734407562 0.3212349833 0.9048856231 0.9814232128 0.9643911681 0.6580872789 0.8417214463 0.9771147877 
         277          278          279          280          281          282          283          284          285          286          287          288 
0.2978349530 0.9143962282 0.7157628430 0.5930814635 0.9879239941 0.1264245176 0.2439692987 0.5754297571 0.9830925403 0.3482510980 0.1472402178 0.9739103703 
         289          290          291          292          293          294          295          296          297          298          299          300 
```

```
0.5361390512 0.2362493659 0.8422593019 0.9172059495 0.8812930882 0.2940174792 0.0822201412 0.1237985139 0.5406925439 0.8156972404 0.7488872367 0.9378722479 
         301          302          303          304          305          306          307          308          309          310          311          312 
0.9125644164 0.8349865425 0.7264483182 0.3093694387 0.2463106599 0.3706914490 0.9850049248 0.9221398204 0.6539067426 0.1238205710 0.4304449557 0.5720867514 
         313          314          315          316          317          318          319          320          321          322          323          324 
0.1238205710 0.9814046688 0.3400719802 0.3081594368 0.9781742488 0.1041821537 0.3497701658 0.1502781823 0.9897244469 0.9406947975 0.3017236483 0.4216832745 
         325          326          327          328          329          330          331          332          333          334          335          336 
0.9665826970 0.2928841693 0.4724580048 0.4755619812 0.5941548142 0.9049206170 0.1698835472 0.9491843547 0.4679443921 0.3538091288 0.9624546177 0.0923803233 
         337          338          339          340          341          342          343          344          345          346          347          348 
0.9985604238 0.2180600129 0.1449797052 0.7751880043 0.9724984978 0.1942273199 0.1962152036 0.5529082908 0.9251241182 0.3160263686 0.9389662874 0.9810786565 
         349          350          351          352          353          354          355          356          357          358          359          360 
0.4008289211 0.2275774885 0.5839615859 0.2791194466 0.6103509910 0.1257997121 0.4228183275 0.9529061396 0.9477197527 0.3907589811 0.4052371748 0.3706914490 
         361          362          363          364          365          366          367          368          369          370          371          372 
0.9198865833 0.7339476434 0.5438836013 0.5119813898 0.7846220734 0.9050255295 0.9354342669 0.8399191235 0.2978349530 0.7231587344 0.9573370502 0.9917804973 
         373          374          375          376          377          378          379          380          381          382          383          384 
0.2928420615 0.9981090037 0.2496516390 0.3280707845 0.6928774706 0.2891458148 0.6073861631 0.8579953025 0.4588831277 0.5317820347 0.9375172688 0.5453904942 
         385          386          387          388          389          390          391          392          393          394          395          396 
0.8728261103 0.9528878873 0.5619483031 0.3496314503 0.4817010666 0.4666343501 0.5317820347 0.3836506554 0.3822149263 0.9643911681 0.1991209873 0.4039693951 
         397          398          399          400 
0.6162818406 0.4485971766 0.0620164326 0.5348805893 
> |
```

```
> cfmatrix
   pred
Act FALSE TRUE
  0   237   20
  1    39  104
```

```
> Acc=(cfmatrix[[1,1]]+cfmatrix[[2,2]])/sum(cfmatrix)
> Acc
[1] 0.8525
> |
```

Inference

After building the logistic regression model, have calculated the accuracy of the model, predicted accuracy for the given model has produced 85 % as its greater than 70% , we can consider this as the **best-fit model** for the given dataset.

Program

```
rm(list=ls())
mydata<-read.csv("D:/6th Sem Works/A2- EDA/LAB/Lab5/Social_Network_Ads.csv")
mydata$Gender<-as.factor(mydata$Gender)
mydata$Purchased<-as.factor(mydata$Purchased)
mymodel <- glm(Purchased ~ Age+Gender+EstimatedSalary, data=mydata, family='binomial')
res<-predict(mymodel,mydata,type='response')
res
cfmatrix<-table(Act=mydata$Purchased, pred=res>0.5)
cfmatrix
Acc=(cfmatrix[[1,1]]+cfmatrix[[2,2]])/sum(cfmatrix)
Acc
```