Doug Woodward

CS613 HW 7

|  |  |
| --- | --- |
| # | Answer |
| 1 | Let e^B0+B1x = y  p(x) = y / 1+y  1/p(x) = (1+y)/y  1/p(x) = 1+1/y  y = 1/(1/p(x))-1  y = 1/1-p(x)/p(x)  y = p(x)/1-p(x)  e^B0+B1x = p(x)/1-p(x) |
| 6 | p(x) = (e^-6+0.05\*40+3.5)/(1+e^-6+0.05\*40+3.5)) = 0.3775  e^-6+0.05\*x+3.5)/(1+e^-6+0.05\*x+3.5)) = 0.5  50 hours |
| 9 | p(x)/1-p(x) = exp(B0 + B1X)  p(x)/1-p(x) = 0.37  p(x) = 0.27  p(x) = 0.16  0.16/1-0.16 = 0.19 |
| 10 | 10.png  Year Lag1 Lag2 Lag3 Lag4 Lag5 Volume Today  Year 1.00000000 -0.032289274 -0.03339001 -0.03000649 -0.031127923 -0.030519101 0.84194162 -0.032459894  Lag1 -0.03228927 1.000000000 -0.07485305 0.05863568 -0.071273876 -0.008183096 -0.06495131 -0.075031842  Lag2 -0.03339001 -0.074853051 1.00000000 -0.07572091 0.058381535 -0.072499482 -0.08551314 0.059166717  Lag3 -0.03000649 0.058635682 -0.07572091 1.00000000 -0.075395865 0.060657175 -0.06928771 -0.071243639  Lag4 -0.03112792 -0.071273876 0.05838153 -0.07539587 1.000000000 -0.075675027 -0.06107462 -0.007825873  Lag5 -0.03051910 -0.008183096 -0.07249948 0.06065717 -0.075675027 1.000000000 -0.05851741 0.011012698  Volume 0.84194162 -0.064951313 -0.08551314 -0.06928771 -0.061074617 -0.058517414 1.00000000 -0.033077783  Today -0.03245989 -0.075031842 0.05916672 -0.07124364 -0.007825873 0.011012698 -0.03307778 1.000000000   1. There is a trend in volume 2. Coefficients:   Estimate Std. Error z value Pr(>|z|)  (Intercept) 0.26686 0.08593 3.106 0.0019 \*\*  Lag1 -0.04127 0.02641 -1.563 0.1181  Lag2 0.05844 0.02686 2.175 0.0296 \*  Lag3 -0.01606 0.02666 -0.602 0.5469  Lag4 -0.02779 0.02646 -1.050 0.2937  Lag5 -0.01447 0.02638 -0.549 0.5833  Volume -0.02274 0.03690 -0.616 0.5377  ---   None are very significant, Lag2 has some small significance, however, 0.02 is still fairly high.   1. Direction  pred.glm Down Up   Down 54 48   Up 430 557  Correct rate is 56%, and the model is correct most of the time when market goes up.   Coefficients:   Estimate Std. Error z value Pr(>|z|)  (Intercept) 0.20326 0.06428 3.162 0.00157 \*\*  Lag2 0.05810 0.02870 2.024 0.04298 \*    Direction2010  glm2.pred Down Up   Down 9 5   Up 34 56  Predictions are correct 62.5% of the time now.  Code:  NORMAL googleAnalytics.r[+] r utf-8[unix] 50% ☰ 3/6㏑ : 1  library(ISLR)  attach(Weekly)  dim(Weekly)  pairs(Weekly)  cor(Weekly[,-9])  glm.fits = glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume, data=Weekly, family=binomial)  summary(glm.fits)  glm.probs = predict(glm.fits, type="response")  pred.glm = rep("Down", length(glm.probs))  pred.glm[glm.probs > 0.5]="Up"  table(pred.glm, Direction)  train=(Year<2009)  Weekly.2010 = Weekly[!train,]  Direction2010 = Direction[!train]  glm.fit2 = glm(Direction ~ Lag2,data=Weekly,family=binomial,subset=train)  summary(glm.fit2)  glm2.probs=predict(glm.fit2,Weekly.2010,type="response")  glm2.pred = rep("Down", length(glm2.probs))  glm2.pred[glm2.probs > 0.5]="Up"  table(glm2.pred, Direction2010)  ~ |
| 5 | E=((1,6),(2,5),(3,4),(4,3),(5,2),(6,1))  P(E) = 6/36 = 16.7%  Odds(E) = P(E)/1-P(E) = 16.7/1-16.7 = 20% |
| 6 | From 4.3, we can see that p(x)/1-p(x) = e^b0+b1x1, if x1 were to increase by1, then 1 more b1 factors into the odds, therefore we can say that if an independent predictor variable increases by 1, its corresponding coefficient is increased by one in the response. |
| 7 | 0.6959309  Code:  ga <- read.csv(file="07-P7Train.csv") glm.fits = glm(Converted~Pageviews+AvgSessionLength+SessionCount, data=ga, family=binomial)  newdata = data.frame(Pageviews = 13, AvgSessionLength = 16.8, SessionCount = 24)  predict(glm.fits, newdata, type="response") |