Assignment 3

Yuqi Zhou
A20423555

Question 1:

a). The Entropy metric for the root node:1.0744900777690845

b). 15 possible binary-splits that can be generated from the CreditCard predictor.

c). The Entropy metric for each possibly binary split that can be generated from the CreditCard predictor.

Index	combination	entropy
0	(American Express,) , (Discover,MasterCard,Visa,Others,)	1.073
1	(Discover,) , (American Express,MasterCard,Visa,Others,)	1.07214
2	(MasterCard,) , (American Express,Discover,Visa,Others,)	1.07443
3	(Visa,) , (American Express,Discover,MasterCard,Others,)	1.07204
4	(Others,) , (American Express,Discover,MasterCard,Visa,)	1.07366
5	(American Express,Discover,) , (MasterCard,Visa,Others,)	1.07088
6	(American Express,MasterCard,) , (Discover,Visa,Others,)	1.07389
7	(American Express,Visa,) , (Discover,MasterCard,Others,)	1.07338
8	(American Express,Others,) , (Discover,MasterCard,Visa,)	1.07375
9	(Discover,MasterCard,) , (American Express,Visa,Others,)	1.0727
10	(Discover,Visa,) , (American Express,MasterCard,Others,)	1.0742
11	(Discover,Others,) , (American Express,MasterCard,Visa,)	1.07329
12	(MasterCard,Visa,) , (American Express,Discover,Others,)	1.07225
13	(MasterCard,Others,) , (American Express,Discover,Visa,)	1.07414
14	(Visa,Others,) , (American Express,Discover,MasterCard,)	1.07084

d). The optimal split for the CreditCard predictor: (Visa,Others) , (American Express,Discover,MasterCard)

e). 63 possible binary-splits that can be generated from the JobCategory predictor.

f). The Entropy metric for each possibly binary split that can be generated from the JobCategory predictor.

Index	combination	entropy
0	(Agriculture,) , (Crafts,Labor,Missing,Professional,Sales,Service,)	1.0743
1	(Crafts,) , (Agriculture,Labor,Missing,Professional,Sales,Service,)	1.07372
2	(Labor,) , (Agriculture,Crafts,Missing,Professional,Sales,Service,)	1.07412
3	(Missing,) , (Agriculture,Crafts,Labor,Professional,Sales,Service,)	1.07381
4	(Professional,) , (Agriculture,Crafts,Labor,Missing,Sales,Service,)	1.07309
5	(Sales,) , (Agriculture,Crafts,Labor,Missing,Professional,Service,)	1.07245
6	(Service,) , (Agriculture,Crafts,Labor,Missing,Professional,Sales,)	1.07413
7	(Agriculture,Crafts,) , (Labor,Missing,Professional,Sales,Service,)	1.07384
8	(Agriculture,Labor,) , (Crafts,Missing,Professional,Sales,Service,)	1.0742
9	(Agriculture,Missing,) , (Crafts,Labor,Professional,Sales,Service,)	1.07435
10	(Agriculture,Professional,) , (Crafts,Labor,Missing,Sales,Service,)	1.0736
11	(Agriculture,Sales,) , (Crafts,Labor,Missing,Professional,Service,)	1.07201
12	(Agriculture,Service,) , (Crafts,Labor,Missing,Professional,Sales,)	1.0744
13	(Crafts,Labor,) , (Agriculture,Missing,Professional,Sales,Service,)	1.07446
14	(Crafts,Missing,) , (Agriculture,Labor,Professional,Sales,Service,)	1.07391
15	(Crafts,Professional,) , (Agriculture,Labor,Missing,Sales,Service,)	1.07274
16	(Crafts,Sales,) , (Agriculture,Labor,Missing,Professional,Service,)	1.07256
17	(Crafts,Service,) , (Agriculture,Labor,Missing,Professional,Sales,)	1.07405
18	(Labor,Missing,) , (Agriculture,Crafts,Professional,Sales,Service,)	1.07399
19	(Labor,Professional,) , (Agriculture,Crafts,Missing,Sales,Service,)	1.07319
20	(Labor,Sales,) , (Agriculture,Crafts,Missing,Professional,Service,)	1.07257
21	(Labor,Service,) , (Agriculture,Crafts,Missing,Professional,Sales,)	1.07385
22	(Missing,Professional,) , (Agriculture,Crafts,Labor,Sales,Service,)	1.07304
23	(Missing,Sales,) , (Agriculture,Crafts,Labor,Professional,Service,)	1.07252
24	(Missing,Service,) , (Agriculture,Crafts,Labor,Professional,Sales,)	1.07403
25	(Professional,Sales,) , (Agriculture,Crafts,Labor,Missing,Service,)	1.07438

Index	combination	entropy
26	(Professional,Service,) , (Agriculture,Crafts,Labor,Missing,Sales,)	1.07244
27	(Sales,Service,) , (Agriculture,Crafts,Labor,Missing,Professional,)	1.07345
28	(Agriculture,Crafts,Labor,) , (Missing,Professional,Sales,Service,)	1.07446
29	(Agriculture,Crafts,Missing,) , (Labor,Professional,Sales,Service,)	1.074
30	(Agriculture,Crafts,Professional,) , (Labor,Missing,Sales,Service,)	1.07314
31	(Agriculture,Crafts,Sales,) , (Labor,Missing,Professional,Service,)	1.07207
32	(Agriculture,Crafts,Service,) , (Labor,Missing,Professional,Sales,)	1.07423
33	(Agriculture,Labor,Missing,) , (Crafts,Professional,Sales,Service,)	1.07412
34	(Agriculture,Labor,Professional,) , (Crafts,Missing,Sales,Service,)	1.0736
35	(Agriculture,Labor,Sales,) , (Crafts,Missing,Professional,Service,)	1.0721
36	(Agriculture,Labor,Service,) , (Crafts,Missing,Professional,Sales,)	1.07408
37	(Agriculture,Missing,Professional,) , (Crafts,Labor,Sales,Service,)	1.07357
38	(Agriculture,Missing,Sales,) , (Crafts,Labor,Professional,Service,)	1.07209
39	(Agriculture,Missing,Service,) , (Crafts,Labor,Professional,Sales,)	1.07434
40	(Agriculture,Professional,Sales,) , (Crafts,Labor,Missing,Service,)	1.07421
41	(Agriculture,Professional,Service,) , (Crafts,Labor,Missing,Sales,)	1.07298
42	(Agriculture,Sales,Service,) , (Crafts,Labor,Missing,Professional,)	1.07308
43	(Crafts,Labor,Missing,) , (Agriculture,Professional,Sales,Service,)	1.07446
44	(Crafts,Labor,Professional,) , (Agriculture,Missing,Sales,Service,)	1.07311
45	(Crafts,Labor,Sales,) , (Agriculture,Missing,Professional,Service,)	1.07291
46	(Crafts,Labor,Service,) , (Agriculture,Missing,Professional,Sales,)	1.07425
47	(Crafts,Missing,Professional,) , (Agriculture,Labor,Sales,Service,)	1.07276
48	(Crafts,Missing,Sales,) , (Agriculture,Labor,Professional,Service,)	1.07269
49	(Crafts,Missing,Service,) , (Agriculture,Labor,Professional,Sales,)	1.07408
50	(Crafts,Professional,Sales,) , (Agriculture,Labor,Missing,Service,)	1.07397
51	(Crafts,Professional,Service,) , (Agriculture,Labor,Missing,Sales,)	1.07212

52	(Crafts,Sales,Service,) , (Agriculture,Labor,Missing,Professional,)	1.07351
53	(Labor,Missing,Professional,) , (Agriculture,Crafts,Sales,Service,)	1.07308
54	(Labor,Missing,Sales,) , (Agriculture,Crafts,Professional,Service,)	1.07258
55	(Labor,Missing,Service,) , (Agriculture,Crafts,Professional,Sales,)	1.07372
56	(Labor,Professional,Sales,) , (Agriculture,Crafts,Missing,Service,)	1.07427
57	(Labor,Professional,Service,) , (Agriculture,Crafts,Missing,Sales,)	1.07222
58	(Labor,Sales,Service,) , (Agriculture,Crafts,Missing,Professional,)	1.07318
59	(Missing,Professional,Sales,) , (Agriculture,Crafts,Labor,Service,)	1.0744
60	(Missing,Professional,Service,) , (Agriculture,Crafts,Labor,Sales,)	1.07235
61	(Missing,Sales,Service,) , (Agriculture,Crafts,Labor,Professional,)	1.07347
62	(Professional,Sales,Service,) , (Agriculture,Crafts,Labor,Missing,)	1.07448

g). The optimal split for the JobCategory predictor:
(Agriculture,Sales) , (Crafts,Labor,Missing,Professional,Service)

h). Because the entropy(1.0708382285522746) of the optimal split for the CreditCard predictor is smaller than the entropy(1.0720111150297396) of the optimal split for the JobCategory predictor, I will choose the CreditCard predictor for producing the second layer of my decision tree.

Question2:
a). Start with a model with only the Intercept term. Two parameters are in this model.

b). The marginal counts of the categories of the target variable A:
category0: 143691
category1: 426067
category2: 95491

c).The maximum likelihood estimates of the predicted probabilities πi1:
0.21599581510081187, 0.64046244338586, 0.1435417415133281
d).The log-likelihood value of this Intercept-only model: -595406.7618844224
e). The maximum likelihood estimates of the Intercept terms β_j0,j=1,2,3: 0, 1.0869314502545038, -0.4086331584354002

f). The contingency table where group_size, homeowner, and married_couple are on the row dimension, and A is on the column dimension.

Index	0	1	2	All
(1, 0, 0)	25.7784	59.1461	15.0755	100
(1, 0, 1)	32.142	51.699	16.1591	100
(1, 1, 0)	18.0242	68.6295	13.3463	100
(1, 1, 1)	22.1715	62.3997	15.4288	100
(2, 0, 0)	27.2708	55.5946	17.1347	100
(2, 0, 1)	20.5508	64.5793	14.8699	100
(2, 1, 0)	25.3318	59.4854	15.1829	100
(2, 1, 1)	16.0672	70.2073	13.7255	100
(3, 0, 0)	27.5591	66.9291	5.51181	100
(3, 0, 1)	23.3002	59.5111	17.1887	100
(3, 1, 0)	25.9953	59.4848	14.5199	100
(3, 1, 1)	26.1579	56.3064	17.5357	100
(4, 0, 0)	100	0	0	100
(4, 0, 1)	18.75	67.8571	13.3929	100
(4, 1, 0)	48.4848	42.4242	9.09091	100
(4, 1, 1)	33.3333	53.0055	13.6612	100
('All', '', '')	21.5996	64.0462	14.3542	100

g). Based on the contingency table in e), just 0 category of the target variable is observed in each and every subpopulation, complete separation will occur but quasi-complete separation will not occur.

h). Category0 of A is used by the MNLogit function as the reference category.
The log-likelihood value of this model: -591936.7938327907
The number of parameters in the model: 18

i). The values of group_size, homeowner, and married_couple such that the odd Prob(A=1)/Prob(A = 0) will attain its maximum are: 2, 1, 1
The maximum odd value: 4.157240574942793

j). The odds ratio for group_size = 3 versus group_size = 1, and A = 2 versus A = 0: 0.8743305637174827

k). The odds ratio for group_size = 1 versus group_size = 3, and A = 2 versus A = 1: 0.6810748474897483