A20474413

Jorge Gonzalez Lopez **Assignment 3**

Question 1

# Answer

CS484 Spring 2021

March 05, 2021

Misclassification Rate = 938

# Explanation





Question 2

# Answer

0.948962149

# Explanation

# The entropy of the root value is calculated as the entropy of the target (‘CAR\_USE’) as follows:

# 

# To do so, the pandas’ function df['CAR\_USE'].value\_counts(normalize=True) gives back both probabilities.

# Answer



# Explanation

# To get the optimal splits, all the possible combinations of values in CAR\_TYPE, OCCUPATION and EDUCATION have been calculated. Then, for all of them, the split Entropy has been estimated by getting the probabilities of ‘CAR\_USE’ being private or commercial for the number of samples that met the left and right split requirements ( df[df[column].isin(split\_xx)).

# Finally, the split entropy is both split entropies escalated by their probability of occurrence.

# Answer

OCCUPATION (the decision rule: value.isin([‘Blue collar’, ‘Unknown’, ‘Student’]))

# Explanation

# The feature selected for the first split is OCCUPATION because it has the lower entropy split out of the three features (or the maximum gain).

# The values in the left branch are: Blue collar, Unknown or Student.

# The values in the right branch are: Professional, Manager, Clerical, Doctor, Lawyer or Home Maker

# Answer

Left Branch: EDUCATION (the decision rule: value.isin([‘Below High School’]))

Right Branch: CAR\_TYPE (the decision rule: value.isin([‘Minivan’, ‘SUV’, ‘Sports Car’]))

# Explanation

# To calculate the splits for the second layer, the same procedure as the first layer has to be followed. However, in this case, the dataset used to analyze the features is different in each branch:

# Left branch -> the subset from the original dataset that meets the first layer decision rule.

# Right branch -> the subset from the original dataset that does not meet the first layer decision rule.

# The features selected for the second layer split (the ones with the lower entropy split out of all the features (or the maximum gain)) are:

# Left Branch: EDUCATION (left branch: Below High School and right branch: High School, Bachelors, Masters, Doctors).

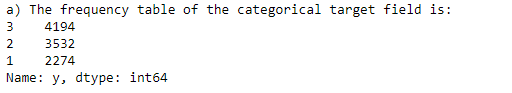
# Right branch: CAR\_TYPE (left branch: Minivan, SUV, Sports Car and right branch: Van, Pickup, Panel Truck.

# Answer



Question 3

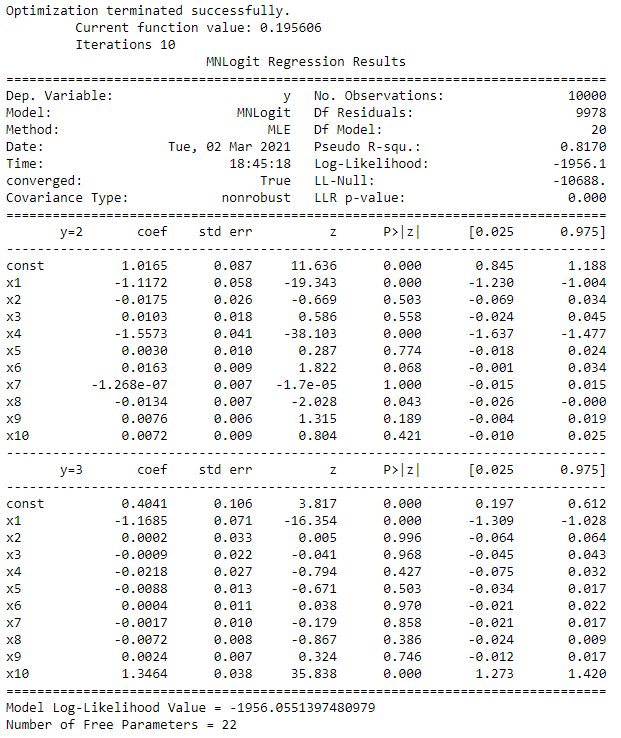
# Command Prompt Output



# Explanation

# Just by doing a value\_counts() of the target column.

# Command Prompt Output



# Explanation

# Get the target values (column ‘y’) and the input dataset (all other columns). Then, the intercept value is added to the input dataset as a constant column (all values equal to 1).

# Now, it is possible to calculate the number of free parameters with the rank of the input matrix multiply by the total number of output categories minus 1.

# Finally, the Multinomial Logistic Regression model is created with the input and target values and trained. Then, applying the function logit.loglike() to the values of the parameters of the model, the Log-Likelihood value is computed.

# 

# Command Prompt Output

Model 1 --> Removed feature: ['x7']

Model Log-Likelihood Value = -1956.0744283318356

Number of Free Parameters = 20

Deviance (Statistic, DF, Significance) 0.03857716747552331 2 0.9808962506876956

Akaike Information Criterion = 3952.1488566636713

Bayesian Information Criterion = 4096.355664103195

Model 2 --> Removed feature: ['x7', 'x3']

Model Log-Likelihood Value = -1956.30232999277

Number of Free Parameters = 18

Deviance (Statistic, DF, Significance) 0.45580332186864325 2 0.7962025538009445

Akaike Information Criterion = 3948.60465998554

Bayesian Information Criterion = 4078.3907866811114

Model 3 --> Removed feature: ['x7', 'x3', 'x2']

Model Log-Likelihood Value = -1956.5870772945038

Number of Free Parameters = 16

Deviance (Statistic, DF, Significance) 0.5694946034677741 2 0.7522043110300307

Akaike Information Criterion = 3945.1741545890077

Bayesian Information Criterion = 4060.5396005406265

Model 4 --> Removed feature: ['x7', 'x3', 'x2', 'x5']

Model Log-Likelihood Value = -1956.9994039437684

Number of Free Parameters = 14

Deviance (Statistic, DF, Significance) 0.8246532985290287 2 0.6621079636455474

Akaike Information Criterion = 3941.9988078875367

Bayesian Information Criterion = 4042.943573095203

Model 5 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9']

Model Log-Likelihood Value = -1957.9473654299018

Number of Free Parameters = 12

Deviance (Statistic, DF, Significance) 1.895922972266817 2 0.38753020449855446

Akaike Information Criterion = 3939.8947308598035

Bayesian Information Criterion = 4026.4188153235177

Model 6 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6']

Model Log-Likelihood Value = -1959.9572147343606

Number of Free Parameters = 10

Deviance (Statistic, DF, Significance) 4.019698608917679 2 0.13400886768610418

Akaike Information Criterion = 3939.914429468721

Bayesian Information Criterion = 4012.017833188483

Model 7 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6', 'x8']

Model Log-Likelihood Value = -1961.9026108507444

Number of Free Parameters = 8

Deviance (Statistic, DF, Significance) 3.8907922327675806 2 0.1429305949727081

Akaike Information Criterion = 3939.805221701489

Bayesian Information Criterion = 3997.4879446772984

Best model --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6', 'x8'] if we try to remove another feature: x1

Model Log-Likelihood Value = -2230.4083138546553

Number of Free Parameters = 6

Deviance (Statistic, DF, Significance) 537.0114060078217 2 2.4516294286020626e-117

Akaike Information Criterion = 4472.8166277093105

Bayesian Information Criterion = 4516.07866994116

# Explanation

# The features x1, x4 and x10 are the most important ones, as if one of them gets removed from the inputs, the Deviance significance values will get lower than 0.05.

# The backward method has been used. For that reason, the significance value is checked when each and every feature of the model is removed. Then, if the greater significance value among all the features is over 0.05, the feature gets eliminated from the model. These steps are carried out until the significance value of all the features remaining is less than 0.05.

# Command Prompt Output

# 

# Explanation

# The final model is computed by subtracting the least relevant columns from the dataset. In this case, those columns are x2, x3, x5, x6, x7, x8 and x9. These columns are found observing the Deviance significance value of the model when some features are removed. If this value is greater than 0.05, subtracting that feature would not greatly impact the model.

# Answer

For the models listed in c:

* The Akaike Information suggests the best model would be removing: ['x7', 'x3', 'x2', 'x5', 'x9'].
* The Bayesian Information suggests the best model would be removing: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6', 'x8']

# Explanation

The Akaike and Bayesian Information are other criteria for choosing the features to get the best model. It is intended to get the lowest possible values in both Criterion.

From the models listed in c), the Bayesian information suggests that the best model is Model 7 -> removing the same columns than while checking the p-value, because every successive model reduces its value. However, for the Akaike information, its value with Model 5 is bigger than with the previous one. Therefore, the model it suggests is Model 5 (even though Model 7 has a lower value).

Model 1 --> Removed feature: ['x7']

Akaike Information Criterion = 3952.1488566636713

Bayesian Information Criterion = 4096.355664103195

Model 2 --> Removed feature: ['x7', 'x3']

Akaike Information Criterion = 3948.60465998554

Bayesian Information Criterion = 4078.3907866811114

Model 3 --> Removed feature: ['x7', 'x3', 'x2']

Akaike Information Criterion = 3945.1741545890077

Bayesian Information Criterion = 4060.5396005406265

Model 4 --> Removed feature: ['x7', 'x3', 'x2', 'x5']

Akaike Information Criterion = 3941.9988078875367

Bayesian Information Criterion = 4042.943573095203

Model 5 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9']

Akaike Information Criterion = 3939.8947308598035

Bayesian Information Criterion = 4026.4188153235177

Model 6 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6']

Akaike Information Criterion = 3939.914429468721

Bayesian Information Criterion = 4012.017833188483

Model 7 --> Removed feature: ['x7', 'x3', 'x2', 'x5', 'x9', 'x6', 'x8']

Akaike Information Criterion = 3939.805221701489

Bayesian Information Criterion = 3997.4879446772984