CS 484: Introduction to Machine Learning

Spring 2021 Assignment 3

# Question 1 (20 points)

An observation is misclassified if the predicted target category is different from the observed target category. The misclassification rate is the proportion of observations that are misclassified. The following diagram shows the classification tree for a binary target variable. The target categories are 0 and 1. Based on the diagram, please calculate the misclassification rate.

Diagram

Description automatically generated

## Ans:

There are five leaf nodes in the tree and in which two of the leaf nodes have misclassified samples of number .

# Question 2 (40 points)

You will train a classification tree to predict the usage of a car. The data is the claim\_history.csv that contains 10,302 observations. The analysis specifications are:

**Target Field**

* **CAR\_USE**. The car’s usage. This field has two categories, namely, *Commercial* and *Private*.

**Nominal Feature**

* **CAR\_TYPE**. The car’s type. This feature has six categories, namely, *Minivan*, *Panel Truck*, *Pickup*, *SUV*, *Sports Car*, and *Van*.
* **OCCUPATION**. The occupation of the car owner. This feature has nine categories, namely, *Blue Collar*, *Clerical*, *Doctor*, *Home Maker*, *Lawyer*, *Manager*, *Professional*, *Student*, and *Unknown*.

**Ordinal Feature**

* **EDUCATION**. The education level of the car owner. This feature has five ordered categories which are *Below High School* < *High School* < *Bachelors* < *Masters* < *Doctors*.

**Decision Tree Specifications**

* Use only the complete records.
* The maximum number of branches is two.
* The maximum depth is two.
* The split criterion is the Entropy metric.

Since the sklearn tree module does not handle string features well, you should write custom Python codes to find the optimal split for a string feature. Also, do not encode the nominal features into dummy columns. It is because your classification tree is not deep enough to let all the dummy columns be used for splitting. Please answer the following questions.

1. (5 points). What is the entropy value of the root node?
2. (10 points). Please list the optimal split (i.e., feature name, values in the two branches, and the split entropy ) for all three features in the first layer.
3. (5 points). Which feature is selected for splitting in the first layer? What are the values in the branches of the first layer?
4. (10 points). Which features are selected for splitting in the second layer? What are the values in the branches of the second layer?
5. (10 points). Describe the leaf (i.e., terminal) nodes in a table. Please include the decision rules, the counts of the target categories, and the predicted probabilities for CAR\_USE.

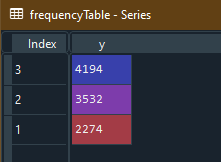
# Question 3 (40 points)

We provide you the sample\_v10.csv that contains 10,000 observations. This data contains a categorical variable **y** and ten continuous features are **x1**, **x2**, **x3**, **x4**, **x5**, **x6**, **x7**, **x8**, **x9**, and **x10**. You will then use this data to train a multinomial logistic regression model that always includes the Intercept term. To include only significant continuous features in the model, you will use the Backward Selection method to determine the list of significant continuous features. The threshold for test significance is 0.05.

1. (5 points). Show the frequency table of the categorical target field.

### Ans:

The frequency table of the categorical target field is shown below:



Figure

1. (5 points). What is the initial model in the Backward Selection method? Please also show the log-likelihood value and the number of free parameters.

### Ans:

* The initial model in the Backward Selection method is the one that includes all the parameters. The R style formula for this model can be shown as:
* From the python code we can get the following information:
* Model Log-Likelihood Value = -1956.055139748098
* Number of Free Parameters = 22

1. (20 points). Please show the step summary of the Backward Selection method. The step summary should include the name of the removed feature, the log-likelihood value of the reduced model, the number of free parameters of the reduced model, the Deviance test statistic, the Deviance degree of freedom, and the Deviance significance value.

### Ans:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| 0 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 | 22 | -1956.05514 |  |  |  |
| 1.1 | Intercept + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 1.2 | Intercept + x1 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 | 20 | -1956.331061 | 0.55184178 | 2 | 0.758872963 |
| 1.3 | Intercept + x1 + x2 + x4 + x5 + x6 + x7 + x8 + x9 + x10 | 20 | -1956.280363 | 0.45044599 | 2 | 0.798338173 |
| 1.4 | Intercept + x1 + x2 + x3 + x5 + x6 + x7 + x8 + x9 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 1.5 | Intercept + x1 + x2 + x3 + x4 + x6 + x7 + x8 + x9 + x10 | 20 | -1956.470194 | 0.83010789 | 2 | 0.660304659 |
| 1.6 | Intercept + x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 1.7 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -1956.074428 | 0.03857717 | 2 | 0.980896251 |
| 1.8 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x9 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 1.9 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x10 | 20 | -1956.953231 | 1.79618191 | 2 | 0.407346562 |
| 1.10 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 1.7 (removing x7) has the largest Chi-Square Significance value. X7 will be removed from the model

| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| --- | --- | --- | --- | --- | --- | --- |
| 2.1 | Intercept + x2 + x3 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 2.2 | Intercept + x1 + x3 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -1956.331061 | 0.55184178 | 2 | 0.758872963 |
| 2.3 | Intercept + x1 + x2 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -1956.280363 | 0.45044599 | 2 | 0.798338173 |
| 2.4 | Intercept + x1 + x2 + x3 + x5 + x6 + x8 + x9 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 2.5 | Intercept + x1 + x2 + x3 + x4 + x6 + x8 + x9 + x10 | 20 | -1956.470194 | 0.83010789 | 2 | 0.660304659 |
| 2.6 | Intercept + x1 + x2 + x3 + x4 + x5 + x8 + x9 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 2.7 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x9 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 2.8 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x8 + x10 | 20 | -1956.953231 | 1.79618191 | 2 | 0.407346562 |
| 2.9 | Intercept + x1 + x2 + x3 + x4 + x5 + x6 + x8 + x9 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 2.3 (removing x3) has the largest Chi-Square Significance value. X3 will be removed from the model

| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| --- | --- | --- | --- | --- | --- | --- |
| 3.1 | Intercept + x2 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 3.2 | Intercept + x1 + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -1956.331061 | 0.55184178 | 2 | 0.758872963 |
| 3.3 | Intercept + x1 + x2 + x5 + x6 + x8 + x9 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 3.4 | Intercept + x1 + x2 + x4 + x6 + x8 + x9 + x10 | 20 | -1956.470194 | 0.83010789 | 2 | 0.660304659 |
| 3.5 | Intercept + x1 + x2 + x4 + x5 + x8 + x9 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 3.6 | Intercept + x1 + x2 + x4 + x5 + x6 + x9 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 3.7 | Intercept + x1 + x2 + x4 + x5 + x6 + x8 + x10 | 20 | -1956.953231 | 1.79618191 | 2 | 0.407346562 |
| 3.8 | Intercept + x1 + x2 + x4 + x5 + x6 + x8 + x9 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 3.2 (removing x2) has the largest Chi-Square Significance value. X2 will be removed from the model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| 4.1 | Intercept + x4 + x5 + x6 + x8 + x9 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 4.2 | Intercept + x1 + x5 + x6 + x8 + x9 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 4.3 | Intercept + x1 + x4 + x6 + x8 + x9 + x10 | 20 | -1956.470194 | 0.83010789 | 2 | 0.660304659 |
| 4.4 | Intercept + x1 + x4 + x5 + x8 + x9 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 4.5 | Intercept + x1 + x4 + x5 + x6 + x9 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 4.6 | Intercept + x1 + x4 + x5 + x6 + x8 + x10 | 20 | -1956.953231 | 1.79618191 | 2 | 0.407346562 |
| 4.7 | Intercept + x1 + x4 + x5 + x6 + x8 + x9 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 4.3 (removing x5) has the largest Chi-Square Significance value. X5 will be removed from the model

| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| --- | --- | --- | --- | --- | --- | --- |
| 5.1 | Intercept + x4 + x6 + x8 + x9 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 5.2 | Intercept + x1 + x6 + x8 + x9 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 5.3 | Intercept + x1 + x4 + x8 + x9 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 5.4 | Intercept + x1 + x4 + x6 + x9 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 5.5 | Intercept + x1 + x4 + x6 + x8 + x10 | 20 | -1956.953231 | 1.79618191 | 2 | 0.407346562 |
| 5.6 | Intercept + x1 + x4 + x6 + x8 + x9 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 5.5 (removing x9) has the largest Chi-Square Significance value. X9 will be removed from the model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| 6.1 | Intercept + x4 + x6 + x8 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 6.2 | Intercept + x1 + x6 + x8 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 6.3 | Intercept + x1 + x4 + x8 + x10 | 20 | -1958.083555 | 4.05683106 | 2 | 0.131543783 |
| 6.4 | Intercept + x1 + x4 + x6 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 6.5 | Intercept + x1 + x4 + x6 + x8 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 6.3 (removing x6) has the largest Chi-Square Significance value. X6 will be removed from the model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| 7.1 | Intercept + x4 + x8 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 7.2 | Intercept + x1 + x8 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 7.3 | Intercept + x1 + x4 + x10 | 20 | -1958.114765 | 4.11924974 | 2 | 0.12750179 |
| 7.4 | Intercept + x1 + x4 + x8 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since for index 7.3 (removing x8) has the largest Chi-Square Significance value. X8 will be removed from the model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Index** | **Model Form** | **Number of Free Parameters** | **Log-Likelihood** | **Deviance** | **Degrees of Freedom** | **Chi-Square Significance** |
| 8.1 | Intercept + x4 + x10 | 20 | -2225.27135 | 538.43242 | 2 | 1.20E-117 |
| 8.2 | Intercept + x1 + x10 | 20 | -5780.494365 | 7648.87845 | 2 | 0 |
| 8.3 | Intercept + x1 + x4 | 20 | -8111.136792 | 12310.1633 | 2 | 0 |

* Since all Chi-Square Significance values are not greater than 0.05, removing any of the rest will reduce model goodness-of-fit.
* Therefore, no predictors can be removed.
* The Backward Selection stops at Model : Intercept + x1 + x4 + x10

1. (5 points). What is the final model suggested by the Backward Selection method?
2. (5 points). Please calculate the Akaike Information Criterion and the Bayesian Information Criterion for all the models that you listed in (c). What model will each criterion suggest?