CS 484 Introduction to Machine Learning

Fall Semester 2022 Final Examination

Instruction

1. Calculate your answers using all the available precision
2. If the final numeric answer has more than four decimal places, then round the numeric answer to the nearest fourth decimal place. Otherwise, please give the exact value.
3. You can attempt this test once.
4. You must submit your answers on Blackboard no later than 4 AM on December 5, 2022.

Question 1 (5 points)

We use the data from the Q1.csv to construct a histogram for the variable x. Suppose the bin width is 5. What is the empirical density value for x = 87.5? Please round your answer to the fourth decimal place.

Question 2 (5 points)

What is the interquartile range of the variable x in the data from the Q1.csv? Please round your answer to the fourth decimal place.

Question 3 (5 points)

I calculated the Elbow values and the Silhouette values for my 1-cluster to 10-cluster solutions. Based on these values, what do you suggest for the number of clusters? An integer answer is expected.

Number of Clusters	Elbow Value	Silhouette
1	579857.9543	N/A
2	532455.2722	0.5391
3	493218.0813	0.5300
4	433215.8150	0.5479
5	430290.4574	0.5411
6	412804.9312	0.5140
7	409729.7423	0.5172
8	404285.7518	0.5081
9	378087.1355	0.5056
10	369686.6227	0.4984

Question 4 (5 points)

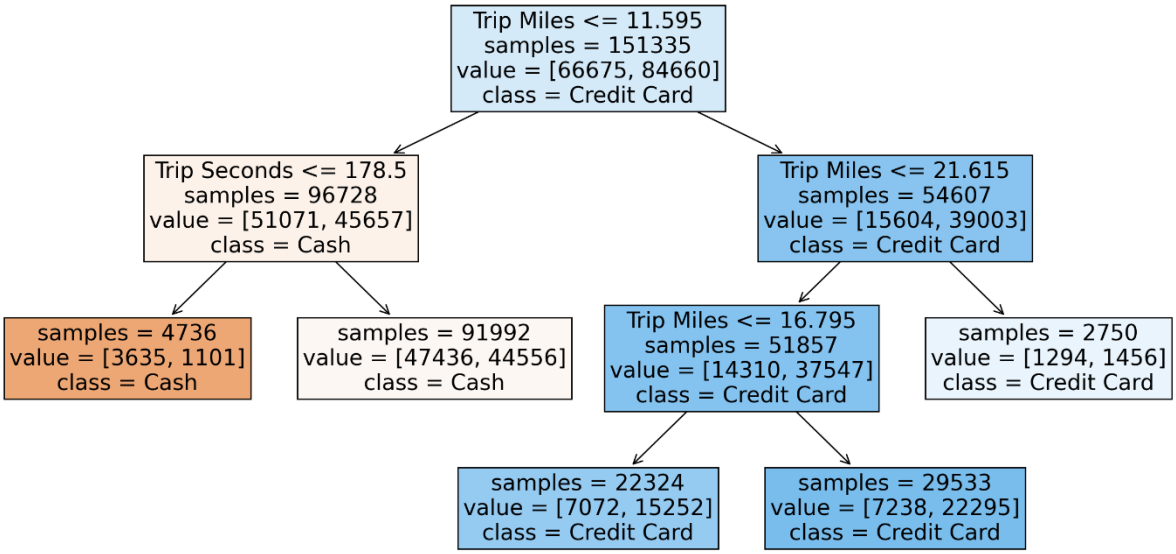
I invited ten friends to my home to watch a basketball game. My friends brought snacks and beverages along. The following table lists the items my friends brought. I noticed that a few of my friends brought Cheese, Soda, and Wings together. I am interested to measure the difference of {Cheese, Wings} and {Soda} appearing together and what would be expected if these two itemsets are statistically independent. Therefore, please calculate for me the Leverage of this association rule {Cheese, Wings} ➔ {Soda}.

Friend	**Items**
Andrew	Cheese, Cracker, Salsa, Soda, Tortilla, Wings
Betty	Cheese, Soda, Tortilla, Wings
Carl	Cheese, Ice Cream, Soda, Wings
Danny	Cheese, Ice Cream, Salsa, Tortilla, Wings
Emily	Pizza, Salsa, Soda, Tortilla, Wings
Frank	Cheese, Cracker, Ice Cream, Soda, Wings
Gary	Cracker, Tortilla
Henry	Ice Cream, Pizza, Tortilla
Irene	Cheese, Cracker, Soda
Jack	Cheese, Cracker, Pizza, Salsa, Wings

Question 5 (5 points)

We provided you with the following classification tree diagram. The label variable is Payment Type. It has two categories, namely, Cash and Credit Card. We did not use the predicted class suggested by the classification tree. Instead, we will predict Credit Card for Payment Type if the predicted probability that Payment Type is Credit Card is greater than 0.6. Otherwise, the predicted Payment Type will be Cash.

What is the accuracy rate? Please round your answer to the fourth decimal place.



Question 6 (5 points)

Suppose a categorical variable has eight categories. If we train a classification tree using this categorical variable as our label, what will be the maximum Gini value in theory?

A. 0.000
B. 0.125
C. 0.500
D. 0.875
E. 3.000

Question 7 (5 points)

Which of the following IS NOT a symptom of the Complete Separation phenomenon in the Logistic Regression model?

A. A Zero Log-Likelihood Value
B. Absolute Values of Parameter Estimates are Relatively Large
C. Standard Errors of Parameter Estimates are Relatively Large
D. Estimation Iteration May Not Converge
E. A Zero Likelihood Value

Question 8 (5 points)

We trained a classification tree to predict the outcome of a binary label variable. Suppose the Root Average Square Error value is zero. What can we expect for the Area Under Curve value?

A. -1.0
B. 0.0
C. 0.5
D. 1.0
E. Any Values Between 0 and 1

Question 9 (5 points)

Our Recommendation System uses a classification model to recommend products for our customers. Which of the followings is the probability that a customer will purchase the product that was recommended?

A. Sensitivity
B. Specificity
C. Precision
D. Recall
E. Accuracy

Question 10 (5 points)

We provide you a training data that have more than ten thousand observations. The label variable is continuous. You can train any prediction model using these observations. After you have settled down on a Multi-Layer Perception neural network model, we give you a test data that has only one observation. Besides asking you to provide us with the predicted value, we also need a 95% confidence interval for the predicted value. What model ensemble method will you use to meet our need?

A. Adaptive or Gradient Boosting on Training Data
B. Adaptive or Gradient Boosting on Testing Data
C. Bagging on Training Data
D. Bagging on Testing Data
E. Nothing Can Be Done

Question 11 (5 points)

You live in the San Francisco Bay area where earthquakes are not uncommon. Your house has a security alarm system against burglary, and it can be set off occasionally by an earthquake. Historically, there is a 6% chance that your house will be burglarized and there is a 2% chance that an earthquake will occur in your area. You can assume that the occurrences of burglary and earthquake are statistically independent. Based on your experience, your alarm will sound if the following events have occurred.

Alarm	Burglary	Earthquake	Prob(Alarm | Burglary, Earthquake)
TRUE	TRUE	TRUE	0.92
TRUE	FALSE	TRUE	0.96
TRUE	TRUE	FALSE	0.98
TRUE	FALSE	FALSE	0.0001

What is Prob(Burglary = True and Earthquake = False | Alarm Sounded = True), i.e., the conditional probability that your house has been burglarized but no earthquake has occurred provided the alarm has been sounded. Please round your answer to the fourth decimal place.

Question 12 (5 points)

Suppose we trained a multinomial logistic regression model using the Q12.csv. The label variable is CreditCard that has five categories. The categorical features in the model are Gender, MaritalStatus, and Retired. When Gender is 'Male', MaritalStatus is 'Married', and Retired is 'No', what is the predicted odds of having a Discovery card versus an American Express card? The predicted odds value is the predicted probability for the 'Discovery' category of CreditCard divided by that for the 'American Express' category of CreditCard.

Question 13 (5 points)

We have trained a Support Vector Machine model for a binary label variable. The continuous features are x and y. The equation of the hyperplane is $3 - 2x + 4y = 0$.

Suppose we created two new features $u = 2x + 1$ and $v = 4y + 2$. What is the formula of the separating hyperplane if we trained a Support Vector Machine model using the same binary label variable but u and v as the new features?

We require the Intercept to be a non-negative integer. Also, the three coefficients must be relatively prime (i.e., the three coefficients have no common factors except the integer one).

[Intercept] + [Coefficient_of_u] * u + [Coefficient_of_v] * v = 0

Question 14 (5 points)

We are planning for our annual marketing campaign to reach our desired respondents. Based on data that we have collected; we trained a classification model. After we calculated the predicted event probabilities from the model, we put respondents into ten deciles. The deciles are determined in descending order of the probabilities. We summarized the results in the following table. Our desired respondents are labelled Event.

Our campaign goal is to reach as many respondents as possible. To use our resources wisely, we only want to reach respondents who will at least twice more likely to respond than the overall sample. Therefore, we need to determine the maximum percent of respondents that we should reach. Starting from Decile 0, which is the last Decile of respondents we should consider? Please enter a non-negative integer.

Decile of Predicted Event Probabilities	Number of Respondents	
	Event	Non-Event
0	873	454
1	94	1234
2	121	1206
3	90	1238
4	135	1192
5	121	1207
6	124	1204
7	100	1227
8	55	1273
9	0	1327

Question 15 (5 points)

Suppose we trained a Support Vector Machine classification model using the Q15.csv. The label variable is group that has two categories. The continuous features in the model are x and y.

Our goal is to achieve the highest possible prediction accuracy (i.e., zero misclassification rate). What is the formula for the Separating Hyperplane or Hypercurve? Please enter the formula as a general equation and round all numeric values to the fourth decimal place.

Question 16 (5 points)

Suppose we trained a Naïve Bayes classification model using the Q12.csv. The label variable is **CreditCard** and it has five categories. The categorical features in the model are **Gender**, **MaritalStatus**, and **Retired**. Which category combination will yield the highest predicted probability that **CreditCard** is *American Express*? Please type the category AS-IS.

[**Gender**] (*Female* or *Male*)

[**MaritalStatus**] (*Married* or *Unmarried*)

[**Retired**] (*No* or *Yes*)

Question 17 (5 points)

We provide you with the following table of counts. Suppose we train a logistic model to predict Claim Indicator by Number of Children Driving. What is the Mcfadden's R-Square value? Please round your answer to the fourth decimal place.

Number of Children Driving	**Claim Indicator**	
	No	Yes
0	6,815	1,254
1	492	312
2	112	339
3+	3	51

Question 18 (5 points)

We will train two models using the 20 observations below. The binary target variable is y, and the interval predictor is x.

x	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
y	0	0	0	0	0	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1

The two models are:

1. Classification Tree, the splitting criterion is Entropy, the maximum depth is 5, and the random state is 60616.
2. Logistic, Intercept term included, the optimization method is newton, the maximum number of iterations is 100, the tolerance level is 1e-8.

After we built these two models, we will apply them on the following test data and then calculate the root average squared error metric.

x	0	1	2	3	4
y	0	0	1	1	1

Finally, compare and recommend the final model.

Please round your RASE values to the fourth decimal place.

RASE for Classification Tree = [RASE_Tree]

RASE for Logistic Model = [RASE_Logistic]

Recommend [Final_Model] (TREE or LOGISTIC)

Question 19 (5 points)

We will study how well a Support Vector Machine model can identify the nose feature on a face. The data is Face.csv. This data contains the two-dimensional coordinates of the face of a robot. The nose is Feature No. 3. It is surrounded by the right eye (Feature No. 1), the left eye (Feature No. 2), and the mouth (Feature No. 4). Since we are only interested in identifying the nose, please create the variable IS_NOSE. If the feature is a nose, then IS_NOSE is 1. Otherwise, IS_NOSE is 0. The predictors are the x-coordinate and the y-coordinate. The random seed is 20221225.

Transform the Cartesian coordinates into the Polar coordinates. The formulas are radius = sqrt(x*x + y*y) and angle = arctan(y/x) (-pi <= angle <= pi). Train another Support on the transformed data to identify the nose. The label variable is still IS_NOSE, but the predictors are radius and angle. What is the accuracy rate? Please round your answer to the fourth decimal place.

Question 20 (5 points)

We will study how well a Gradient Boosting model can identify the nose feature on a face. The data is Face.csv. This data contains the two-dimensional coordinates of the face of a robot. The nose is Feature No. 3. It is surrounded by the right eye (Feature No. 1), the left eye (Feature No. 2), and the mouth (Feature No. 4). Since we are only interested in identifying the nose, please create the variable IS_NOSE. If the feature is a nose, then IS_NOSE is 1. Otherwise, IS_NOSE is 0. The predictors are the x-coordinate and the y-coordinate.

The random seed is 20221225. The probability threshold for classification is 0.5. Please use the sklearn.gbm function. The incremental booster is a regression tree with maximum four leaves. After 10 iterations, what is the accuracy value? Please round your answer to the fourth decimal place.