CS 484 Introduction to Machine Learning

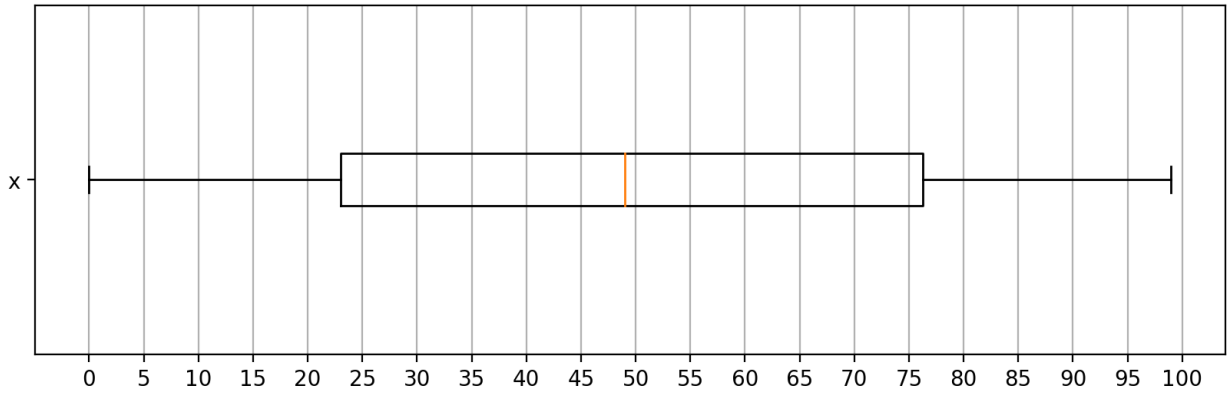
Spring 2021 Final Examination

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.

Question 1 (5 points)

Based on the following boxplot, which of the following value is the Interquartile Range?



Multiple Choice:

(A) 23

(B) 49

(C) 53.25

(D) 76.25

(E) 99

Question 2 (5 points)

Which of the following points is nearest to this focal point (2, 2) according to the Cosine Distance?

Multiple Choice:

(A) (0, 0)

(B) (0, 4)

(C) (4, 0)

(D) (4, 4)

(E) All of the Above

Question 3 (5 points)

We will train a Nearest Neighbor Regressor model on the following data. All three variables x1, x2, and y are continuous variables. The label variable is y. The input features are x1 and x2. The distance measure is Chebyshev. The error is y minus the predicted y. The criterion is the sum of absolute error. What is the number of neighbors that yields the smallest criterion?

x1	x2	y
0	0.6	-0.6
0.4	0.4	-0.6
0.7	0.8	0.6
0.5	0.2	1.8
0.5	0.8	1.2
0.6	0	1.2
0.3	0.2	1.4
0.1	0.6	0.6
0.8	0.8	1.8
0.8	0	1.6

Multiple Choice:

(A) 2

(B) 3

(C) 4

(D) 5

(E) 6

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.

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

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}.

Multiple Choice:

(A) 0.04

(B) 0.4

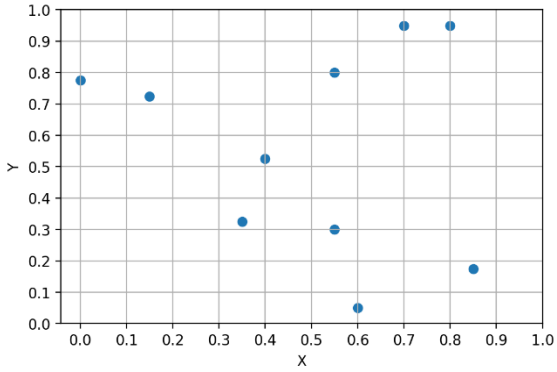
(C) 0.6667

(D) 1.1111

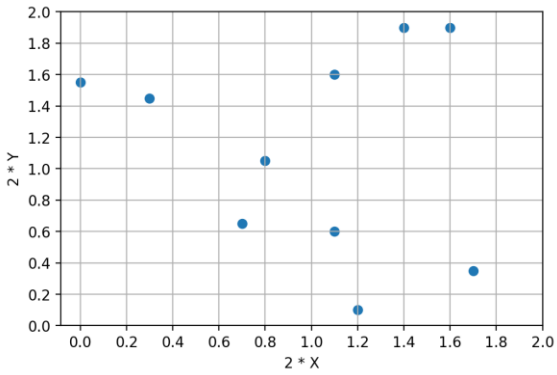
(E) 1.2

Question 5 (5 points)

A data table has 10 rows and 2 columns. The columns are named X and Y. The following scatterplot shows the spread of the ten points.



We discovered two clusters using the Euclidean distance. The Silhouette value of the two-cluster solution is 0.3843. Suppose we magnify both X and Y by a factor of two, i.e., X → 2*X and Y → 2*Y. The following scatterplot shows the resulting spread of the ten points.



Suppose we, again, discovered two clusters on the magnified data. Which of the following values would be the new Silhouette value?

Multiple Choice:

(A) -0.3843

(B) 0.1921

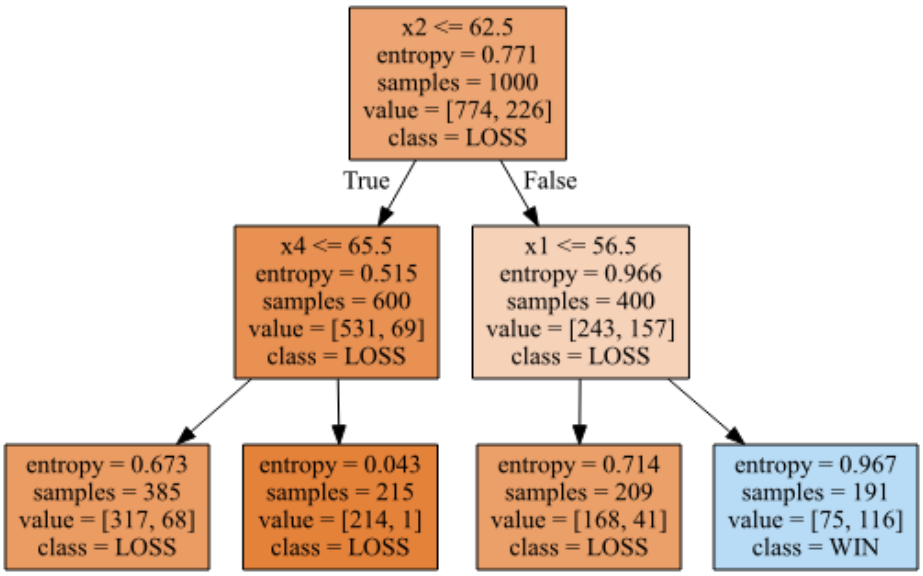
(C) 0.3843

(D) 0.7685

(E) Cannot Be Determined

Question 6 (5 points)

A research paper shows the following classification tree diagram. The label variable is binary and has two categories LOSS and WIN. The event category is WIN. Since the research paper does not provide me the Area Under Curve metric, I need to calculate the metric. What is the Area Under Curve value?



Multiple Choice:

(A) 0.1388

(B) 0.4584

(C) 0.5

(D) 0.7780

(E) Cannot Be Determined

Question 7 (5 points)

You will calculate the Cramer's V statistic to measure the association between two categorical features, namely, *Row* and *Column*. Instead of the original training data, you are given the following crosstabulation table.

Number of Observations		Column			
		1	2	3	4
Row	A	4,340	5,403	2,456	353
	B	8,095	16,156	10,798	2,371
	C	4,761	14,154	14,103	4,597
	D	813	3,636	5,307	2,657

Multiple Choice:

(A) 0.0289

(B) 0.1471

(C) 0.1699

(D) 53.7200

(E) Cannot Be Determined

Question 8 (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.

Earthquake	True	True	False	False
Burglary	True	False	True	False
Probability that the Alarm will sound	0.99	0.15	0.95	0.0001

Please calculate this quantity 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.

Multiple Choice:

(A) 0.0559

(B) 0.0684

(C) 0.5000

(D) 0.9316

(E) 0.9400

Question 9 (5 points)

The following table shows the observed target values and the predicted event probabilities from a model. What is the probability threshold that yields the highest Kolmogorov–Smirnov statistic? Please provide the exact answer.

Target	Predicted Event Probability	Target	Predicted Event Probability
Non-Event	0.0814	Event	0.4974
Non-Event	0.1197	Event	0.6732
Non-Event	0.1969	Event	0.6744
Non-Event	0.3505	Event	0.6836
Non-Event	0.3878	Event	0.7475
Non-Event	0.3940	Event	0.7828
Non-Event	0.4828		
Non-Event	0.4889		
Non-Event	0.5587		
Non-Event	0.5614		
Non-Event	0.6175		
Non-Event	0.6342		
Non-Event	0.6527		
Non-Event	0.6668		

Multiple Choice:

(A) 0.6342

(B) 0.6732

(C) 0.8333

(D) 0.8750

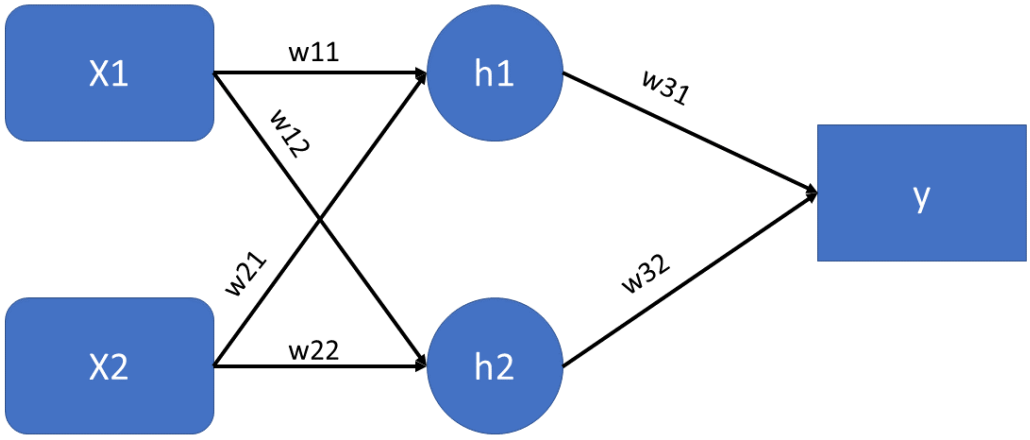
(E) 0.9583

Question 10 (5 points)

The logical XAND gate returns TRUE only when both arguments are the same (i.e., either both are TRUE, or both are FALSE). Otherwise, it returns FALSE. This can be represented by the following table.

X1	0	0	1	1
X2	0	1	0	1
X1 XAND X2	1	0	0	1

Consider a XAND neural network that has two neurons in a single hidden layer.



In the above diagram, h1 = (w11 * X1 + w21 * X2) >= c1, h2 = (w12 * X1 + w22 * X2) >= c2, and y = (w31 * h1 + w32 * h2) >= c3. Specify the six synaptic weights and the three threshold values such that the above neural network can implement the XAND function. The parameters must be integers, but we allow negative integers and zero. Check all the correct answers.

Multiple Answers:

Answer	w11	w12	w21	w22	w31	w32	c1	c2	c3
(A)	1	-1	1	-1	1	1	2	0	1
(B)	-1	1	-1	1	1	1	0	0	0
(C)	-1	1	-1	1	-1	-1	0	2	0
(D)	1	1	1	1	1	-1	2	1	0
(E)	1	1	1	1	-1	1	0	1	2

Question 11 (5 points)

Train a multinomial logistic regression model using the Q11.csv. The label variable is CreditCard that has five categories. The categorical features are ActiveLifestyle, Gender, MaritalStatus, and Retired. Use only non-missing values of these variables for training the model. Enter the categorical features into the model using the Forward Selection method. The entry threshold for significance is 0.01.

When the Forward Selection method stops, what is the last feature that enters into the model?

Multiple Choice:

(A) ActiveLifestyle
(B) CreditCard
(C) Gender
(D) MaritalStatus
(E) Retired

Question 12 (5 points)

Suppose we trained a multinomial logistic regression model using the Q11.csv. The label variable is CreditCard that has five categories. The categorical features in the model are Gender, MaritalStatus, and Retired.

When Gender is 'Female', MaritalStatus is 'Unmarried', and Retired is 'Yes', what are the predicted probabilities for each CreditCard category?

(A) American Express
(B) Discover
(C) MasterCard
(D) Others
(E) Visa

Question 13 (5 points)

Suppose we trained a Naïve Bayes classification model using the Q11.csv. The label variable is CreditCard that has five categories. The categorical features in the model are Gender, MaritalStatus, and Retired.

When Gender is 'Female', MaritalStatus is 'Unmarried', and Retired is 'Yes', what are the predicted probabilities for each CreditCard category?

(A) American Express

(B) Discover

(C) MasterCard

(D) Others

(E) Visa

Question 14 (5 points)

This is a follow-up to Question 13.

Suppose we trained a Naïve Bayes classification model using the Q11.csv. The label variable is CreditCard that has five categories. The categorical features in the model are Gender, MaritalStatus, and Retired.

Which category combination will yield the highest predicted probability of CreditCard is American Express?

Multiple Choice:

(A) Gender = Female, MaritalStatus = Unmarried, and Retired = Yes

(B) Gender = Female, MaritalStatus = Unmarried, and Retired = No

(C) Gender = Male, MaritalStatus = Married, and Retired = Yes

(D) Gender = Male, MaritalStatus = Married, and Retired = No

(E) Gender = Female, MaritalStatus = Married, and Retired = No

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?

Multiple Choice:

(A) 1000003.2 – 1.1014302 * x + 3.5540221 * y = 0

(B) -2.9960 + 1.6245 * x – 0.0005 * y = 0

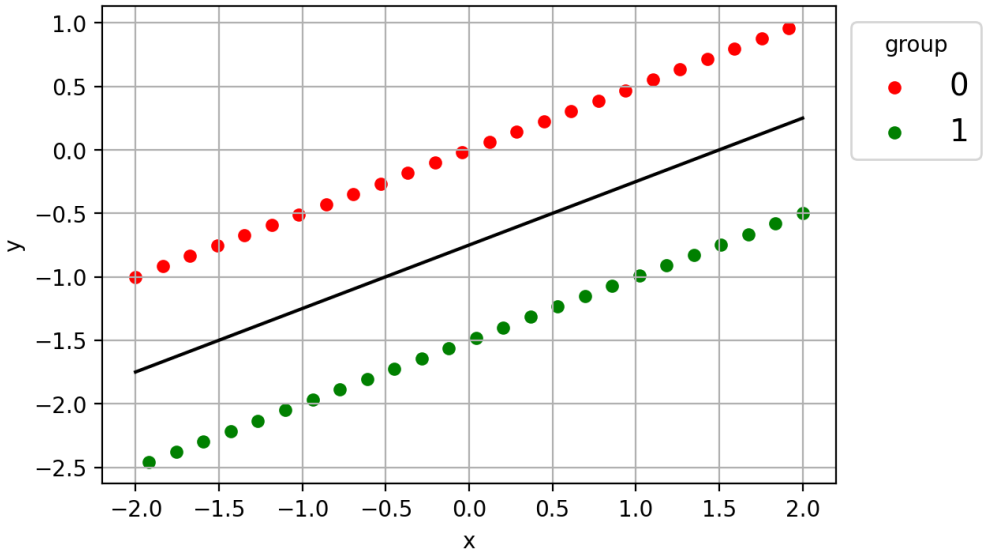
(C) x * x + y * y - 3.4051 = 0

(D) x * x + y * y - 1.8453 = 0

(E) 1.8453 – y = 0

Question 16 (5 points)

Suppose we have trained a Support Vector Machine model on a data. The label variable is group that has two categories 0 and 1. The continuous features are x and y. The following figure shows the observations and the separating hyperplane. The equation of the hyperplane is 3 - 2*x + 4*y = 0.



Suppose we created two new features u = 2 * x + 1 and v = 4 * y + 2. What is the formula of the separating hyperplane if we trained a Support Vector Machine model using the same group as the label 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 for u	
Coefficient for v	

Question 17 (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 provide you one test observation and ask you to provide us the predicted value. Besides, we also need a 95% confidence interval for the predicted value. What model ensemble method will you use to fulfill our request?

Multiple Choice:

(A) Adaptive or Gradient Boosting on Training Data

(B) Adaptive or Gradient Boosting on Testing Data

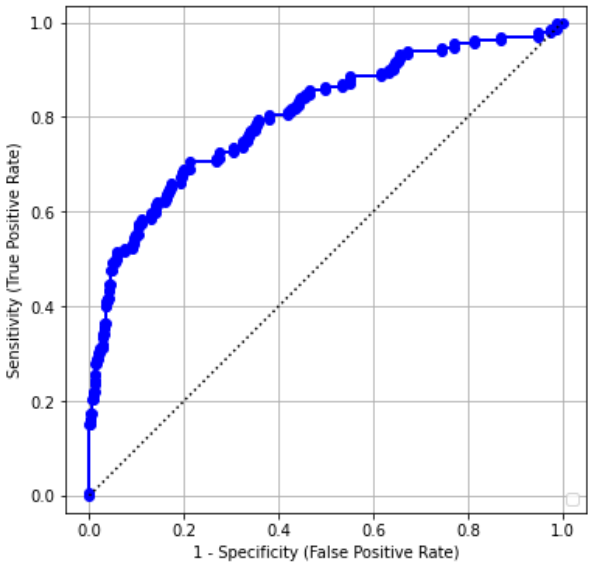
(C) Bagging on Training Data

(D) Bagging on Testing Data

(E) Nothing Can Be Done

Question 18 (5 points)

We ran a marketing campaign to promote a product. We offered the product to 1,431 persons where 132 persons responded positively to the product. We first trained a binary logistic regression model. Afterward, we generated the following Receiver Operating Characteristics curve to measure the performance of this model. Which of the following values might be the Area Under Curve metric of the model?



Multiple Choice:

(A) 0.3

(B) 0.5

(C) 0.6

(D) 0.8

(E) 0.95

Question 19 (5 points)

Given the following confusion matrix

	Predicted Event	Predicted Non-Event
Observed Event	18	7
Observed Non-Event	3	22

Please calculate the F1 Score.

Question 20 (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. Because of privacy concerns, we cannot give you access to the original data. Instead, we provide you the following table. Our desired respondents are labelled Event.

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

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. In what deciles the respondents are in should we reach?

Multiple Choice:

(A) Decile 0

(B) Decile 0 and Decile 1

(C) Decile 0, Decile 1, and Decile 2

(D) Decile 0, Decile 1, Decile 2, and Decile 3

(E) Decile 0, Decile 1, Decile 2, Decile 3, and Decile 4