March 2021 Exam
Sample Questions Answer Key with LO reference

Part 1 - 113 Multiple Choice Questions

1. C LO 1.1.2, pp 24
2. B LO 2.1.2, pp 15-28
3. A LO 2.1.3, pp 36
4. B LO 2.1.3, pp 34 $(0.052+(4.57)^2+0.033$
5. B LO 2.2.2, pp 7
6. A LO 2.1.3, pp 33-34
7. C LO 2.1.3, pp 31
8. B LO 2.1.3, pp 31
9. A LO 2.2.3, pp 7
10. A LO 2.2.6, pp 23-32 $\nu \rightarrow \nu' = \nu - \eta \nabla C$
11. A LO 7.2.1, pp 2
12. B LO 3.1.2, pp 50-55 $(0.98 - 0.45*0.35 - 0.55*0.75 = 0.41)$
13. B LO 2.1.3, pp 33-34
14. B LO 4.1.1, pp 116-118
15. C LO 7.1.1, pp 239-242
16. B LO 3.1.2, pp 61-62 $RSS = (2.5 - 0.5 - 0.7X3)^2 + (3.5 - 0.5 - 0.7X4)^2 + (3.9 - 0.5 - 0.7X5)^2 = 0.06$
17. C LO 3.2.1, pp 65-68 $(1.4 - 0)/0.49 = 2.86$
18. B LO 3.2.3, pp 83-84
19. B LO 3.2.3, pp 86-89
20. C LO 3.2.3, pp 88-89
21. C LO 3.2.3, pp 92-102
22. B LO 4.2.1, pp 210-213
23. A LO 4.2.2, pp 219-229
24. A LO 4.2.1, pp 214-217
25. A LO 4.2.2, pp 219-220
26. A LO 3.3.2, pp 8-9 $\frac{\sigma^2}{1-\alpha^2} = .50/(1-.3^2) - .55$
27. A LO 3.3.3, pp 13-14
28. A LO 5.1.1, pp 149-153
29. C LO 5.1.2, pp 159-161 $1 - \frac{7X5+4X2+5X1+3X4}{\sqrt{7^2+4^2+5^2+3^2}\sqrt{5^2+2^2+1^2+4^2}} = 1 - \frac{7X5+4X2+5X1+3X4}{\sqrt{9.95}\sqrt{6.78}}$
30. A LO 5.1.2, pp 165-170
31. C LO 5.1.3, pp 188-190
32. C LO 6.1.1, pp 214-219
33. C LO 6.2.1, pp 67-68
34. B LO 7.1.3, pp 263-265
35. C LO 7.1.4, pp 265
36. B LO 7.2.5, pp 11
37. A LO 7.2.5, pp 11
38. A LO 9.4.4, pp 24-25
39. B LO 9.4.1, pp 6
40. A LO 9.4.1, pp 21
41. C LO 9.5.1, pp 16 & 28
42. B LO 9.5.1, pp 21-22

43. B LO 9.6.1, pp 34-35
44. A LO 9.6.1, pp 34
45. C LO 9.6.1, pp 32-33
46. C LO 9.6.1, pp 34-35
47. B LO 9.7.1, pp 7-8
48. A LO 9.7.1, pp 11
49. A LO 9.7.1, pp 17
50. A LO 9.8.1, pp 10-11
51. B LO 9.8.1, pp 11
52. A LO 9.58.1, pp 11-12
53. A LO 9.8.1, pp 12-13
54. A LO 9.10.1, pp 110
55. C LO 9.10.1, pp 111-112
56. C LO 9.10.1, pp 112-113
57. A LO 3.1.2, pp 56-62
58. B LO 7.1.1, pp 244-246
59. C LO 3.2.3, pp 95-96
60. B LO 4.2.2, pp 217-218
61. A LO 4.2.2, pp 219-220
62. B LO 3.3.2, pp 7
63. C LO 5.1.3, pp 189-191
64. A LO 5.1.4, pp 203-204
65. C LO 2.2.6, pp. 29-30
66. A LO 3.1.5, pp 101
67. A LO 3.2.1, pp 62
68. A LO 5.1.1, pp 143-144
69. D LO 5.1.3, pp 190
70. A LO 6.1.1, pp 211
71. C LO 6.1.1, pp 216
72. B LO 9.5.1, pp 15
73. C LO 9.6.1, pp 22
74. B LO 9.8.1, pp 26
75. B LO 9.8.1, pp 28
76. D LO 9.10.1, pp 36
77. C LO 9.11.1, pp 40
78. A LO 6.1.1, pp 213
79. C LO 2.2.3, pp 12
80. C LO 3.1.2, pp 53
81. B LO 3.1.2, pp 53-59
82. B LO 3.1.5, pp 102
83. D LO 5.1.2, pp 159
84. C LO 6.1.1, pp 213
85. B LO 6.3.1, pp 2
86. B LO 6.3.1, pp 15
87. B LO 9.10.1, pp 34
88. A LO 9.11.1, pp 43
89. B LO 2.2.2, pp 7
90. A LO 3.1.2, pp 51
91. A LO 3.1.2, pp 53
92. C LO 3.1.3, pp 73
93. C LO 4.2.3, pp 312
94. C LO 5.1.1, pp 150
95. D LO 5.1.4, pp 195-196
96. D LO 6.3.1, pp 2
97. B LO 6.3.1, pp 4
98. C LO 7.1.1, pp 237
99. B LO 9.10.1, pp 35
100. C LO 1.2.1, pp 7-8
101. A LO 1.2.3, pp 9
102. C LO 1.3.3, pp 10
103. B LO 1.4.2, pp 6
104. B LO 1.4.4, pp 12
105. A LO 1.4.6, pp 16
106. A LO 1.4.6, pp 17
107. B LO 1.4.7, pp 19
108. A LO 9.1.1.A, pp 6
109. C LO 9.1.1.B, pp 6
110. A LO 9.1.1.E, pp 7-8
111. B LO 9.1.1.H, pp 12-13
112. B LO 9.2.1.D, pp 9
113. A LO 4.2.3, pp 312

Part 2 - Ten Constructive Responses

Constructive Response Question 1

1. Marilyn Taylor is a quantitative analyst responsible for finding new investment ideas in the equity space. In the past, she has used unsupervised learning techniques to filter firms from a larger list, and this technique seems to have worked well for her. She mainly considers firm characteristics, such as P/E ratio, P/B ratio, and size, for her analysis. Recently, she has come across Volta Electric Company, which looks interesting to her. Currently, it has a P/E ratio of 16.5, a P/B ratio of 2.3, and a size value of 4.6 billion. To compare Volta with some other companies, she pulled out information on 3 companies that she had analyzed in the past. Out of the 3 companies, 2 were recommended for investment and 1 was not recommended. Following table provides information on the three companies.

Name	P/E Ratio	P/B Ratio	Size ($ billions)	Recommendation
Northern Healthcare	14	2.1	5.9	Invest
Wholesome Foods	17	1.5	6.7	Invest
Real Tech	21	1.8	7.8	Do not invest

   A. Marilyn uses Euclidean distance to measure the difference between a new company and the companies she has analyzed in the past. What are the Euclidean distances between Volta and the 3 companies listed in the table above?

      If we take a majority vote of the distances, what would be the recommendation by Marilyn for Volta? Explain how you decided on this.

(5 Points)

Answer:

Euclidean distance between Volta and	Distance
Northern Healthcare	$\sqrt{(16.5-14)^2+(2.3-2.1)^2+(4.6-5.9)^2}=2.82$
Wholesome Foods	$\sqrt{(16.5-17)^2+(2.3-1.5)^2+(4.6-6.7)^2}=2.30$
Real Tech	$\sqrt{(16.5-21)^2+(2.3-1.8)^2+(4.6-7.8)^2}=5.54$

The recommendation would be to invest in Volta. If we consider the distances between Volta and the three companies analyzed earlier, we see that two of the companies are quite close in terms of the Euclidean distance and the recommendations for both of these companies were to invest in them. Majority voting requires us to look at all the points that are close to point we are trying to classify and assigns the majority class from the closest points to the new point.

Source: LO 5.1.1, Reading 5.1 pp 142-144, pp 147-149

Grading guidelines: 3 points for calculating the distances. 1 point for the correct recommendation. 1 point for explanation on the recommendation.

B. Marilyn also calculates weighted voting or similarity moderated voting for any new company using the companies she has analyzed in the past. What are the similarity weights and contribution to the probability of investing or not investing for the 3 companies? What would be the estimated probability of investing and not investing for Volta?

(5 points)

Sample Answer:

Name	Euclidean Distance	Similarity Weight	Contribution
Northern Healthcare	2.82	$\frac{1}{2.82^2} = 0.125748$	$\frac{0.125748}{0.347366} = 0.362$
Wholesome Foods	2.30	$\frac{1}{2.30^2} = 0.189036$	$\frac{0.189036}{0.347366} = 0.544$
Real Tech	5.54	$\frac{1}{5.54^2} = 0.032582$	$\frac{0.032582}{0.347366} = 0.094$
Sum		0.347366	

Estimated probability of investing = 0.362 + 0.544 = 0.906
Estimated probability of not investing = 0.094

Source: LO 5.1.1, Reading 5.1 pp 149-151
Grading guidelines: 1.5 points for similarity weights. 1.5 points for contributions to probability. 2 points for the estimated probabilities.

Constructive Response Question 2

2. George works at Tentacle, a hedge fund focused on marine industries. Tentacle is particularly interested in using data collected by Jellyfish Technologies (JT), a private biotechnology firm, as inputs to its neural networks. Tentacle aims to predict equity price changes in various marine industries involved in or impacted by changing marine life conditions and ecosystems.

   JT uses automated diving cameras to collect data and machine learning algorithms to identify new species of marine life, primarily for medical applications. The data they collect is diverse and can provide information that is valuable to the government and a host of industries. Both managers and scientists at Jellyfish technologies are very protective of the data that they collect, although they are also dependent on data collected by government agencies and through industrial pursuits by other firms. This is known as the "selfish scientist paradox." Nevertheless, certain confidentiality agreements with pharmaceutical firms prevent the sharing of some types of data whereas government regulations of the United States National Oceanic and Atmospheric Administration (NOAA) mandate the sharing of some types of data. It is JT's policy to share data without compensation only if mandated by law.

   Answer each question based on your readings of the Business Ethics Briefings by the Institute for Business Ethics (IBEs).

   A. Does Jellyfish Technologies' use and lack of willingness to disseminate marine data constitute a 'data trust deficit' issue for Tentacle, Jellyfish Technologies or NOAA? If so, why or why not? In your answer, be sure to define the term "data trust deficit" as used by the IBEs.

(3 points)

Sample Answer:

"Data trust deficit" is a term to describe research that shows the public has a lower level of trust in companies to use data appropriately than their level of trust in general.

In the scenario above, the data is not of a personal or individual nature so that the public is not directly impacted by the use or dissemination of the marine data. Therefore, one could say that the data trust deficit is not an issue, in the context above, for any of the three parties - if data is defined as individual's personal data as implied (but not stated) by the reading.

In the context of openness and transparency, i.e., the appropriate use of data (e.g., for the purpose of innovation and public welfare) the "selfish scientist paradox" displayed by Jellyfish Technologies may constitute a data trust deficit from the perspective of NOAA.

Source: LO 8.1.1, Reading 8.1, p 2

Grading Guidelines:

a. 1 point for the correct definition of "data trust deficit"

b. 2 points for answering that none of the three parties are concerned about the data deficit in the context of the vignette (unless a valid case can be made, such as that of the NOAA.)

If a different case is made for part b., it may be hard to judge without a grading committee consensus, such as when using members to grade.

B. How consistent are JT's policies consistent with IBEs' values that form the cornerstone of an ethical framework of ARTIFICIAL intelligence in business?

Discuss the values of Transparency and Control in the context of JT's attitude towards data dissemination. For each value determine whether JT's values are consistent, or inconsistent with IBEs framework, or not relevant to the information in the vignette. What approach might George use to convince JT to provide the data that interests Tentacle?

(7 points)

Sample Answer:

Transparency in the IBEs framework refers to having source code freely available and open to scrutiny so that it builds trust with the public. An additional benefit is that code can be improved. This contrasts with traditional models whereby companies treated code as a corporate asset that should not be shared.

JT: JT is reluctant to share their data. If the transparency value is extended to include data, then JT's values are not consistent with IBEs framework.

Tentacle: Not relevant. We do not know if Tentacle wishes to share its algorithms or data.

Control in the IBEs framework refers to the fear that humans may lose control over AI machines. The recommendation is to make sure that it is understood how the algorithms work. Being able to explain how machines work will help to build trust with all stakeholders.

JT: Control as used by IBEs is not relevant to the vignette since it does not relate to controlling the terms of possession.

Tentacle: Control as used by IBEs is not relevant to the vignette.

George is likely willing to pay for the data if Mary is not willing to share it for free. George might remind Mary that Tentacle is not a competitor and needn't be protective of the data.

George could also offer a promise to not share the data with anyone outside of Tentacle. Yet, in that case, both Tentacle and JT would be acting inconsistently with IBEs framework.

Source: LO 8.3.1 Reading 8.3, p 1.

Grading Guidelines: 3 points for the discussion of transparency, 3 points for the discussion of control and 1 point for a recommendation to George.

Constructive Response Question 3

3. Lilian is a fund manager who has two quantitative modelers working for her. She wants to evaluate the performance of each model. She has gathered the following information.

	Number of Times the Signal Given	
	Model A	Model B
Buy signals	25	30
Buy signals that were correct	19	22
Do not buy signals	20	22
Do not buy signals that were correct	12	15

A. Lilian wants to form the confusion matrices for models A and B. What are the values for X1, X2, X3, and X4 for each of the models?

(2 points)

	Actual Buy	*Actual Do not Buy*
Predicted Buy	*X1*	*X2*
Predicted Do not Buy	*X3*	*X4*

Sample Answer:

Model A

	Actual Buy	*Actual Do not Buy*
Predicted Buy	*19*	*6*
Predicted Do not Buy	*8*	*12*

Model B

	Actual Buy	*Actual Do not Buy*
Predicted Buy	*22*	*8*
Predicted Do not Buy	*7*	*15*

Source: LO 5.1.3, Reading 5.1 pp 189-191
Grading guidelines: 1 point each for the correct confusion matrix.

B. Define true positive rate and false positive rates.

(2 points)

Sample Answer:
True positive rate refers to the frequency of being correct. It is the proportion of values that are correctly predicted to have a particular trait out of all the values that actually have the trait.

False positive rate refers to the frequency of incorrectly predicting values that are supposed to have a trait but ultimately does not have the trait.
Source: LO 5.1.3, Reading 5.1 pp 190, 203.
Grading guidelines: 1 point for each definition.

C. What are the true positive rates for the two (2) models?

(2 points)

- *True positive rate, Model A = 19/(19+8) = 70.37%*
- *True positive rate, Model B = 22/(22+7) = 75.86%*

Source: LO 5.1.3, Reading 5.1 pp 189-191
Grading guidelines: 1 point each for the true positive rate.

D. Suppose the benefit of a true positive in $100 and the benefit of a true negative is 0 while the cost of a false positive is $50 and the cost of a false negative is $100. What is the expected value of each model?

(4 points)

Sample Answer:

Expected value of model A = $\frac{19}{45}X100 - \frac{6}{45}X50 - \frac{8}{45}X100 = 17.78$

Expected value of model B = $\frac{22}{52}X100 - \frac{8}{52}X50 - \frac{7}{52}X100 = 21.15$

Source: LO 5.1.4, Reading 5.1 pp 196-202
Grading guidelines: 2 points for each expected value.

E. Define a profit curve and list two (2) characteristics of a profit curve.

(3 points)

Sample Answer:

- A profit curve plots the cumulative amount of expected profit from all test instances for various classifiers. This essentially involves ranking the list of instances by score from the highest to the lowest and sweeping down through it, recording the expected profit after each instance.
- Two characteristics:
  - All curves begin and end at the same point. At the left side, there are no expenses and zero profit for every curve; at the right side, every test instance is targeted, so every classifier performs the same.
  - The curves can show that sometimes profit can be negative. This happens when the profit margin is thin, the number of responders is small, and one works too far down the list.
  - Budget constraint determines the profit curve that is used and the choice of the classifier to do the ranking.

Source: LO 6.1.1, Reading 6.1 pp 212-214
Grading guidelines: 1 point for the definition. 2 points for the characteristics.

Constructive Response Question 4

4. The following information is collected from historical census data. Use the natural logarithm to perform all calculations involving logarithm.

	Features				Target
ID	Age	Education	Marital Status	Occupation	Salary
1	39	bachelors	never married	transport	35K-65K
2	50	bachelors	married	professional	35K-65K
3	18	high school	never married	agriculture	< 35K
4	28	bachelors	married	professional	35K-65K
5	37	high school	married	agriculture	35K-65K
6	24	high school	never married	transport	< 35K
7	52	high school	divorced	transport	35K-65K
8	40	graduate	married	professional	>65

There are four descriptive features and one target feature in this dataset:

- AGE: a continuous feature listing the age of the individual
- EDUCATION: a categorical feature listing the highest education award achieved by the individual (high school, bachelors, graduate school)
- MARITAL STATUS: a categorical feature listing current marital status (never married, married, divorced)
- OCCUPATION: a categorical feature listing current occupation (transport = works in the transportation industry; professional = works as a lawyer, accountant, etc.; agriculture = works in the agricultural industry)
- SALARY: the target feature with 3 levels (less than 35,000; between 35,000 and 65,000; more than 65,000)

A. Using the values of the target feature, what is the entropy of the above dataset? **(3 points)**
B. Based on the information provided below, briefly explain what the information gain for Split by Age more than 39.5 should be. **(3 points)**
C. When building a decision tree, a convenient way to handle a continuous feature is to define a threshold at which splits will be made. Briefly explain the optimal threshold to split on the continuous AGE feature based on the information provided below. **(2 points)**

Split by Age Feature	IDs in Each Partition	Partition Entropy	Information Gain
> 26	3,6 1,2,4,5,7,8	0.000 0.451	0.619
> 39.5	1,3,4,5,6 2,7,8	0.673 0.637	??
>45	1,3,4,5,6,8 2,7	1.011 0.000	0.268

Source: LO: 3.1.2.C-D, 4.2.3.E,
Sample Answer

A. Entropy = -((5/8)×ln(5/8) + (2/8)×ln(2/8) + (1/8)×ln(1/8)) = 0.90
B. IG = 0.90- (5/8)*0.673 – (3/8)*0.637 = 0.241
C. Split by > 26 because it has the highest information gain.

Constructive Response Question 5

5. A bag-of-words approach is used to detect spam emails. The following table displays the information about 5 emails from the training set.

	Number of Times the Words Appear in Each Email							
ID	MONEY	FREE	FOR	GAMBLING	FUN	MACHINE	LEARNING	SPAM
1	3	0	0	0	0	0	0	TRUE
2	1	2	1	1	1	0	0	TRUE
3	0	0	1	1	1	0	0	TRUE
4	0	0	1	0	3	1	1	FALSE
5	0	1	2	0	0	1	1	FALSE

A new email containing the words "machine learning for free" is received, and a nearest-neighbor model (k-NN) using Euclidean distance is used to determine if the email is a spam. The following table displays the components of the Euclidean distance as well as the Euclidean distance between the new email and those in the training set:

	Components for Calculating the Euclidean Distance Between the New Email and Training Set							
ID	MONEY	FREE	FOR	GAMBLING	FUN	MACHINE	LEARNING	DISTANCE
1	9	1	1	0	0	1	1	3.61
2	1	1	0	1	1	1	1	2.45
3	0	1	0	1	1	1	1	2.24
4	0	1	0	0	9	0	0	3.16
5	0	0	??	0	0	0	0	??

A. The Euclidean distance between email #5 and the new email and the component for calculating the Euclidean distance for the feature "FOR" are missing. Briefly explain what values should appear in these cells. **(2 points)**

B. A k-NN approach with k = 1 is used to classify the new email. Briefly explain what the new email's classification should be. **(2 points)**

C. A k-NN approach with k=3 and majority rule are used to classify the new email. Briefly explain what the new email's classification should be. **(2 points)**

D. A weighted voting k-NN with k=2 is used to classify the new email. Briefly explain what the new email's classification should be. **(3 points)**

Source: LO: 5.1.1.A-F
Sample Answer

A. A value of 1 should appear in the cell for "FOR" since $(2-1)^2 = 1$. The distance between the new email and email #5 $\sqrt{0+0+(2-1)^2+0+0+0+0} = 1$.

B. The nearest neighbor is #5 because it has the smallest Euclidean distance. Using just 1 neighbor means the classification will be that of #5 = FALSE. This implies that the new email will not be classified as a SPAM.

C. When k= 3, the 3 nearest neighbors are with IDs 2, 3, and 5. Two of these IDs have TRUE as the target variable and one of the ID has FALSE as the target variable. So, the classification will be TRUE since there are more TRUES among the nearest neighbors. This means the email will be classified as a SPAM.

D. When K = 2, one of the nearest neighbors will have TRUE as the value of the target variable and the other neighbor will have FALSE as the value of the target variable. Of these two neighbors, the closer neighbor has FALSE as the value of the target variable. Since the weighting function weights the points inversely to the distance, nearer point will be more heavily weighted. This will mean that the overall value of the target variable will be FALSE. So, the email will not be classified as a SPAM.

Constructive Response Question 6

6. The following table displays the return on an investment depending on the values of 3 factors observed in the previous year.

Factor 1	Factor 2	Factor 3	Return
1	1	1	-15
1	1	1	-5
2	1	1	7
2	1	2	1
1	2	1	0
1	2	2	13
2	2	2	12

The mean and the residual sum of squares of future returns when a predictor (i.e., factor) takes on a specific value are provided below:

	Mean Return	Residual Sum of Squares
Factor 1 = 1	-1.8	407
Factor 1 = 2	6.7	61
Factor 2 = 1	-3.0	264
Factor 2 = 2	8.3	105
Factor 3 = 1	-3.3	257
Factor 3 = 2	8.7	89

A. Using a recursive binary splitting approach, explain which factor should be used in the top node to split the response, i.e., return. **(2 points)**

Source: LO 4.2.3
Sample Answer

Total residual sum of square for Factor 3 is 257 + 89 = 346 and is the lowest of the three factors. Factor 3 should be used for the split.

Constructive Response Question 7

7. A sample of 120 monthly rates of return on Bitcoin is used as the training set. The most recent two observations from the training set (Observations 119 and 120) and other estimates are presented in the following table.

Observation	Realized Return % at time t-1	Unexpected Returns % at time t-1	Expected Standard Deviation % at time t
119	9.40	5.90	7.96
120	-5.10	-8.60	10.26
121	8.30	4.80	??

The standard deviations appearing in the last column are obtained using the following GARCH(1,1) model:

$$p_t = 0.035 + a \times p_{t-1} + \epsilon_t$$
$$\epsilon_t \sim N(0, \sigma_t^2)$$
$$\sigma_t^2 = 0.002 + 0.01\epsilon_{t-1}^2 + 0.8\sigma_{t-1}^2$$

Here, p_t is the log of Bitcoin's price.

A. Given the information presented in the above table and the estimated GARCH(1,1) model, what is the forecast of standard deviation given the new observations (i.e., month 121)? **(3 points)**
B. Briefly explain the conditions under which the log of Bitcoin price will follow a random walk. **(3 points)**
C. Suppose the following dataset is used to create a nearest neighbor model that predicts whether the Bitcoin price will rise or decline next month. Manhattan distance is used to find the nearest neighbor:

Observations	S&P 500 Return %	Treasury Bill Rate %	Next Month's Perfomance of Bitcoin
1	4.00	2.00	Positive
2	3.00	1.50	Positive
3	-2.00	1.00	Negative
4	1.00	1.00	??

Using the most recent observation, #4, and a k-NN model with k =1, briefly explain your forecast of the Bitcoin's next month. **(4 points)**

Source: LO 3.3.3.A-C
Sample Answer

A. $\sigma_{121} = (0.002 + 0.01 \times 4.8\%^2 + 0.83 \times 10.26\%^2)^{0.5} = 10.37\%$
B. The process will be a random walk if $a = 1$ and the errors are independently and identically distributed.
C. The Manhattan distances of observation #4 from the other 3 observations are:

Observations	Distance	
1	4.00	
2	2.50	Positive
3	3.00	

The nearest neighbor is observation #2. Therefore, we expect the change in the price of Bitcoin to be positive.

Constructive Response Question 8

8. The following table represents our entire sample of some emails. Naïve Bayesian approach will be used to design a spam filter that is trained using this sample. The table indicates whether an email contained the words "Free" or "Urgent." For example, email #1 was Spam and contained the word "Free," but did not contain the word "Urgent."

Observations	Free	Urgent	Spam
1	1	0	Yes
2	0	0	Yes
3	1	0	No
4	0	1	No
5	1	1	Yes
6	0	1	No

A. Based on these observations, what is the unconditional probability of receiving Spam? **(2 points)**
B. Based on these observations, what is the probability of receiving a Spam email that contains the word "Free"? **(2 points)**
C. A new email arrives, and it contains the word "Free." What is the conditional probability that the email is Spam? **(4 points)**

Source: LO 7.1.1.A-F
Sample Answer

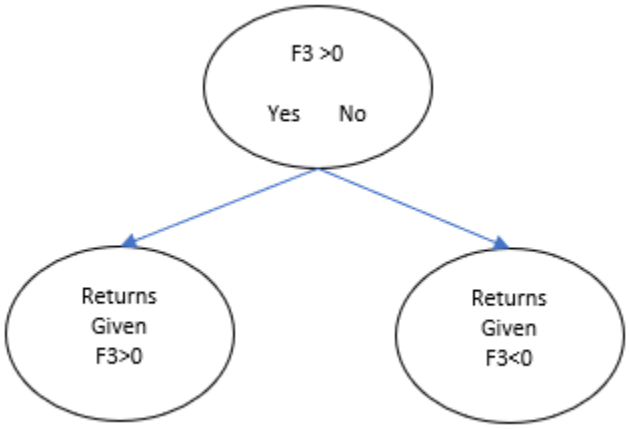
A. The unconditional probability of receiving Spam email is 3/6 = 50%
B. The joint probability of receiving a Spam email with the word "Free" is 2/6 = 33.3%
C. The conditional probability than an email is Spam given that it contains the word "Free" is P(Spam|Free)=Prob(Free|Spam)*P(Spam)/P(Free) = P(Spam & Free)/P(Free).
Since P(Free) = 3/6 = 50%, P(Spam|Free) = 33.3%/50% = 66.7%

Constructive Response Question 9

9. The following table displays a set of 10 observations on the annual values of 3 factors (i.e., features) and the following year's return on an investment (i.e., target variable). For example, for observation #1, F1 = 1.5, F2 = -4.1 and F3 = -1.2 and the return on the investment in the following year was Negative. The last row displays each factor's entropy and the target's entropy using a split of positive versus negative values. A supervised segmentation approach will be applied to this data set to predict our investment return, given new factors' observations.

Observations	Factor 1	Factor 2	Factor 3	Next Year's return
1	1.5	-4.1	-1.2	Negative
2	-2.8	-4.2	4.6	Positive
3	-0.1	-5.8	2.5	Positive
4	4.6	1.6	6.8	Positive
5	-0.8	-4.1	0.8	Negative
6	-4.0	-0.5	0.1	Negative
7	3.9	-3.2	3.8	Positive
8	-3.5	2.1	6.6	Positive
9	-1.7	-5.8	-1.8	Negative
10	1.5	0.4	2.4	Positive
Entropy	0.67	0.61	0.50	0.67

A. Suppose the threshold F3>0 is used to split the target variable (i.e., next year's return) at the root node. What will be the entropy of the two children displayed below? **(3 points)**



B. What will be the information gain from the above split? **(3 points)**

Source: LO 3.1.2.A-G

Sample Answer

A. Conditional on F3 > 0: there will be 8 observations in the node, of which 2 will be negative next year's returns. Thus, its entropy will be -(2/8)*ln(2/8)-(6/8)*ln(6/8) = 0.56
Conditional on F3 < 0: there will be 2 observations in the node, of which 2 will be negative and no positive returns. The entropy will be zero.

B. The information gain = 0.67 – (8/10)*0.56 – (2/10)*0 = 0.22

Constructive Response Question 10

10. Suppose a sample of credit card customers accounts is used to determine how relative account balances and marital status affect the likelihood of delay in their monthly payments. The following table presents 4 observations from the training set.

Customer ID	Balance/Annual Salary	Marital Status	Late Payment
1105	0.14%	0	0
1106	7.15%	1	1
1107	0.51%	1	0
1108	1.31%	1	0

For example, Customer 1105's ratio of credit card balance to annual income was 0.14%, the customer was not married, and was not late with the payment. The parameter values of the two equations appearing below were estimated using a large training set.

$$Late\ Payment = 0.13 + 16 \times \left({Balance}/{Salary}\right) - 0.27 \times Marital \qquad Eq(1)$$

$$Late\ Payment = \frac{1}{1 + exp(0.5 - 19 \times (Balance\ /\ Salary) + 0.31 \times Marital)} \qquad Eq(2)$$

Both equations had high explanatory power, and the coefficients were statistically significant.

A. Briefly explain which model is better suited for the current application. **(3 points)**
B. A customer with the following information appears in the testing database.

Customer ID	Balance/Annual Salary	Marital Status
1510	6.50%	0

Briefly explain whether the customer is likely to be late on his payment. **(3 points)**

C. What is the expression for the class's log-odds (i.e., the customer is late with the payment) based on Equation (2)? **(2 points)**

Source: LO 3.1.5.A-F

Sample Answer

A. Logistic regression (Eq 2) is better suited because the output will be between 0 and 1.
B. For Equation 1: Late Payment Prob = 0.13 + 16*0.065 – 0.27*0 =1.17
For Equation 2: Late Payment Prob = 1/(1+exp(0.5-19*0.065+0.31*0)) = 0.68
The customer is likely to be late. Using only equation 1 is incorrect.
C. The log-odds expression is ln(p(x)/(1-p(x)) = -0.5 + 19*Balance/Salary – 0.31*Marital