Midterm Version B

CS540

July 15, 2019

1 Instruction

1. Each incorrect answer receives -0.25, each correct answer receives 1, blank answers receives 0.
2. Check to make sure your name and (numerical) student ID (if you have it) is on the (Scantron) answer sheet. Also write your Wisc email ID on the answer sheet.
3. Check to make sure you completed question 41 and 42.
4. If you think none (or more than one) of the answers are correct, choose the best (closest) one.
5. Please submit this midterm, the answer sheet, the formula sheet, and all your additional notes when you finish.
6. Good luck!

2 Questions

41. Calculator?

    - A: Yes.
    - B: No.

42. Number of pages of additional notes? Please submit them at the end of the exam.

    - A: 0
    - B: 1
    - C: 2
    - D: 3
    - E: 4 or more.

3 Questions

1. Consider a linear threshold perceptron $\hat{y}_i = a_i = \mathbb{1}_{\{wx_i+b\geqslant 0\}}$ with initial weights $w = 1$ and bias $b = -1$. Given a new input $x_i = 1$ and $y_i = 1$. Let the learning rate be 1, what is the updated weight w after one iteration of the perceptron algorithm?

    - A: -1
    - B: 0
    - C: 1
    - D: 2
    - E: 3

2. Continue from the previous question, what is the updated bias term b?

    - A: -3
    - B: -2
    - C: -1
    - D: 0
    - E: 1

3. Let $C(w) = \det\left(ww^T\right), w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$. Here, $\det(A) = ad - bc$ for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the determinant of a matrix. What is the Hessian matrix of C at $w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$?

    - A: $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
    - B: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
    - C: $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$
    - D: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
    - E: $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

4. Consider a linear model without bias term $a_i = wx_i$ with log cost function $-y_i \log(a_i) - (1 - y_i)\log(1 - a_i)$. The initial weight is w_0. What is the updated weight after one stochastic gradient descent step for w if the chosen training data is $x_1 = 1, y_1 = 0$? The learning rate is α.

- A: $w_0 - \dfrac{\alpha}{1 + w_0}$
- B: $w_0 - \dfrac{\alpha}{1 - w_0}$
- C: w_0
- D: $w_0 + \dfrac{\alpha}{1 + w_0}$
- E: $w_0 + \dfrac{\alpha}{1 - w_0}$

5. Continue from the previous question, what if the chosen training data is $x_1 = -1, y_1 = 0$? Everything else is the same.

- A: $w_0 - \dfrac{\alpha}{1 + w_0}$
- B: $w_0 - \dfrac{\alpha}{1 - w_0}$
- C: w_0
- D: $w_0 + \dfrac{\alpha}{1 + w_0}$
- E: $w_0 + \dfrac{\alpha}{1 - w_0}$

6. Given the following weights for $a2$ layer neural network, which of following logical operators does it represent?

$$w_{11}^{(1)} = +2, w_{21}^{(1)} = -2, b_1^{(1)} = -1$$
$$w_{12}^{(1)} = -2, w_{22}^{(1)} = +2, b_2^{(1)} = +1$$
$$w_{11}^{(2)} = -2, w_{21}^{(2)} = -1, b_1^{(2)} = +1$$

The activation functions are LTU, $\mathbb{1}_{\{w^T x + b \geqslant 0\}}$ for all units. The notation $w_{ij}^{(l)}$ represents the weight in layer l from unit i in the previous layer to unit j in the next layer.

x_1	x_2	$\wedge$ AND	$\vee$ OR	$\neg$ NOT	XOR	NOR	XNOR	$\Rightarrow$	$\Leftarrow$
1	1	1	1	0	0	0	1	1	1
1	0	0	1	0	1	0	0	0	1
0	1	0	1	1	1	0	0	1	0
0	0	0	0	1	0	1	1	1	1

- A: XOR
- B: NOR
- C: XNOR
- D: $\Rightarrow$
- E: $\Leftarrow$

7. Which ones of the following expressions is equivalent to XOR?

$$\begin{aligned}
1 &: (x_{i1} \wedge \neg x_{i2}) \vee (\neg x_{i1} \wedge x_{i2}) \\
2 &: (x_{i1} \vee x_{i2}) \wedge (\neg x_{i1} \vee \neg x_{i2}) \\
3 &: (\neg x_{i1} \wedge x_{i2}) \vee (x_{i1} \wedge x_{i2}) \\
4 &: (\neg x_{i1} \vee x_{i2}) \wedge (x_{i1} \vee \neg x_{i2}) \\
5 &: (\neg x_{i1} \wedge \neg x_{i2}) \vee (x_{i1} \wedge x_{i2})
\end{aligned}$$

- A: 1 and 2
- B: 1 and 3
- C: 2 and 3
- D: 3 and 4
- E: 4 and 5

8. Continue from the previous question, which ones of those expressions is equivalent to XNOR?

- A: 1 and 2
- B: 1 and 3
- C: 2 and 3
- D: 3 and 4
- E: 4 and 5

9. Perceptron algorithm will terminate on which one of the following training set? The data set is construsted using $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, x_4 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ on these operators. Note: the operators are the same as the ones in the previous question.

$$\begin{aligned}
y_i^{(A)} &= (x_{i1} \wedge \neg x_{i2}) \vee (\neg x_{i1} \wedge x_{i2}) \\
y_i^{(B)} &= (x_{i1} \vee x_{i2}) \wedge (\neg x_{i1} \vee \neg x_{i2}) \\
y_i^{(C)} &= (\neg x_{i1} \wedge x_{i2}) \vee (x_{i1} \wedge x_{i2}) \\
y_i^{(D)} &= (\neg x_{i1} \vee x_{i2}) \wedge (x_{i1} \vee \neg x_{i2}) \\
y_i^{(E)} &= (\neg x_{i1} \wedge \neg x_{i2}) \vee (x_{i1} \wedge x_{i2})
\end{aligned}$$

- A: Training set $\left\{\left(x_i, y_i^{(A)}\right)\right\}_{i=1}^{4}$
- B: Training set $\left\{\left(x_i, y_i^{(B)}\right)\right\}_{i=1}^{4}$
- C: Training set $\left\{\left(x_i, y_i^{(C)}\right)\right\}_{i=1}^{4}$
- D: Training set $\left\{\left(x_i, y_i^{(D)}\right)\right\}_{i=1}^{4}$
- E: Training set $\left\{\left(x_i, y_i^{(E)}\right)\right\}_{i=1}^{4}$

10. For a convolutional neural network with 10×10 input images, one convolutional layer with one 3×3 filter (one activation map, with zero padding, i.e. convolution preserves the size of the original image by adding zeroes around the border of the image), then one 2×2 max pooling layer (non-overlapping or stride 2), and then one fully connected layers with three output units (for three-category classification problems). How many weights (not including bias terms) are updated during training?

- A: $3 \cdot 3 \cdot 2 + 10 \cdot 10 \cdot 3$
- B: $3 \cdot 3 + 2 \cdot 2 + 10 \cdot 10 \cdot 3$
- C: $3 \cdot 3 + 2 \cdot 2 + 5 \cdot 5 \cdot 3$
- D: $3 \cdot 3 + 10 \cdot 10 \cdot 3$
- E: $3 \cdot 3 + 5 \cdot 5 \cdot 3$

11. Continue from the previous question. Suppose the 3×3 filter is trained to be the following filter F. Given an image full of 1's (every pixel is white), how many 1's are there in the activation map (activation matrix) in the convolutional layer, $a^{(1)}$?

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- A: $10 \cdot 10$
- B: $9 \cdot 9$
- C: $8 \cdot 8$
- D: $10 \cdot 9$
- E: $9 \cdot 8$

12. Continue from the previous question. How many 1's are there in the activation map in the pooling layer, $a^{(2)}$?

- A: $10 \cdot 10$
- B: $10 \cdot 9$
- C: $5 \cdot 5$
- D: $4 \cdot 4$
- E: $5 \cdot 4$

13. Given a fully connected neural network with three hidden layers, $a^{(1)}, a^{(2)}, a^{(3)}$, of 10 units in each hidden layer. The notation $w_{j'j}^{(l)}$ is the weight from unit j' in layer $l-1$ to unit j in layer l, and the notation $a_j^{(l)}$ is the unit j activation on layer l for the current instance. Which of the following derivatives is always 0 for any training data?

- A: $\dfrac{\partial a_1^{(2)}}{\partial w_{12}^{(2)}}$
- B: $\dfrac{\partial a_2^{(2)}}{\partial w_{12}^{(2)}}$
- C: $\dfrac{\partial a_1^{(3)}}{\partial w_{12}^{(2)}}$
- D: $\dfrac{\partial a_2^{(3)}}{\partial w_{12}^{(2)}}$
- E: $\dfrac{\partial a_3^{(3)}}{\partial w_{12}^{(2)}}$

14. Given the following training data. What is the 2 fold cross validation accuracy (percentage of correct classificationn) if 1 nearest neighbor classifier with Manhattan distance is used? The first fold is the first five data points.

x_{i1}	1	1	2	2	3	3	4	4	5	5
y	1	2	3	3	2	2	3	3	2	1

- A: 0 percent
- B: 20 percent
- C: 30 percent
- D: 40 percent
- E: 60 percent

15. Continue from the previous question. What is the 10 fold cross validation accuracy (percentage of correct classificationn)?

- A: 20 percent
- B: 40 percent
- C: 60 percent
- D: 80 percent
- E: 100 percent

16. Find the weights w_1, w_2, b for the support vector machine classifier $\mathbb{1}_{\{w_1 x_{i1} + w_2 x_{i2} + b \geqslant 0\}}$ given the following training data.

x_1	x_2	y
0	2	0
4	0	1

- A: $w_1 = 1, w_2 = -1, b = 0$
- B: $w_1 = 2, w_2 = -1, b = 0$
- C: $w_1 = 2, w_2 = -1, b = -3$
- D: $w_1 = 2, w_2 = -1, b = -5$
- E: $w_1 = 1, w_2 = -1, b = -2$

17. Continue from the previous question. What if the labels are flipped?

x_1	x_2	y
0	2	1
4	0	0

- A: $w_1 = -1, w_2 = 1, b = 0$
- B: $w_1 = -2, w_2 = 1, b = 0$
- C: $w_1 = -2, w_2 = 1, b = 3$
- D: $w_1 = -2, w_2 = 1, b = 5$
- E: $w_1 = -1, w_2 = 1, b = 2$

18. Continue from the previous question. What is the maximum margin?

- A: $\sqrt{2}$
- B: $\sqrt{5}$
- C: $\sqrt{8}$
- D: $\sqrt{10}$
- E: $\sqrt{20}$

19. Consider a filter discribed by $F_{t,t'} = t^2 + (t')^2, t = -1, 0, 1, t' = -1, 0, 1$. What is the convolution between the following matrix and this filter? Use zero padding, i.e. set nonexistent values to 0 around the edges of the first matrix.

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} * F$$

- A: $\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$, B: $\begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$, C: $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$, D: $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, E: $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

20. Continue from the previous question, what is the convolution between the following two matrices?

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} * F$$

- A: $\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$, B: $\begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$, C: $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$, D: $\begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, E: $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$

21. In a convolutional nueral network, suppose the resulting activation matrix of the convolution layer is the following $A^{(1)}$. What is the activation matrix after a non-overlapping 2×2 average pooling layer? (There are no scale and bias terms for the pooling layer.)

$$A^{(1)} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

- A: $\begin{bmatrix} 6 & 8 \\ 14 & 16 \end{bmatrix}$
- B: $\begin{bmatrix} 1 & 3 \\ 9 & 11 \end{bmatrix}$
- C: $\begin{bmatrix} 2.5 & 6.5 \\ 10.5 & 14.5 \end{bmatrix}$
- D: $\begin{bmatrix} 3.5 & 5.5 \\ 11.5 & 13.5 \end{bmatrix}$
- E: $\begin{bmatrix} 7 & 8 \\ 9 & 10 \end{bmatrix}$

22. Given the gradient magnitude and gradient orientation (in degrees) for a cell in an image, compute the histogram of oriented gradient (HOG) feature vector assuming 3×3 cells and 2 bins. The bins are split according to the following partitions. (Different from original HOG paper: do not split a single gradient magnitude into two bins.)

$$\{[0, 90), [90, 180)\}$$

$$G = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \Theta = \begin{bmatrix} 10 & 30 & 50 \\ 70 & 90 & 110 \\ 130 & 150 & 170 \end{bmatrix}$$

- A: $\begin{bmatrix} 4 & 5 \end{bmatrix}$
- B: $\begin{bmatrix} 5 & 4 \end{bmatrix}$
- C: $\begin{bmatrix} 1+2+3+4+5 & 6+7+8+9 \end{bmatrix}$
- D: $\begin{bmatrix} 1+2+3+4 & 5+6+7+8+9 \end{bmatrix}$
- E: $\begin{bmatrix} 1+2+3+4+2.5 & 2.5+6+7+8+9 \end{bmatrix}$

23. Continue from the previous question. What is the histogram if the there are 5 bins instead?

$$\{[0, 36), [36, 72), [72, 108), [108, 144), [144, 180)\}$$

- A: $\begin{bmatrix} 2 & 2 & 1 & 2 & 2 \end{bmatrix}$
- B: $\begin{bmatrix} 3 & 7 & 11 & 15 & 19 \end{bmatrix}$
- C: $\begin{bmatrix} 3 & 7 & 5 & 13 & 17 \end{bmatrix}$
- D: $\begin{bmatrix} 1 & 5 & 9 & 13 & 17 \end{bmatrix}$
- E: $\begin{bmatrix} 3 & 9 & 15 & 21 & 17 \end{bmatrix}$

24. What is the entropy of the random variable with the following distribution? Remember to use log base 2.

y	1	2	4	8
$\mathbb{P}\{Y = y\}$	$\dfrac{1}{2}$	$\dfrac{1}{4}$	$\dfrac{1}{8}$	$\dfrac{1}{8}$

- A: $\dfrac{1}{2} + \dfrac{1}{2} + \dfrac{1}{2} + \dfrac{1}{2}$
- B: $\dfrac{1}{2} + \dfrac{1}{2} + \dfrac{3}{4} + \dfrac{3}{4}$
- C: $\dfrac{1}{2} + \dfrac{1}{2} + \dfrac{3}{8} + \dfrac{3}{8}$
- D: $1 + 1 + 1 + 1$
- E: $1 + 1 + \dfrac{3}{2} + \dfrac{3}{2}$

25. What is the entropy of the random variable with the following distribution? $1024 = 2^{10}$.

$$\mathbb{P}\{Y = y\} = \frac{1}{1024}, \text{ for } y = 1, 2, 3, ..., 1024$$
$$\mathbb{P}\{Y = y\} = 0, \text{ otherwise}$$

- A: 0
- B: 1
- C: 2
- D: 10
- E: 1024

26. What is the accuracy (on the training set) of the decision tree that first splits on x_1 then splits on x_2 trained on the following training set?

$$(x_{i1}, x_{i2}, y_i)_{i=1...5} = \{(0,0,0), (0,0,0), (1,0,1), (1,1,1), (1,1,0)\}$$

- A: 20 percent
- B: 40 percent
- C: 60 percent
- D: 80 percent
- E: 100 percent

27. Continue from the previous question, what is accuracy if the decision tree is first split on x_2 then split on $x_{1?}$

- A: 20 percent
- B: 40 percent
- C: 60 percent
- D: 80 percent
- E: 100 percent

28. Given the following training set S, suppose n instances are removed, and 1 nearest neighbor with Manhattan distance is trained on the remaining set S', and tested on the original training set S. If the accuracy is 100 percent on S, what is the maximum possible value for n?

$$S = (x_i, y_i)_{i=1...5} = \{(-3, 0), (-1, 0), (0, 1), (1, 2), (3, 2)\}$$

- A: 0
- B: 1
- C: 2
- D: 3
- E: 4

29. Continue from the previous question. Same assumptions as the previous question: what is the maximum possible value for n if the training set S is changed to the following?

$$S = (x_i, y_i)_{i=1...5} = \{(-3, 2), (-1, 1), (0, 1), (1, 1), (3, 2)\}$$

- A: 0
- B: 1
- C: 2
- D: 3
- E: 4

30. Suppose there are 2 training instances $x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and the new feature vector for a support vector machine is given by $\phi(x_i) = \phi\left(\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}\right) = \begin{bmatrix} x_{i1} \\ x_{i1} \cdot x_{i2} \\ x_{i2} \end{bmatrix}$. What is the kernel (Gram) matrix?

- A: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- B: $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
- C: $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
- D: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
- E: $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

31. Three documents A, B and C. Document A is $HHTTTTTTTT$ and document B is $HHHHHTTTTT$ and document C is $HHHHHHHHTT$. One document is chosen at random (each document with equal probability) and one word is chosen at random (each word with equal probability). What is the probability that the word is H?

- A: $\dfrac{1}{3}\left(\dfrac{2}{10}+\dfrac{5}{10}+\dfrac{8}{10}\right)$
- B: $\dfrac{1}{3}\left(\dfrac{3}{10}+\dfrac{6}{10}+\dfrac{9}{10}\right)$
- C: $\dfrac{1}{3}\left(\dfrac{2}{11}+\dfrac{5}{11}+\dfrac{8}{11}\right)$
- D: $\dfrac{1}{3}\left(\dfrac{3}{11}+\dfrac{6}{11}+\dfrac{9}{11}\right)$
- E: $\dfrac{1}{3}\left(\dfrac{3}{12}+\dfrac{6}{12}+\dfrac{9}{12}\right)$

32. Continue from the previous question, suppose the answer to the question is p. Given the chosen word is H, what is the probability that the document is A?

- A: $\dfrac{1}{3}\dfrac{\frac{2}{10}}{p}$
- B: $\dfrac{1}{3}\dfrac{\frac{3}{10}}{p}$
- C: $\dfrac{1}{3}\dfrac{\frac{2}{11}}{p}$
- D: $\dfrac{1}{3}\dfrac{\frac{3}{11}}{p}$
- E: $\dfrac{1}{3}\dfrac{\frac{3}{12}}{p}$

33. Given the counts, what is the maximum likelihood estimate of $\mathbb{P}\{A|\neg C\}$ with Laplace smoothing? The event C means $C = T$ and the event $\neg C$ means $C = F$ in the following table.

A	B	C	count
F	F	F	1
F	F	T	0
F	T	F	0
F	T	T	4
T	F	F	1
T	F	T	1
T	T	F	1
T	T	T	2

- A: $\dfrac{1+1}{3+1}$
- B: $\dfrac{1+1}{3+2}$
- C: $\dfrac{1+1}{3+4}$
- D: $\dfrac{2+1}{3+2}$
- E: $\dfrac{2+1}{3+4}$

34. Continue from the previous question. What is the MLE of $\mathbb{P}\{A|\neg B, \neg C\}$ with Laplace smoothing?

- A: $\dfrac{1+1}{2+1}$
- B: $\dfrac{1+1}{2+2}$
- C: $\dfrac{1+1}{2+4}$
- D: $\dfrac{2+1}{2+2}$
- E: $\dfrac{2+1}{2+4}$

35. Suppose A, B and C form a causal chain, which means the Bayesian network is $A \to B \to C$. All variables are binary. What is $\mathbb{P}\{A = 1, C = 1\}$?

$$\mathbb{P}\{A = 1\} = 0.4$$
$$\mathbb{P}\{B = 1|A = 1\} = 0.8, \mathbb{P}\{B = 1|A = 0\} = 0.1$$
$$\mathbb{P}\{C = 1|B = 1\} = 0.3, \mathbb{P}\{C = 1|B = 0\} = 0.7$$

- A: $0.4 \cdot 0.3$
- B: $0.4 \cdot 0.8 \cdot 0.3$
- C: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.2 \cdot 0.7$
- D: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.1 \cdot 0.7$
- E: $0.4 \cdot 0.8 \cdot 0.3 + 0.4 \cdot 0.2 \cdot 0.3$

36. Continue from the previous question, suppose the answer to the previous question is p, what is $\mathbb{P}\{C = 1|A = 1\}$?

- A: $\dfrac{p}{0.4}$
- B: $\dfrac{p}{0.4 \cdot 0.8}$
- C: $\dfrac{p}{0.4 \cdot 0.2}$
- D: $\dfrac{p}{0.4 \cdot 0.8 \cdot 0.3}$
- E: $\dfrac{p}{0.4 \cdot 0.2 \cdot 0.7}$

37. Continue from the previous question, what is $\mathbb{P}\{A = 1|C = 1\}$?

- A: $\dfrac{p}{0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.6 + 0.7 \cdot 0.8 \cdot 0.4 + 0.7 \cdot 0.1 \cdot 0.6}$
- B: $\dfrac{p}{0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.6 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.9 \cdot 0.6}$
- C: $\dfrac{p}{0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.4 + 0.7 \cdot 0.8 \cdot 0.4 + 0.7 \cdot 0.1 \cdot 0.4}$
- D: $\dfrac{p}{0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.1 \cdot 0.4 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.9 \cdot 0.4}$
- E: $\dfrac{p}{0.3 \cdot 0.8 \cdot 0.4 + 0.3 \cdot 0.8 \cdot 0.6 + 0.7 \cdot 0.2 \cdot 0.4 + 0.7 \cdot 0.2 \cdot 0.6}$

38. Given the following (incomplete) transition matrix for a bigram model with characters "a" "b" "c", in a random string generated with the model, what is the probability that the second character is "c" given the first is "a". For example, row c column b of the matrix is the probability that a "b" follows an "c": $\mathbb{P}\{b|c\} = 0.3$.

−	a	b
a	0.4	0.1
b	0.3	0.2
c	0.2	0.3

- A: 0.2
- B: 0.3
- C: 0.4
- D: 0.5
- E: 0.6

39. Continue from the previous question, what is the probability that third character is "c" given the first is "a".

- A: 0.5
- B: $0.5 \cdot 0.5$
- C: $0.5 \cdot 0.5 \cdot 0.5$
- D: $0.4 \cdot 0.5 + 0.3 \cdot 0.5 + 0.2 \cdot 0.5$
- E: $0.1 \cdot 0.5 + 0.2 \cdot 0.5 + 0.3 \cdot 0.5$

40. Continue from the previous question, what is the probability that the second character is "c" given the first is "a" and the third is "c"? Assume the answer from the previous question is p.

- A: $\dfrac{0.5}{p}$
- B: $\dfrac{0.5 \cdot 0.5}{p}$
- C: $\dfrac{0.1 \cdot 0.5}{p}$
- D: $\dfrac{0.4 \cdot 0.5}{p}$
- E: $\dfrac{1}{p}$