Quiz 1: Probability Theory

Question 1

1 / 1 pts

Assume that the random variable X takes on values $\{1, 2.5, 4\}$ with corresponding probabilities $\{0.5, 0.2, 0.3\}$, what is the expected value of $E\left[2X - 3\right]$?

- 2.2
- 4.4
- 2.5
- 1.4

$E[X] = 1 * 0.5 + 2.5 * 0.2 + 4 * 0.3 = 2.2$

$E[2X - 3] = 2 * 2.2 - 3 = 1.4$

Question 2

1 / 1 pts

If you have three bags, labeled A and B. Bag A contains 3 red marbles and 5 blue marbles. Bag B contains 6 red marbles and 2 blue marbles. Bag C contains an equal number of red and blue marbles. The probability to pick each bag is 1/3. If you randomly draw a marble from one of the bags, what is the probability that the marble is red?

- $\frac{3}{8}$
- $\frac{13}{24}$
- $\frac{5}{8}$
- $\frac{7}{24}$

$$P(red) = P(A) * P(red|A) + P(B) * P(red|B) + P(C) * P(red|C) = \frac{1}{3} * \frac{3}{8} + \frac{1}{3} * \frac{6}{8} + \frac{1}{3} * \frac{1}{2} = \frac{13}{24}$$

Question 3

1 / 1 pts

If you have two bags, labeled A and B. Bag A contains 3 red marbles and 5 blue marbles. Bag B contains 6 red marbles and 2 blue marbles. Suppose a bag is randomly selected with equal probability(i.e. P(A) = P(B) = 1/2), but you don't know which it is. You randomly draw a marble and observe it is blue. What is the probability that the bag you selected this marble from is A? i.e., find P(A | blue).

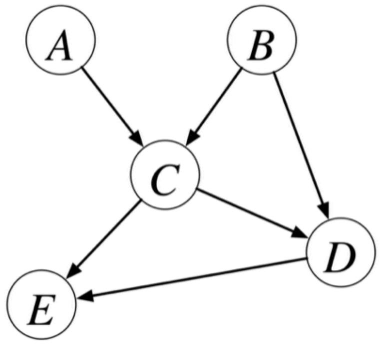
- ( ) $\frac{5}{16}$
- (•) $\frac{5}{7}$
- ( ) $\frac{3}{7}$
- ( ) $\frac{1}{8}$

$$P(blue|A) = \frac{5}{8}, P(A) = \frac{1}{2}, P(blue) = \frac{7}{16}$$

$$P(A|blue) = \frac{P(blue|A)*P(A)}{P(blue)} = \frac{\frac{5}{8}*\frac{1}{2}}{\frac{7}{16}} = \frac{5}{7}$$

Question 4

1 / 1 pts



The joint probability of variables ABCDE in the graphical model can be represented as:

A: P(A,B,C,D,E) = P(A)P(B | C,D)P(C | E,D)P(D | C,B)P(E | C,D)

B: P(A,B,C,D,E) = P(A)P(B | A)P(C | A,B)P(D | A,B,C)P(E | A,B,C,D)

C: P(A,B,C,D,E) = P(A)P(B)P(C | A,B)P(D | B,C)P(E | C,D)

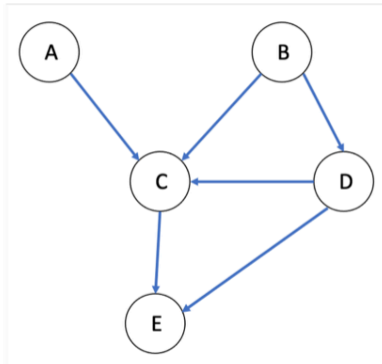
D: P(A,B,C,D,E) = P(A)P(B)P(C)P(D)P(E)

- [ ] A
- [x] B
- [x] C
- [ ] D

Question 5

1 / 1 pts

For the following structured probabilistic model:



Assume:

1. P(A) =0.5, P(B) = P(C) = 0.8, P(D) = P(E) = 0.6
2. P(D | B) = 0.5, P(E | C,D) = P(C | A,B,D) = 0.4

What is the probability of P(A, B, C, D, E)?

- 0.4
- 0.2
- 0.032
- 0.016

$$P(A, B, C, D, E) = P(A) * P(B) * P(C|A, B, D) * P(D|B) * P(E|C, D) = 0.5 * 0.8 * 0.4 * 0.5 * 0.4 = 0.032$$

Question 6

1 / 1 pts

With a sample $[(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)]$, you estimate $p(Y = y_i | x_i)$ by a normal distribution $N(\mu, \sigma)$. What is the Maximum Likelihood Estimate (MLE) of μ?

A: $\sum_i x_i$

B: $\frac{\sum_i (x_i - \sigma)^2}{n}$

C: $n \sum_i x_i$

D: $\frac{\sum_i x_i}{n}$

- ○ A
- ○ B
- ○ C
- ◉ D

Question 7

1 / 1 pts

Suppose you are classifying 3 samples into three possible classes. For each sample $i\,(1 \le i \le 3)$, its ground truth values (y_i), and prediction probabilities $(\hat{y}_i)$ for the three classes (c_1, c_2, c_3) are shown below.

$1 : y_1 = c_1, \hat{y}_1 = [0.7, 0.1, 0.2]$

$2 : y_2 = c_2, \hat{y}_2 = [0.2, 0.6, 0.2]$

$3 : y_3 = c_3, \hat{y}_3 = [0.1, 0.5, 0.4]$

Which equation shows the total cross entropy cost of these 3 samples?

(Hint: Consider using one-hot encoding to represent the ground truth label)

- ○ $-(log(0.7) + log(0.6) + log(0.5))$
- ◉ $-(log(0.7) + log(0.6) + log(0.4))$
- ○ $-(log(0.7) + log(0.1) + log(0.2) + log(0.2) + log(0.6) + log(0.2) + log(0.1) + log(0.5) + log(0.4))$
- ○ $-(log(0.7) + log(0.9) + log(0.8) + log(0.8) + log(0.6) + log(0.8) + log(0.9) + log(0.5) + log(0.6))$

According to the one-hot, the ground truth probabilities should be:

$y1 = [1, 0, 0], y2 = [0, 1, 0], y3 = [0, 0, 1]$

So: $CrossEntropy = -(log0.7 + log0.6 + log0.4)$

Question 8

1 / 1 pts

In a **multi-label classification** problem with three classes (c_1, c_2, c_3), a sample's the predicted probabilities for these three classes are [0.3, 0.8, 0.7] respectively, and it has ground truth labels [c_2, c_3]. Which of the following calculates the total cross entropy cost of this sample:

- `- (log(0.8) + log(0.7))`
- `- (log(0.7) + log(0.2) + log(0.3))`
- `- (log(0.3) + log(0.8) + log(0.7))`
- `- (log(0.7) + log(0.8) + log(0.7))` (selected)

This is multiple-label classification. We can each class separately.

For C1, ground truth probability: p(y=1) = 0 , p(y=0) = 1
cross entropy: - p(y=1) log 0.3 - p(y=0) log (1-0.3) = - log 0.7

For C2, ground truth probability: p(y=1) = 1 , p(y=0) = 0
cross entropy: - p(y=1) log 0.8 - p(y=0) log (1-0.8) = - log 0.8

For C3, ground truth probability: p(y=1) = 1 , p(y=0) = 0
cross entropy: - p(y=1) log 0.7 - p(y=0) log (1-0.7) = - log 0.7

So, the total cross entropy is D.

Question 9

1 / 1 pts

Suppose events = ['red', 'green', 'blue'], and P1, P2, P3 are three probability distributions over this event set:

P1 = [0.5, 0.3, 0.2]

P2 = [0.2, 0.2, 0.6]

P3 = [0.33, 0.33, 0.34]

Which one has the largest Shannon entropy?

- P1
- P2
- P3

Shannon entropy measures the uncertainty. If a distribution is more evenly distributed, it is more difficult to guess which outcome will happen, so there is higher entropy. Notice the probabilities in P3 are almost equal.

You can also validate as follows:

```python
from math import log

P1  = [0.5, 0.3, 0.2]
P2  = [0.2, 0.2, 0.6]
P3 = [0.33, 0.33, 0.34]

def Shannon(p):
    result = 0
    for i in range(len(p)):
        result -= p[i]*log(p[i])

    return result

"""
>>> Shannon(P1)
1.0296530140645737
>>> Shannon(P2)
0.9502705392332347
>>> Shannon(P3)
1.0985126170507196
"""
```

Question 10

1 / 1 pts

Suppose events = ['red', 'green', 'blue'], the probabilities of these events p = [0.5, 0.3, 0.2]. q_1, q_2, q_3 are three estimated probability distributions for this event set:

A: q1 = [0.6, 0.3, 0.1]

B: q2 = [0.3, 0.2, 0.5]

C: q_3 = [0.3, 0.3, 0.4]

Which one has the smallest cross entropy H(p, q) for q = q_1, q_2, q_3?

- (selected) A
- B
- C

Cross entropy measures how a probability distribution approximates another. The more similar the two distributions are , the smaller the cross entropy.

You can also validate as follows:

```
from math import log

p = [0.5, 0.3, 0.2]
q1 = [0.6, 0.3, 0.1]
q2 = [0.3, 0.2, 0.5]
q3 = [0.3, 0.3, 0.4]

def crossEntropy(p, q):
    result = 0
    for i in range(len(p)):
        result -= p[i]*log(q[i])

    return result

"""
>>> crossEntropy(p, q1)
1.0771216717795853
```

```
>>> crossEntropy(p, q2)
1.2234472120051871
>>> crossEntropy(p, q3)
1.1464363898355798
"""
```

Question 11

1 / 1 pts

Suppose events = ['red', 'green', 'blue'], and P1, P2, P3 are three probability distributions over this event set:

P1 = [0.4, 0.3, 0.3]

P2 = [0.75, 0.2, 0.05]

P3 = [0.3, 0.3, 0.4]

$D_{KL}(P1 \,||\, P2) > D_{KL}(P1 \,||\, P3)$?

- (•) True
- ( ) False

KL divergence measures how a probability distribution approximates another. The more similar the two distributions are , the smaller the divergence.

You can also validate as follows:

```python
from math import log

p1 = [0.4, 0.3, 0.3]
p2 = [0.75, 0.2, 0.05]
p3 = [0.3, 0.3, 0.4]

def KL(p, q):
    result = 0
    for i in range(len(p)):
        result -= p[i] * log(q[i]/p[i])

    return result

"""
>>> KL(p1, p2)
0.40772390943191605
```

```
>>> KL(p1, p3)
0.028768207245178132
"""
```

Question 12

1 / 1 pts

Which of the following are true:

A: For supervised classification problems, minimizing total cross entropy over all samples is equivalent to the maximum likelihood estimation of the samples

B: For supervised classification problems, minimizing cross entropy over all samples is equivalent to minimizing the KL divergence between the distributions of the ground truth labels and the predicted labels

C: $D_{KL}(p \,||\, q) = D_{KL}(q \,||\, p)$

D: $H(p, q) = D_{KL}(p \,||\, q) + H(p)$

- [x] A
- [x] B
- [ ] C
- [x] D