Q1.1 On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

A. $\frac{1}{4}+\frac{1}{3}$

B. $\frac{1}{4}\times\frac{1}{3}$

C. $\frac{1}{4}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{3}$

D. None of the above

Q1.1 On a multiple choice test, problem A has 4 choices, while problem B has 3. Assume that each problem has 1 correct answer. What is the probability of guessing the correct answer to both of the problems?

A. $\frac{1}{4} + \frac{1}{3}$

B. $\frac{1}{4} \times \frac{1}{3}$ 

C. $\frac{1}{4} \times \frac{1}{3} + \frac{1}{3} \times \frac{2}{3}$

D. None of the above

Q1.2 Consider a fair die, and the following three events:

X = rolling any of {1, 2}

Y = rolling any of {2, 4, 6}

Z = rolling any of {1, 4}

In other words, P(X) = 1/3, P(Y) = 1/2, P(Z) = 1/3.

Are events X and Y independent? Are events X and Y independent given event Z?

A. Yes, Yes

B. No, No

C. Yes, No

D. No, Yes

Q1.2 Consider a fair die, and the following three events:

X = rolling any of {1, 2}

Y = rolling any of {2, 4, 6}

Z = rolling any of {1, 4}

In other words, P(X) = 1/3, P(Y) = 1/2, P(Z) = 1/3.

Are events X and Y independent? Are events X and Y independent given event Z?

A. Yes, Yes

B. No, No

C. Yes, No 

D. No, Yes

Q2.1 We have a piece a text:

```
It was the best of times, it was the worst of times.
```

Suppose our vocabulary is `["it", "was", "best", "of", "times", "worst"]`

What is the bag of words representation of this text?

A. [2, 2, 1, 2, 2, 1]

B. [2, 2, 1, 2, 2, 1] / 6

C. [2, 2, 1, 2, 2, 1] / 10

D. [1, 1, 2, 1, 1, 2] / 10

Q2.1 We have a piece a text:

```
It was the best of times, it was the worst of times.
```

Suppose our vocabulary is `["it", "was", "best", "of", "times", "worst"]`

What is the bag of words representation of this text?

A. [2, 2, 1, 2, 2, 1]

B. [2, 2, 1, 2, 2, 1] / 6

C. [2, 2, 1, 2, 2, 1] / 10 

D. [1, 1, 2, 1, 1, 2] / 10

Q2.2 We have a a corpus containing only the following documents.

Document ID 1: "A time to plant and a time to reap"

Document ID 2: "Time for you and time for me"

Document ID 3: "Time flies"

Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

A. log(3)

B. log(3)/3

C. log(2)

D. log(2)/2

Q2.2 We have a a corpus containing only the following documents.

Document ID 1: "A time to plant and a time to reap"

Document ID 2: "Time for you and time for me"

Document ID 3: "Time flies"

Given that the stemmed version of the word "flies" is the term "fly", what is the tf-idf of "fly" in document 3?

A. log(3) 

B. log(3)/3

C. log(2)

D. log(2)/2

Q2.3 Given the following two document vectors, what is their cosine similarity?

v_a = [0.5, 1, 2]

v_b = [2, 1, 0.5]

A. 0.571

B. 0.99

C. 1.909

D. -0.99

Q2.3 Given the following two document vectors, what is their cosine similarity?

v_a = [0.5, 1, 2]

v_b = [2, 1, 0.5]

A. 0.571 

B. 0.99

C. 1.909

D. -0.99

Q3.1 Suppose “the dog ran away” is our training corpus.

What is P(ran away) if we use a unigram model?

A. 0

B. 1/2

C. 1/4

D. 1/16

Q3.1 Suppose “the dog ran away” is our training corpus.

What is P(ran away) if we use a unigram model?

A. 0

B. 1/2

C. 1/4

D. 1/16 

Q 3.2: Suppose “the dog ran away” is the training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

A. 1/4

B. 1

C. 2/5

D. 1/2

Q 3.2: Suppose “the dog ran away” is the training corpus. What is P(ran|dog) if we use a bigram model with Laplace Smoothing?

A. 1/4

B. 1

C. 2/5 

D. 1/2