Q 1. Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

a) 0.4
b) 0.5
c) 0.6
d) 0.3

Q 1. Consider a biased coin toss. If P(heads) = 0.6, then P(tails) = ?

a) 0.4
b) 0.5
c) 0.6
d) 0.3

Solution:
P(tails) = 1 - P(heads) = 1 - 0.6 = 0.4

Q 2. In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

a) 50%
b) 70%
c) 40%
d) 100%

Q 2. In a presidential election, there are 3 candidates, A, B and C. Based on our polling analysis, we estimate that A has a 30 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that either A or B win the election?

a) 50%
b) 70%
c) 40%
d) 100%

Solution:

P(A wins or B wins)
= P({A wins} U {B wins})
= P(A wins) + P(B wins) … (Note that A and B cannot win at the same time)
= 70 percent

Q 3. What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

a) 26 / 52
b) 4 / 52
c) 30 / 52
d) 28 / 52

Q 3. What is the probability of selecting a black card or a number 6 from a deck of 52 cards?

a) 26 / 52

b) 4 / 52

c) 30 / 52

d) 28 / 52 ⬅

Solution:

We need to find out P(card is black or card has number 6)

P(card is black) = 26/52 (either red or black)

P(card has number 6) = 4/52 (6 of clubs or 6 of diamonds or 6 of hearts or 6 of spades)

P(card is black and has number 6) = 2/52 (6 of clubs or 6 spades)

P({card is black} U {card has number 6})

= P(card is black) + P(card has number 6) - P({card is black} and {card has number 6})

= 26 / 52 + 4 / 52 - 2 / 52

= **28 / 52**

Q 4. Consider the joint probability distribution given below.
What is the probability that the temperature is hot given the weather is cloudy?

a) 40/365
b) 2/5
c) 3/5
d) 195/365

	weather = sunny	weather = cloudy	weather = rainy
temp= hot	150/365	40/365	5/365
temp = cold	50/365	60/365	60/365

Q 4. Consider the joint probability distribution given below.
What is the probability that the temperature is hot given the weather is cloudy?

a) 40/365
b) 2/5 ←
c) 3/5
d) 195/365

	weather = sunny	weather = cloudy	weather = rainy
temp= hot	150/365	40/365	5/365
temp = cold	50/365	60/365	60/365

Solution:

P(temp = hot | weather = cloudy)
= P(temp = hot, weather = cloudy) / P(weather = cloudy)

From the table, P(temp = hot, weather = cloudy) = 40/365
P(weather = cloudy) = P(temp = hot, weather = cloudy) + P(temp = cold, weather = cloudy) = 100/365

Hence, P(temp = hot | weather = cloudy) = (40/365) / (100/365) = **2/5**

Q 5. Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

a) 0.06
b) 0.3
c) 0.2
d) 0.24

Q 5. Of a company's employees, 30% are women and 6% are married women. Suppose an employee is selected at random. If the employee selected is a woman, what is the probability that she is married?

a) 0.06
b) 0.3
c) 0.2 ←
d) 0.24

Solution:

P(Employee selected is Married | Employee selected is a woman)
= P(Employee selected is Married and Employee selected is a woman) / P(Employee selected is a woman)
= 0.06 / 0.30 = 0.2

Q 6. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

a) 5 / 104
b) 95 / 100
c) 1 / 100
d) 1 / 2

Q 6. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

a) 5 / 104
b) 95 / 100
c) 1 / 100
d) 1 / 2

Solution:

Define events

A = event that an email is detected as spam,

B = event that an email is spam,

Bc = event that an email is not spam.

We are given that, P(B) = P(Bc) = 0.5,

$P(A \mid B) = 0.99$,

$P(A \mid B^c) = 0.05$.

Hence by the Bayes's formula, we have

$P(B^c \mid A)$

$= P(A \mid B^c)*P(B^c) / (P(A \mid B)*P(B) + P(A \mid B^c)*P(B^c))$

$= 0.05 \times 0.5 / (0.05 \times 0.5 + 0.99 \times 0.5)$

= **5 / 104**

Q 7. If a fair coin is tossed three times, find the probability of getting 2 heads and a tail.

a) 1 / 8
b) 2 / 8
c) 3 / 8
d) 5 / 8

Q 7. If a fair coin is tossed three times, find the probability of getting 2 heads and a tail.

a) 1 / 8
b) 2 / 8
c) 3 / 8 ←
d) 5 / 8

Solution:

P(H) = P(T) = 0.5
Each coin toss is independent of each other.
Hence, probability of getting 2 heads and a tail is given by

P(THH) + P(HTH) + P(HHT) = 3 * P(H) * P(H) * P(T) = **3 / 8**

Q 8. 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) 1/4 + 1/3
b) 1/4 * 1/3
c) 1/4*3/4 + 1/3*2/3
d) None of the above

Q 8. 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) 1/4 + 1/3
b) 1/4 * 1/3
c) 1/4*3/4 + 1/3*2/3
d) None of the above

Solution: The two events are independent.

Q 9. 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

Q 9. 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

$$P(X, Y) = P(X)P(Y) = \frac{1}{6}$$

So, X and Y are independent.

$$P(X|Z) = \frac{1}{2}, P(Y|Z) = \frac{1}{2}, P(X, Y|Z) = P(\{2\}|Z) = 0$$

So, X and Y are not conditionally independent given event Z.

Q 10. Bag-of-Words

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

Q 10. Bag-of-Words

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

$$Z = \sum_{w} c(w, d) = 2 + 2 + 1 + 2 + 2 + 1 = 10$$

Q 11. tf-idf

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) 1/3 log(3)
c) log(2)
d) 1/2 log(2)

Q 11. tf-idf

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) 1/3 log(3)
c) log(2)
d) 1/2 log(2)

Solution:
tf = 1
idf = log(3)

Q 12. Given the following two document vectors, what is their cosine similarity?

$$v_a = \begin{bmatrix} 0.5 \\ 1 \\ 2 \end{bmatrix} \quad v_b = \begin{bmatrix} 2 \\ 1 \\ 0.5 \end{bmatrix}$$

a) 0.571

b) 0.99

c) 1.909

d) -0.99

Q 12. Given the following two document vectors, what is their cosine similarity?

$$v_a = \begin{bmatrix} 0.5 \\ 1 \\ 2 \end{bmatrix} \quad v_b = \begin{bmatrix} 2 \\ 1 \\ 0.5 \end{bmatrix}$$

a) 0.571 ⬅

b) 0.99

c) 1.909

d) -0.99

Solution:

$$\cos\theta = \frac{v_a^T . v_b}{||v_a||_2 * ||v_b||_2} = \frac{v_a^T . v_b}{\sqrt{v_a^T . v_a} * \sqrt{v_b^T . v_b}} = \frac{0.5*2+1*1+2*0.5}{\sqrt{0.25+1+4} * \sqrt{4+1+0.25}} = \frac{3}{5.25} = 0.571$$

Q 13. Unigram

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 13. Unigram

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 ⬅

$$P(ran\ away) = P(ran)P(away) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$

Q 14. Smoothing

Suppose *the dog ran away* is our 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 14. Smoothing

Suppose *the dog ran away* is our 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

Solution:

P(ran|dog) = $\frac{|ran\ dog| + \alpha}{|dog| + \alpha * |V|} = \frac{1+1}{1+1*4} = \frac{2}{5}$

Q 15. Consider we choose number Uniformly from a set {1,2,3,4, …θ} (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

a) 3
b) 5
c) 7
d) 9

Q 15. Consider we choose number Uniformly from a set {1,2,3,4, …θ} (θ is an integer) with replacement. Suppose the numbers we choose are 2,5,7, then based on MLE, what value of θ do you estimate?

a) 3
b) 5
c) 7 ←
d) 9

Solution:

'Uniform' => $P(x|\theta) = 1/\theta$ if $x <= \theta$;
$P(x|\theta) = 0$ if $x > \theta$.

$P(2|\theta) * P(5|\theta) * P(7|\theta) = (1/\theta)^3$ if $\theta >= 7$

$P(2|\theta) * P(5|\theta) * P(7|\theta) = 0$ otherwise

Thus, to make it largest, we choose $\theta = 7$.

Q 16. For the above example, if beforehand we know that $\theta \sim Pr$, that $Pr(\theta=5) = 1/2$, $Pr(\theta=8) =1/6$, $Pr(\theta=9) = 1/2$. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

a) 5
b) 7
c) 8
d) 9

Q 16. For the above example, if beforehand we know that θ ~ Pr, that Pr(θ=5) = 1/2, Pr(θ=8) =1/6, Pr(θ=9) = 1/2. Then after we see the numbers we choose are 2,5,7, what value of θ do you think it is most likely to be?

a) 5
b) 7
c) 8
d) 9 ←

Solution:

θ = 5: P(2|5) * P(5|5) * P(7|5) * Pr(θ=5) = 0   since P(7|5) = 0
θ = 7: P(2|7) * P(5|7) * P(7|7) * Pr(θ=7) = 0   since P(θ=7) = 0
θ = 8: P(2|8) * P(5|8) * P(7|8) * Pr(θ=8) = $(1/8)^3$ * 1/6 = 0.000326
θ = 9: P(2|9) * P(5|9) * P(7|9) * Pr(θ=9) = $(1/9)^3$ * 1/2 = 0.000686
Thus, we choose θ = 9 to maximize

Q 17. Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \dots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 | Y = y) = y/46$, $P(x_2 = 1 | Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

a) 16
b) 26
c) 31
d) 32

Q 17. Consider a classification problem with $n = 32$, $y \in \{1, 2, 3, \ldots, n\}$, and two binary features, $x_1, x_2 \in \{0,1\}$. Suppose $P(Y=y) = 1/32$, $P(x_1 = 1 | Y = y) = y/46$, $P(x_2 = 1 | Y = y) = y/62$. Which class will naive Bayes classifier produce on a test item with $x_1 = 1$ and $x_2 = 0$?

a) 16

b) 26

c) 31

d) 32

Solution:

$P(y|x_1 = 1, x_2 = 0) \propto P(x_1 = 1, x_2 = 0 \mid y) * P(y) = P(x_1 = 1 \mid y) P(x_2 = 0 \mid y) * P(y)$

$= y/46 * (1-y/62) * 1/32$

Maximize above formula => $y = 31$

Q 18. Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data:

$P\{V=1\} = 0.92$,
$P\{A=1|V=1\} = 0.65$,
$P\{A=1|V=0\} = 0.9$,
$P\{B=1|V=1\} = 0.32$,
$P\{B=1|V=0\} = 0.78$,
$P\{C=1|V=1\} = 0.12$,
$P\{C=1|V=0\} = 0.94$.

Compute $P\{A=a,B=b,C=c\}$ for a,b,c = 1, 0, 1.

a) 0.0637
b) 0.0149
c) 0.0488
d) 0.0766

Q 18. Consider the problem of detecting if an email message contains a virus. Say we use four random variables to model this problem: Boolean (binary) class variable V indicates if the message contains a virus or not, and three Boolean feature variables: A,B,C. We decide to use a Naive Bayes Classifier to solve this problem so we create a Bayesian network with arcs from V to each of A,B,C. Their associated CPTs (Conditional Probability Table) are created from the following data:

P{V=1} = 0.92,
P{A=1|V=1} = 0.65,
P{A=1|V=0} = 0.9,
P{B=1|V=1} = 0.32,
P{B=1|V=0} = 0.78,
P{C=1|V=1} = 0.12,
P{C=1|V=0} = 0.94.

Compute P{A=a,B=b,C=c} for a,b,c = 1, 0, 1.

a) 0.0637 ←
b) 0.0149
c) 0.0488
d) 0.0766

Solution:

P(A=1,B=0, C=1) = P(A=1,B=0, C=1, V = 1) + P(A=1,B=0,C=1,V=0)
= P(A=1|V=1)P(B=0|V=1)P(C=1|V=1)P(V=1) + P(A=1|V=0)P(B=0|V=0)P(C=1|V=0) P(V=0)
= 0.65 * 0.68 * 0.12 * 0.92 + 0.9 * 0.22 * 0.94 * 0.08 = 0.0637

Q 19. Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

a) Pass
b) Fail

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Q 19. Consider the below dataset showing the result whether a person is pass or fail in the exam based on various factors. We want to classify an instance 'X' with Confident=Yes, Studied=Yes and Sick=No. Suppose the factors are independent to each other.

a) Pass ⬅
b) Fail

Confident	Studied	Sick	Result
Yes	No	No	Fail
Yes	No	Yes	Pass
No	Yes	Yes	Fail
No	Yes	No	Pass
Yes	Yes	Yes	Pass

Solution:

First we need to calculate the class probabilities i.e. P(Pass)=3/5 and P(Fail)=2/5

Now we need to calculate individual probability with respect to each features. For example,

P(Confident=Yes| Result=Pass) = 2/3

P(Studied=Yes| Result=Pass) = 2/3

P(Sick=No| Result=Pass) = 1/3

P(Confident=Yes| Result=Fail) = 1/2

P(Studied=Yes| Result=Fail) = 1/2

P(Sick=Yes| Result=Fail) = 1/2

P(Confident, Studied, not Sick | Result=Pass) * P(Result=Pass) = (2/3) * (2/3) * (1/3) * (3/5) = **0.089**

P(Confident, Studied, not Sick | Result=Fail) * P(Result=Fail) = (1/2) * (1/2) * (1/2) * (2/5) = **0.05**

P(Result=Pass | Confident, Studied, not Sick) > P(Result=Fail | Confident, Studied, not Sick)

⇨ Pass

Q 20. In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-6)^2}{2}\right)$

Their weights are also both in Normal distribution:

P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-9)^2}{2}\right)$,

P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$.

When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

a) Salmon
b) Tilapia

Q 20. In a lake, there are 2 kinds of fish, salmon and tilapia. Their lengths are both in Normal distribution: P(length=x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-3)^2}{2}\right)$, P(length=x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-6)^2}{2}\right)$

Their weights are also both in Normal distribution:

P(w = x | salmon) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-9)^2}{2}\right)$,

P(w = x | tilapia) = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-5)^2}{2}\right)$.

When you catch a fish, the chance to be a salmon is 0.8. If now you catch a fish with length 5 and weight 7, which fish do you think it would be? (suppose weight and length are independent).

a) Salmon
b) Tilapia ⬅

Solution:

P(salmon | w = 7, l = 5) ∝ P(w=7 | salmon) * P(l = 5 | salmon) * P(salmon)

$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(7-9)^2}{2}\right) * \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(5-3)^2}{2}\right) * 0.8$

= 0.0023

P(tilapia | w = 7, l = 5) ∝ P(w=7 | tilapia) * P(l = 5 | tilapia) * P(tilapia)

$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(7-5)^2}{2}\right) * \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(5-6)^2}{2}\right) * 0.2$

= 0.0026

So it’s more likely a tilapia.

Q21. Suppose x is a column vector. Is the equation $\| x \|_2^2 = x^T x$ correct?

a) Yes
b) No

Q21. Suppose x is a column vector. Is the equation $\| x \|_2^2 = x^T x$ correct?

a) Yes ⬅

b) No

Q22. Which following statements are correct? (I is the identity matrix)
(1) For any square matrix X, $XI = IX = X$
(2) For any square matrix X, $XX^Tv - \lambda v = (XX^T - \lambda I)v$
(3) If u_i is an eigenvector of square matrix A, then $Au_i = u_i$

a) (1)
b) (2)
c) (3)
d) (1)(2)
e) (1)(3)
f) (2)(3)
g) (1)(2)(3)

Q22. Which following statements are correct? (I is the identity matrix)
(1) For any square matrix X, $XI = IX = X$
(2) For any square matrix X, $XX^T v - \lambda v = (XX^T - \lambda I)v$
(3) If u_i is an eigenvector of square matrix A, then $Au_i = u_i$

a) (1)
b) (2)
c) (3)
d) (1)(2) ⟵
e) (1)(3)
f) (2)(3)
g) (1)(2)(3)

Solution:
The eigenvalue is missing in option (3).

Q23. If v is a unit column vector, which one is correct?

a) $v^T v = 1$
b) $\| v \|_2 = 1$
c) Both are correct

Q23. If v is a unit column vector, which one is correct?

a) $v^T v = 1$
b) $\| v \|_2 = 1$
c) Both are correct

Q24. If v_1, v_2, ..., v_d are principal components, which one is correct?

a) $v_1^T v_2 = 0$
b) $v_2^T v_d = 0$
c) $v_1^T v_1 = 1$
d) All are correct

Q24. If v_1, v_2, ..., v_d are principal components, which one is correct?

a) $v_1^T v_2 = 0$
b) $v_2^T v_d = 0$
c) $v_1^T v_1 = 1$
d) All are correct

Q25. Suppose we have a data matrix $X \in \mathbb{R}^{n \times p}$ where n is the number of data points and p is the number of features. After applying PCA, we keep the first k eigenvectors with largest eigenvalues and project the data. What is the dimension of the projected data?

a) $n \times k$
b) $n \times p$
c) $k \times p$

Q25. Suppose we have a data matrix $X \in \mathbb{R}^{n \times p}$ where n is the number of data points and p is the number of features. After applying PCA, we keep the first k eigenvectors with largest eigenvalues and project the data. What is the dimension of the projected data?

a) $n \times k$ ⬅
b) $n \times p$
c) $k \times p$

Solution:
After applying PCA, the data feature is reduced from p-dimension to k-dimension since we keep k principal components.

Q26. Consider the same setting as the previous question. We apply PCA on the data $X \in \mathbb{R}^{n \times p}$ and keep the first k principal components with largest eigenvalues. What is the dimension of each principal component?

a) n

b) p

c) k

Q26. Consider the same setting as the previous question. We apply PCA on the data $X \in \mathbb{R}^{n \times p}$ and keep the first k principal components with largest eigenvalues. What is the dimension of each principal component?

$a)$ n

$b)$ p ⬅

$c)$ k

Solution:
Each principal component has the same dimension as the feature of original data.