Suppose $\vec{X} = (X_1, X_2)$ is a random vector with a multivariate Gaussian distribution, and let C be that distribution's covariance matrix. Six random draws from the distribution are shown below:

X_1	X_2
1	4
3	2
2	5
-1	1
2	2
5	0

What is the maximum likelihood estimate for C_{12}?

The maximum likelihood estimate for the covariance of feature 1 and 2 is

$$C_{12} = \left(\frac{1}{6} \sum_{k=1}^{6} \vec{x}_1^{(k)} \vec{x}_2^{(k)} \right) - \mu_1 \mu_2$$
$$= \frac{1}{6}(4 + 6 + 10 - 1 + 4) - \frac{14}{3}$$
$$= -\frac{5}{6} \approx -0.833$$

Suppose a Gaussian with a full covariance matrix is fit to a data set of 1000 points, each a vector in $\mathbb{R}^{20}$. What will be the shape of the estimated covariance matrix?

( ) 1000×1000
(X) 20×20
( ) 1000×20
( ) 20×1000

Remember that covariance measures how much two variables vary together. It is also a square matrix. Thus, it must be 20×20, as there are 20 features.

Alternatively, one can recall that the covariance matrix can be calculated with the formula, $C = \frac{1}{n} X^T X$ where X is the centered data, and X is an $n \times d$ matrix. In this case, we have $n = 1000$ and $d = 20$. If we multiply $X^T X$, we will get a $d \times d$, or a 20×20 matrix.

Suppose X_1 measures the amount of natural gas used for heating on a given day, and X_2 measures the temperature on that day. What is likely to be the sign of the covariance of X_1 and X_2?

( ) Positive
(X) Negative

We know that as temperature decreases, i.e. the weather gets cooler, one may want to turn up the heater (under reasonable assumptions). This means a decrease in X_2 corresponds to an increase in X_1; there is a negative correlation or a negative covariance between X_1 and X_2.

True or False: the diagonal entries in a covariance matrix must be ≥ 0.

(X) True
( ) False

The diagonal entries in a covariance matrix are variances of each feature. Variances must be non-negative, as it does not make sense to have a negative measure of spread.

Suppose a data set of points in $\mathbb{R}^3$ consists of equally-many points from two classes: Class 1 and Class 0. The mean of the points in Class 1 is $(2, 1, 1)^T$, and the mean of the points in Class 0 is $(4, 0, 0)^T$. Suppose Linear Discriminant Analysis is applied using the same covariance matrix $C = \sigma^2 I$, for both classes, where σ is some constant.

What is the predicted label of a new point, $(3, 0, 0)^T$?

( ) Class 1
(X) Class 0

Because the class-conditional Gaussians have the same covariance matrix, and because $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0)$, the decision boundary is a straight line consisting of all points that are equidistant from the Class 0 mean and the Class 0 mean. That is, the predicted label of a new point will be given by whichever class mean is closer. In this case, (3, 0, 0) is closer to (4, 0, 0) than it is to (2, 1, 1), and so the prediction is for class 0.

Given the following events, use intuition to determine if A and B are conditionally independent given event C.

We said in lecture that real world events are very rarely *exactly* independent. Here, you should consider events to be independent if the dependence is very weak.

- Event A: People are using umbrellas.
- Event B: The sidewalk is wet.
- Event C: It is raining outside.

Is A independent of B given C?

(X) Yes
( ) No

If I know that it is raining, additionally knowing that the sidewalk is wet does not change my belief about whether people are using umbrellas. Thus $P(A|B, C) = P(A|C)$.

- Event A: There was an earthquake.
- Event B: There was a burglary.
- Event C: The house's alarm has gone off.

You can assume that a house alarm can be triggered by a burglary, but also by an earthquake.

Is A independent of B given C?

( ) Yes
(X) No

If the alarm sounds, there's some chance that it is was because of an earthquake and some chance that it was due to a burglary. These probabilities are $P(A|C)$ and $P(B|C)$. However, if I know that a burglary happened, it is less likely that an earthquake caused the alarm to sound; thus $P(A|B, C) < P(A|C)$, and the events are *not* conditionally independent.

The table below shows whether or not (Y/N) it was sunny, warm, and windy on ten days.

Sunny	Warm	Windy
Y	Y	Y
Y	Y	Y
Y	Y	N
Y	Y	Y
Y	N	N
Y	N	Y
N	N	Y
N	N	N
N	N	N
N	Y	Y

Suppose a new day is not warm but is windy. Does a Naive Bayes classifier using probabilities estimated from this table predict that the new day is sunny or not sunny?

( ) Sunny
(X) Not Sunny

In this scenario, we want to make predictions based on which is larger: $\mathbb{P}(\text{sunny}|\neg\text{warm, is windy})$ vs. $\mathbb{P}(\neg\text{sunny}|\neg\text{warm, is windy})$.

To do this, we can just compare the numerators of the Bayes' formula and also leverage the conditional independence assumption in Naïve Bayes. So,

$$\begin{aligned}\mathbb{P}(\text{sunny}|\neg\text{warm, is windy}) &\propto \mathbb{P}(\neg\text{warm}|\text{sunny}) \cdot \mathbb{P}(\text{windy}|\text{sunny}) \cdot \mathbb{P}(\text{sunny}) \\ &= \frac{2}{6} \cdot \frac{4}{6} \cdot \frac{6}{10} \\ &= \frac{2}{15} \\ \mathbb{P}(\neg\text{sunny}|\neg\text{warm, is windy}) &\propto \mathbb{P}(\neg\text{warm}|\neg\text{sunny}) \cdot \mathbb{P}(\text{windy}|\neg\text{sunny}) \cdot \mathbb{P}(\neg\text{sunny}) \\ &= \frac{3}{4} \cdot \frac{2}{4} \cdot \frac{4}{10} \\ &= \frac{3}{20}\end{aligned}$$

Since $\frac{3}{20} > \frac{2}{15}$, Naïve Bayes will predict "Not Sunny".

A survey asks each person to rate their happiness level on a scale from 1 to 10. It also asks them how much homework they have. Is the covariance between these two variables likely to be positive or negative (assuming that they are a normal person)?

( ) Positive
(X) Negative

Perhaps one can relate that as one's homework increases, one's happiness decreases 😦
This means there is a negative covariance between these two variables.

Suppose Gaussian Naive Bayes is trained on a data set of 200 points in $\mathbb{R}^4$ from two classes: 1 and 0.

How many one-dimensional Gaussians will be learned (by estimating parameters with maximum likelihood)?

____

Since there are four features and two classes, we will have $4 \times 2 = 8$ univariate Gaussians to learn.

Let X be the number of ice cream cones sold on a particular day, and let Y be the number of people at the beach. Are these random variables dependent or independent?

(X) Dependent
( ) Independent

These variables are dependent because knowing that there are more vs less people on the beach can tell us that there will be more vs less ice cream sales respectively. Moreover, a potential confounder could be the weather.

Suppose you have two, 6-sided dice, each labeled with the numbers 1 through 6. You roll both dice; let X be the number on the first die, and let Y be the number on the second. Let Z be the event that the sum of the two die is odd. Which is true?

Are X and Y are conditionally independent given Z?

( ) Yes
(X) No

If Z is odd, knowing what X is gives additional information about what Y could be, namely that Y must be odd if X is even and vice versa.

Suppose a Gaussian Naive Bayes classifier is trained on a data set in $\mathbb{R}^2$.

True or False: the decision boundary is guaranteed to be linear.

( ) True
(X) False

Gaussian Naive Bayes is equivalent to Quadratic Discriminant Analysis when the covariance matrices are diagonal. In general, the decision boundary is therefore not linear, but quadratic.