1 Probability Review

You should be familiar with basic probability definitions, but as a quick run-down:

(a) *Random Variables*: In this class, we will be using random variable notation. A random variable can take on any state in a distribution. For example, if Z represents the number you roll in a six-sided fair dice, the probability of rolling a 2 will be denote as the following:

$$P(Z = 2) = \frac{1}{6}$$

(b) *Indicator Variables*: If A is an event, 1_A is a random variable such that

$$1_A = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false} \end{cases}$$

Important!

$E[1_A] = P(A)$ because by the definition of expected value, $E[1_A] = 1 * P(A) + 0 * P(\overline{A}) = P(A)$

(c) *Conditional probability*: Probability distribution of a variable given another variable takes on a certain value

$$P(A = a|Y = b) = \frac{P(A = a, Y = b)}{P(Y = b)}$$

(d) *Marginalization (Sum Rule)*: Probability distribution of a single variable in a joint distribution

$$P(A = a) = \sum_{b\,=\,\text{all values of } Y} P(A = a, Y = b)$$

(e) *Law of Total Probability*: Using conditional probability we can rewrite the above expression for marginalization as:

$$P(A = a) = \sum_{b} P(A = a|Y = b) * P(Y = b)$$

(f) *Continuous vs discrete random variables*: Discrete random variables can only take a countable number of values (eg, values we can roll in a dice) while continuous can take on infinitely many values. Our definitions for marginalization and law of total probability assumed Y was a discrete random variable. If Y is not:

$$P(A = a) = \int_{Y} P(A = a, Y = b) * db$$

2 Independence

2.1 Independent Random Variables

1. Two random variables A and B are independent and can each take on two values a_1, a_2 and b_1, b_2 respectively.
   - What is $P(A = a_1, B = b_2)$?
   - What is $P(A = a_1, B = b_1)$?
   - What is $P(A = a_1 | B = b_1)$?

2. Let X, Y, and Z be random variables taking values in $\{0, 1\}$. The following table lists the probability of each possible assignment of 0 and 1 to the variables X, Y, and Z:

	$Z = 0$		$Z = 1$	
	$X = 0$	$X = 1$	$X = 0$	$X = 1$
$Y = 0$	0.1	0.05	0.1	0.1
$Y = 1$	0.2	0.1	0.175	0.175

   Is X independent of Z?

2.2 Conditional Independence

Two random variables A and B are conditionally independent given C if they are independent after conditioning on C

$$P(A, B|C) = P(B|A, C) * P(A|C) = P(B|C) * P(A|C)$$

1. Revisiting the table from the previous section, is X conditionally independent of Y given Z?

2.3 Bayes Rule

We can derive Bayes Rule using our definition of conditional probability, product rule, and LoTP

$$P(A|B) = \frac{P(A, B)}{P(B)} = \frac{P(B|A) * P(A)}{P(B)}$$

1. At a university, 51% of the students are commuters while the rest live on campus. One student is randomly selected for a survey. It later turns out that the selected student is studying sciences. Also, 10% of commuter students study sciences while 5% of other students study sciences. What's the probability that the selected student is a commuter?

3 Correlation

Two random variables A and B are correlated if and only if

$$Cov(X, Y) \neq 0$$

Recall that

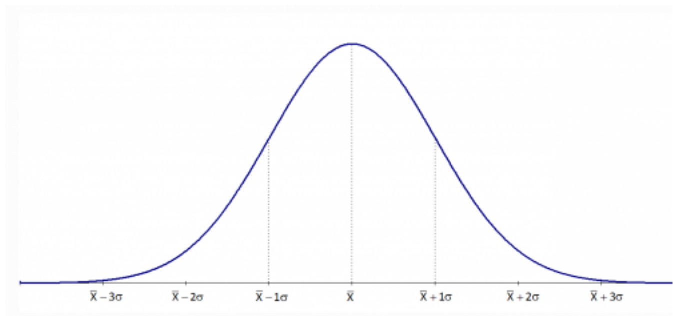
$$Cov(X, Y)! = E[XY] - E[X]E[Y]$$

1. Suppose X and Y can take on the joint values (0,0), (1,0), and (1,1) with equal probability. Are X and Y correlated?

2. Suppose X and Y can take on the joint values (-1,1), (0,0), and (1,1) with probability $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ respectively. Are X and Y correlated? Are they independent?

4 Visualizing Gaussian Distributions

The general shape of a Gaussian is: $a * e^{-b*(x-\mu)^2}$, where a, b are some factors. Gaussian distributions should look something like this:



4.1 What affects "spread" of the distribution?

The "spread" of a Gaussian is determined by the factor multiplied to $(x-\mu)^2$, which in our case is $b = \frac{1}{2\sigma^2}$. A smaller the σ value will create a distribution that is more "concentrated" to the center. Take some time to think why that is.

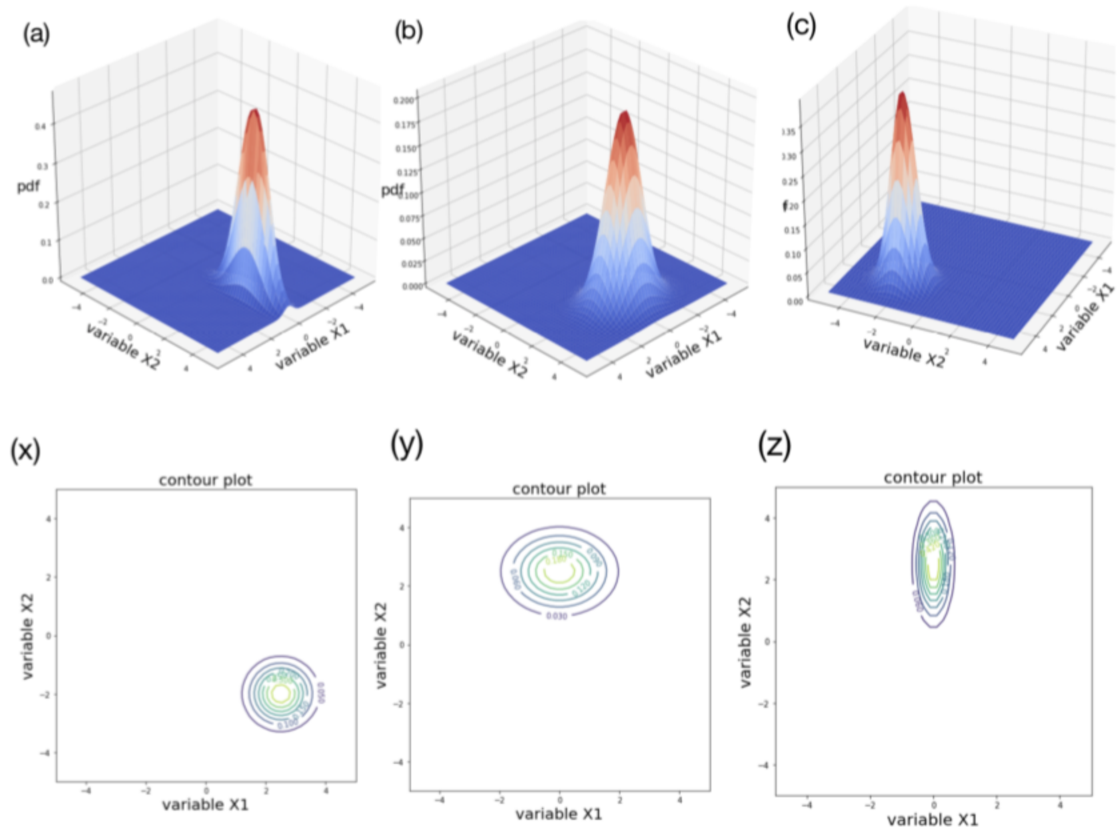
4.2 What affects the position of the distribution?

The Gaussian will be centered around μ. At $\mu = 5$, the tip of the Gaussian would be x when $-b(x-5)^2 = 0$, hence the distribution will be centered around 5.

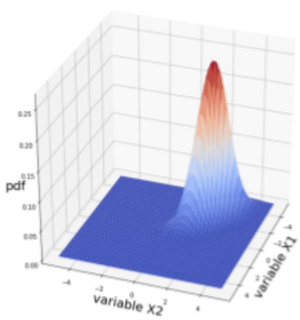
4.3 Multivariate Gaussian: What affects the "spread" of the distribution across the diagonal?

The covariance between variables (v_1, v_2) affects the spread of the distribution across the diagonal.
A distribution that is more concentrated along the diagonal would have larger covariance between the two variables A negative covariance between two variables would reverse the direction of the diagonal.

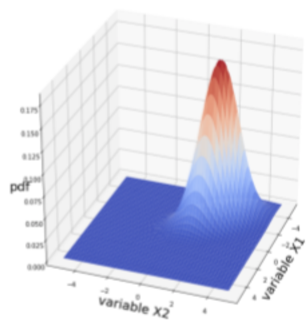
1. For each surface plot, (1) find the corresponding contour plot (2) use the plotting tool provided to find the parameter($\boldsymbol{\mu}, \boldsymbol{\Sigma}$) of the distribution.



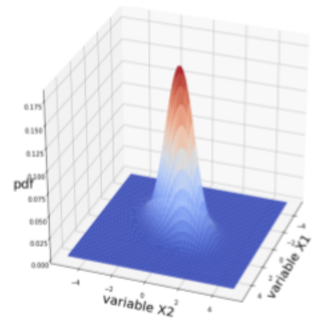
2. For each surface plot, find the corresponding contour plot and the corresponding parameters.



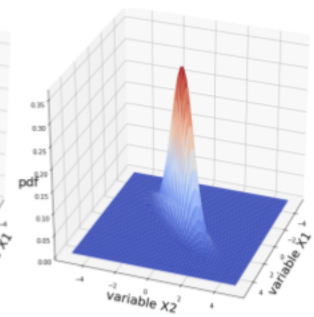
(a)



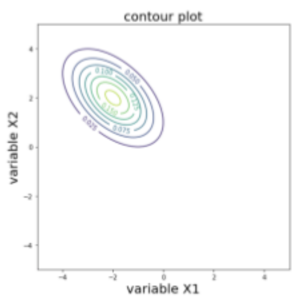
(b)



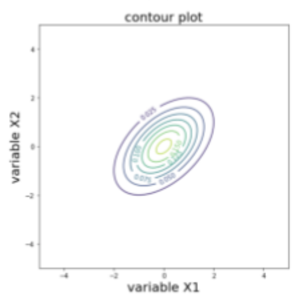
(c)



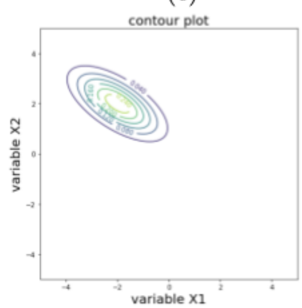
(d)



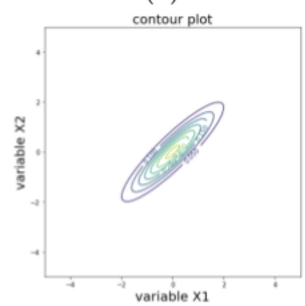
(h)



(i)



(j)



(k)

(x)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

(y)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

(z)

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

(w)

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 0.6 \end{bmatrix}$$