Probability Review

- I'm assuming you know all of these words:
  - Probability
  - Random Variable, Indicator RV
  - Joint Probability → [fancy way of saying $P(X \cap Y)$]
  - Conditional Probability
  - Complementary events
  - Intersection of events
  - Law of total probability → (partitions!)

$$P(x) = \sum_{y} \underbrace{P(y)P(X|y)}_{P(x \cap y)}$$

  - Independence of events

$$P(x \cap y) = P(X)P(y)$$

or

$$P(x|y) = P(X).$$

★ Correlated / Uncorrelated events

X and Y are uncorrelated iff. Covariance$(X,Y) = 0$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Correlated Example:

For an example, suppose X and Y can take on the joint values (expressed as ordered pairs (0,0), (1,0), and (1,1) with equal probability. Then for any of the three possible points (x, y), $P((X, Y) = (x, y)) = 1/3$. We will find the covariance between these two random variables.

The first step is to calculate the mean of each individual random variable. X only takes on two values, 0 and 1, with probability 1/3 and 2/3 respectively. (Remember that two of the points have $X = 1$, with each of those probabilities as 1/3.) Then

$$E[X] = 0 \cdot 1/3 + 1 \cdot 2/3 = 2/3$$

Similarly, $E[Y] = 0 \cdot 2/3 + 1 \cdot 1/3 = 1/3$. Now, we must calculate the expected value of the *product* of X and Y. That product can take on values 0 or 1 (multiply the elements of each ordered pair together) with respective probabilities 2/3 and 1/3. These probabilities are obtained the same way as for the individual expectations. Thus,

$$E[XY] = 0 \cdot 2/3 + 1 \cdot 1/3 = 1/3$$

Finally, we put it all together:

$$\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{2}{3} \cdot \frac{1}{3} = \frac{1}{3} - \frac{2}{9} = \frac{1}{9}$$

Uncorrelated Example

Let's take X and Y to exist as an ordered pair at the points (-1,1), (0,0), and (1,1) with probabilities 1/4, 1/2, and 1/4. Then $E[X] = -1 \cdot 1/4 + 0 \cdot 1/2 + 1 \cdot 1/4 = 0 = E[Y]$ and

$$E[XY] = -1 \cdot 1/4 + 0 \cdot 1/2 + 1 \cdot 1/4 = 0 = E[X]E[Y]$$

and thus X and Y are uncorrelated.

Exercise: See if they're independent!!

- Bayes' Rule → Important in this class [and life?]

$$P(A_1|B) = \frac{P(A_1) \cdot P(B|A_1)}{P(B)}$$

and if we have $A_1, A_2 \cdots A_n$ form a partitioning

then $P(B) = \sum_i P(A_i) \cdot P(B|A_i)$

Bayes Example:

In a public university, 51% of the students are females. One adult is randomly selected for a survey. It turned out later that the selected survey subject was studying sciences. Also, 10% of female students study sciences while 5% of males study sciences. What is the probability that the selected subject is a female? Let's use the following notations:

F = female
F' = male
S = study sciences
S' = study other fields

$P(F)$=
$P(F')$=
$P(S|F)$=
$P(S|F')$=

We want ???

SOL:
$P(F)$= 0.51 because 51% of students are females
$P(F')$= 0.49 because 49% of students are males
$P(S|F)$=0.1 because 10% of the female students study sciences
$P(S|F')$=0.05 because 5% of the male students study sciences

We want P(F|S)

answer = (0.1 * 0.51) / (0.1*0.51 + 0.05*0.49) = 0.6755

3 Gaussian Distribution

Review: 1-D Gaussian Distribution

The probability density function of $\mathcal{N}(\mu, \sigma^2)$ is given by:

$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

Multivariate Gaussian Distribution

The multivariate Gaussian distribution in M dimensions is parameterized by a **mean vector** $\boldsymbol{\mu} \in \mathbb{R}^M$ and a **covariance matrix** $\boldsymbol{\Sigma} \in \mathbb{R}^{M \times M}$, where $\boldsymbol{\Sigma}$ is a symmetric and positive-definite. This distribution is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and its probability density function is given by:

$$p(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^M |\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

where $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$.

Let $\boldsymbol{X} = [X_1, X_2, ..., X_m]^T$ be a vector-valued random variable where $\boldsymbol{X} = [X_1, X_2, ..., X_m]^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, we have:

$$\boldsymbol{\Sigma} = Cov[\boldsymbol{X}] = \begin{bmatrix} Cov[X_1, X_1] = Var[X_1] & Cov[X_1, X_2] & \dots & Cov[X_1, X_M] \\ Cov[X_2, X_1] & Cov[X_2, X_2] = Var[X_2] & \dots & Cov[X_2, X_M] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_M, X_1] & Cov[X_M, X_2] & \dots & Cov[X_M, X_M] = Var[X_M] \end{bmatrix}$$

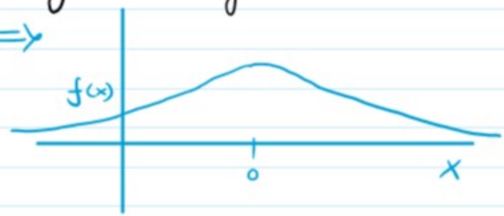
Note: Any arbitrary covariance matrix is positive semi-definite. However, since the pdf of a multivariate Gaussian requires Σ to have a strictly positive determinant, Σ has to be positive definite.

In order to get get an intuition for what a multivariate Gaussian is, consider the simple case where $M = 2$. Then, we have:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & Cov[X_1, X_2] \\ Cov[X_1, X_2] & \sigma_2^2 \end{bmatrix}$$

* General shape of Gaussian: Something something e^{-x^2}

$\Rightarrow$



* What affects the pointy-ness? : The pointy-ness of the peak is determined by the factor of multiplication with x^2 (so here, $\frac{1}{2\sigma^2}$)

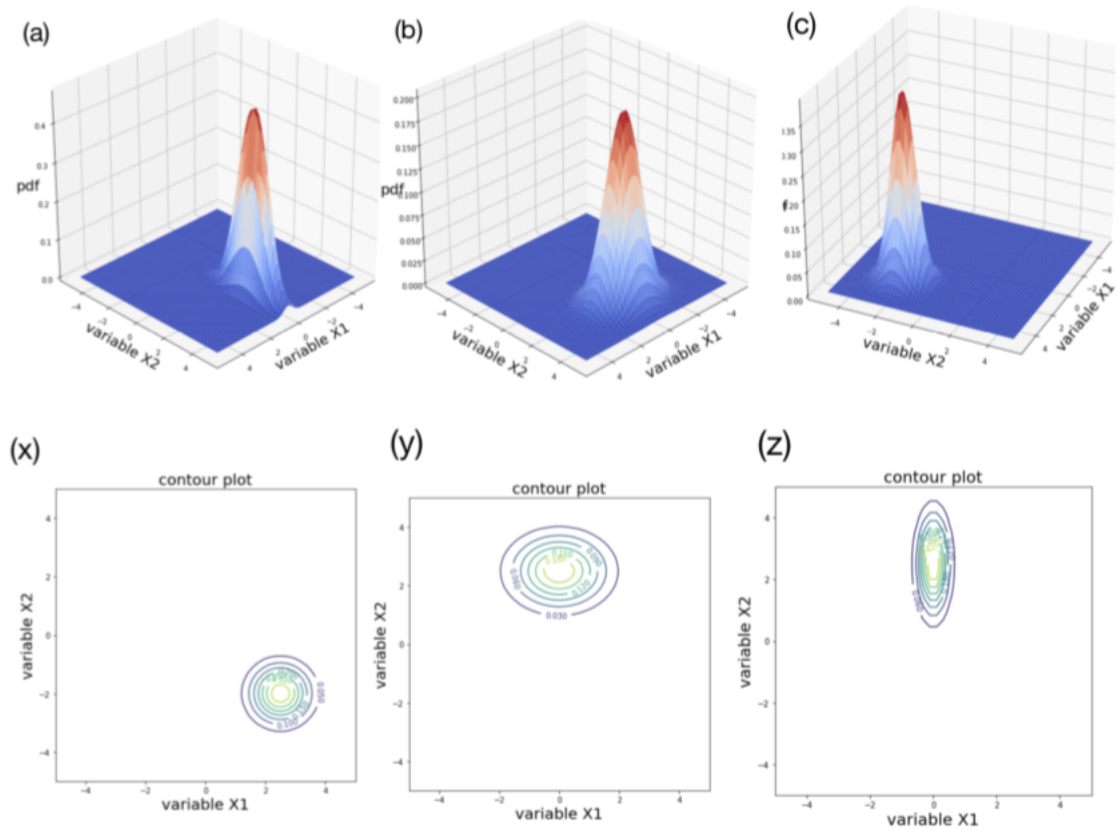
Makes sense: smaller σ $\Rightarrow$ concentrated in one place, pointier graph.

* What affects position? : The addition factor on x. (so here $-\mu$)

$\Rightarrow$ if mean is 5, then at $x-5=0$ we will have the tip of the curve.

Makes sense.

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.



(a) → (z)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 2.5 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix}$$

(b) → (y)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 2.5 \end{bmatrix}$$

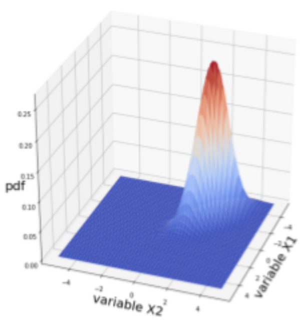
$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix}$$

(c) → (x)

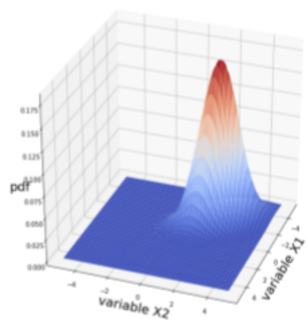
$$\boldsymbol{\mu} = \begin{bmatrix} 2.5 \\ -2 \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix}$$

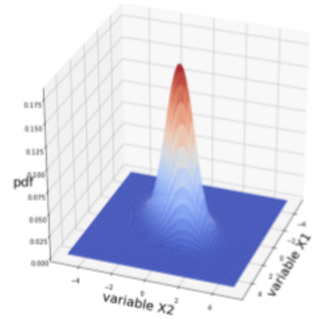
2. For each surface plot, find the corresponding contour plot and the corresponding parameters.



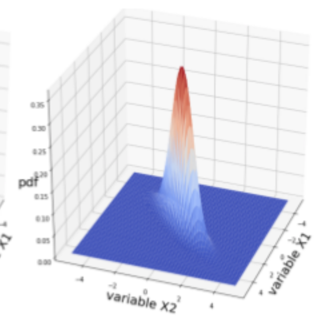
(a)



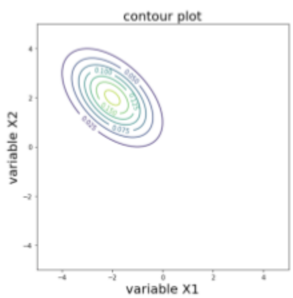
(b)



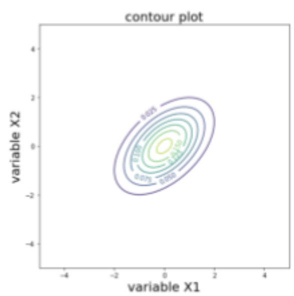
(c)



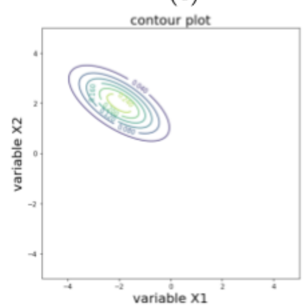
(d)



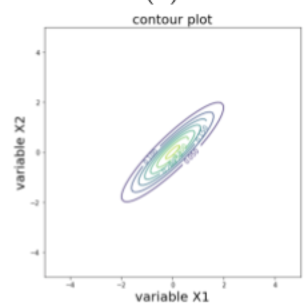
(h)



(i)



(j)



(k)

(x)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

(y)

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$$

(z)

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

(w)

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 0.6 \end{bmatrix}$$

(a)→(j)→(w) (b)→(h)→(z) (c)→(i)→(x) (d)→(k)→(y)

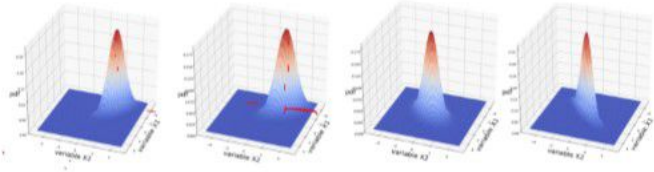
* This type of question can come in your exams/quizzes:

"Predict $\vec{\mu}$ and Σ based on graph of Gaussian". Just remember what makes it pointy and what displaces it!

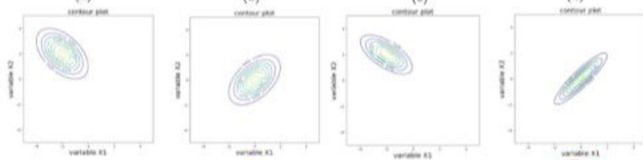
* Also, visualization beyond 3D is impossible for most people so don't worry about it.

Source: S20 Rec2

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)
$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$
$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 0.6 \end{bmatrix}$

(a)→(j)→(w) (b)→(h)→(z) (c)→(i)→(x) (d)→(k)→(y)

The pointy-ness along the diagonal representing $(X_1 = X_2)$ shows the magnitude of $Cov(X_1, X_2)$. If $cov(X_1, X_2)$ is high the graph is more concentrated along that diagonal. If $cov(X_1, X_2)$ is negative, direction of diagonal is reversed, ie, $(X_1 = -X_2)$