CSC 477 Tutorial Probability Refresher

Radian Gondokaryono

October 5, 2022

Slides adapted from:
CSC 2515 Introduction to Machine Learning. Credit: Amir Massoud Farahmand, et. Al CSC 2515 Fall 2021: Introduction to Machine Learning | IntroML-Fall2021 (amfarahmand.github.io)
and CSC477 (toronto.edu) Week 7

Outline

- Probability Overview
- Bayes Rule
- Expectation and Variance
- Gaussian Distributions
- Covariance

Motivation

Uncertainty arises through:

- Noisy measurements
- Variability between samples
- Finite size of data sets

Probability provides a consistent framework for the quantification and manipulation of uncertainty.

Sample Space

Sample space Ω is the set of all possible outcomes of an experiment.

Observations $\omega \in \Omega$ are points in the space also called sample outcomes, realizations, or elements.

Events $E \subset \Omega$ are subsets of the sample space.

In this experiment we flip a coin twice:

Sample space All outcomes $\Omega = \{HH, HT, TH, T\,T\}$

Observation $\omega = HT$ valid sample since $\omega \in \Omega$

Event Both flips same $E = \{HH, T\,T\}$ valid event since $E \subset \Omega$

Probability

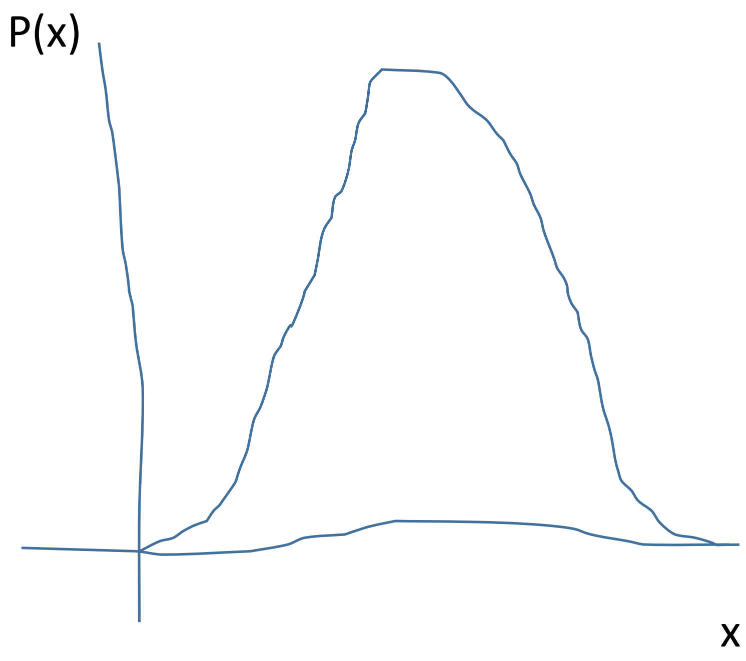
The probability of an event E, P(E), satisfies three axioms:

1: P(E) ≥ 0 for every E

2: P(Ω) = 1

3: If E1, E2, . . . are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

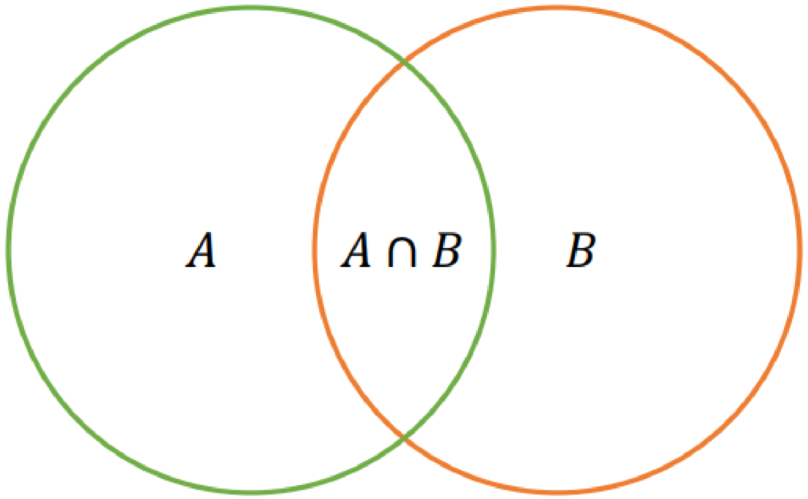


Joint and Conditional Probabilities

Joint Probability of A and B is denoted P(A, B).

Conditional Probability of A given B is denoted P(A|B).

Joint: $p(A, B) = p(A \cap B)$



Conditional: $p(A|B) = \frac{p(A \cap B)}{p(B)}$

$$p(A, B) = p(A|B)p(B) = p(B|A)p(A)$$

Conditional Example

Probability of passing the midterm is 60% and probability of passing both the final and the midterm is 45%. What is the probability of passing the final given the student passed the midterm?

$$
\begin{aligned}
P(F|M) &= P(M,F)/P(M) \\
&= 0.45/0.60 \\
&= 0.75
\end{aligned}
$$

Independence

$P(A|B) = P(A)$

Events A and B are independent if $P(A, B) = P(A)P(B)$.

* Independent: A: first toss is HEAD; B: second toss is HEAD;

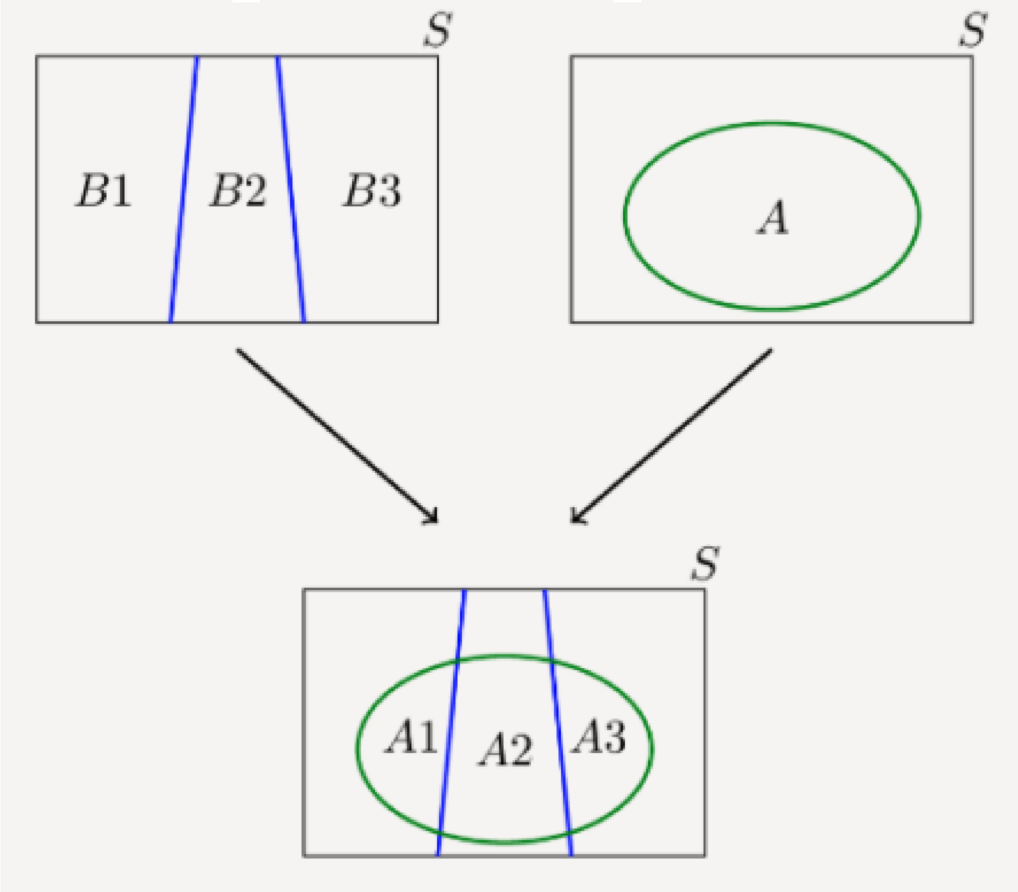
$$P(A, B) = 0.5 * 0.5 = P(A)P(B)$$

- Not Independent: A: first toss is HEAD; B: first toss is HEAD;

$$P(A, B) = 0.5 \,!= P(A)P(B)$$

Marginalization and Law of Total Probability

Law of Total Probability $$P(X) = \sum_Y P(X,Y) = \sum_Y P(X|Y)P(Y)$$



[Law of Total Probability | Partitions | Formulas (probabilitycourse.com)]

Bayes' Rule

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

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

Probability Distribution Statistics

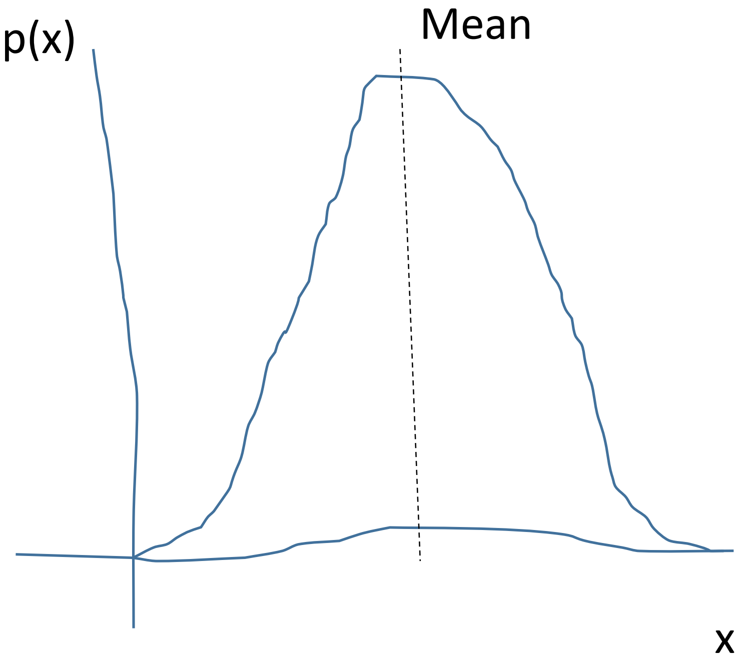
Mean: First Moment, μ

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i p(x_i) \qquad \text{(univariate discrete r.v.)}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x p(x) dx \qquad \text{(univariate continuous r.v.)}$$

Variance: Second (central) Moment, σ^2

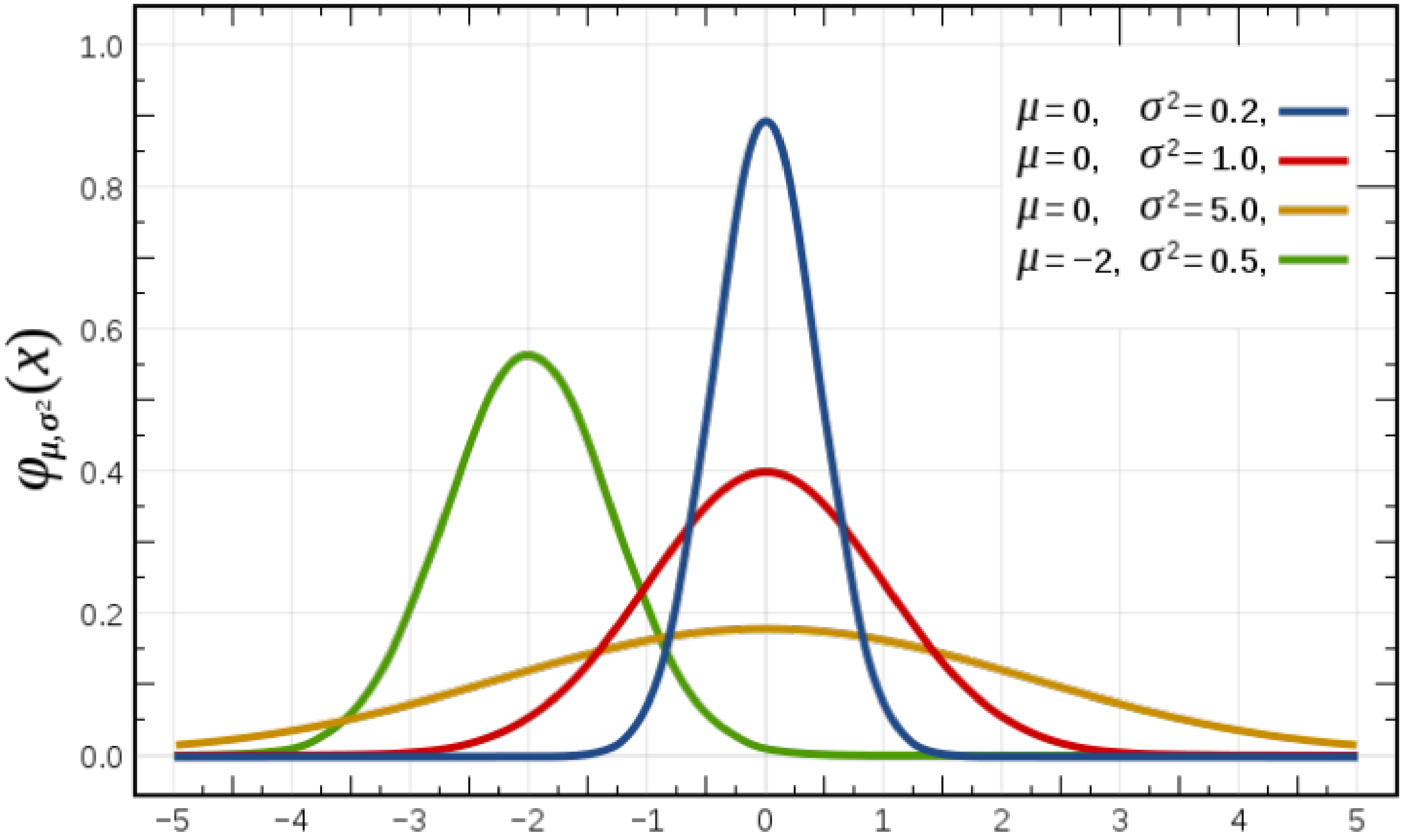
$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2\right] \\
&= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2
\end{aligned}$$



Univariate Gaussian Distribution

Also known as the Normal Distribution, $\mathcal{N}(\mu, \sigma^2)$

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Multivariate Gaussian Distribution

Multidimensional generalization of the Gaussian.

$\mathbf{x}$ is a D-dimensional vector

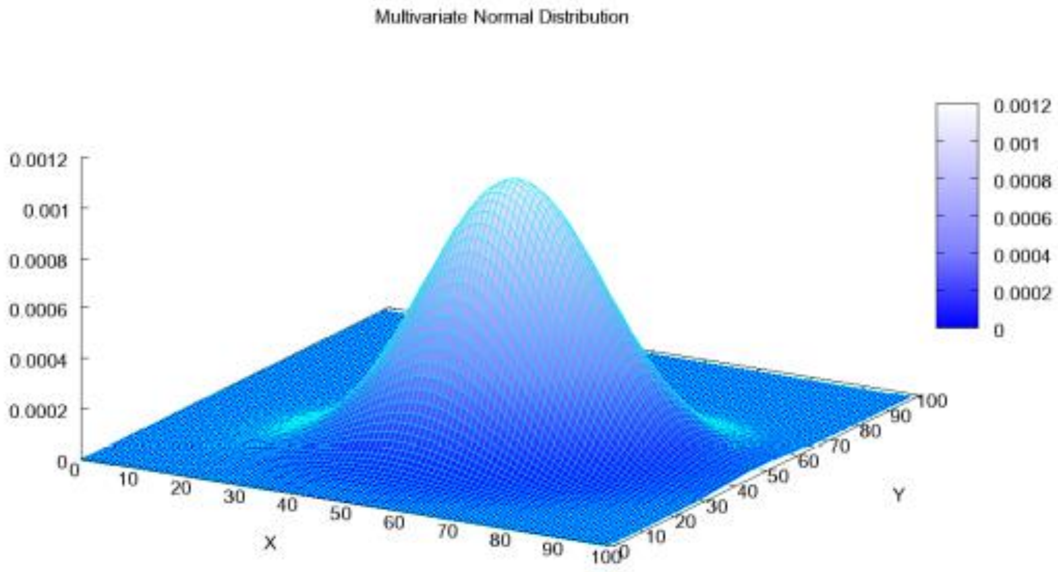
μ is a D-dimensional mean vector

Σ is a $D \times D$ covariance matrix with determinant $|\Sigma|$

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right)$$

$$-\frac{1}{2}\left(\begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix}\right)^T \begin{bmatrix}5 & 2\\2 & 4\end{bmatrix} \quad \begin{bmatrix}2\\3\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix}$$

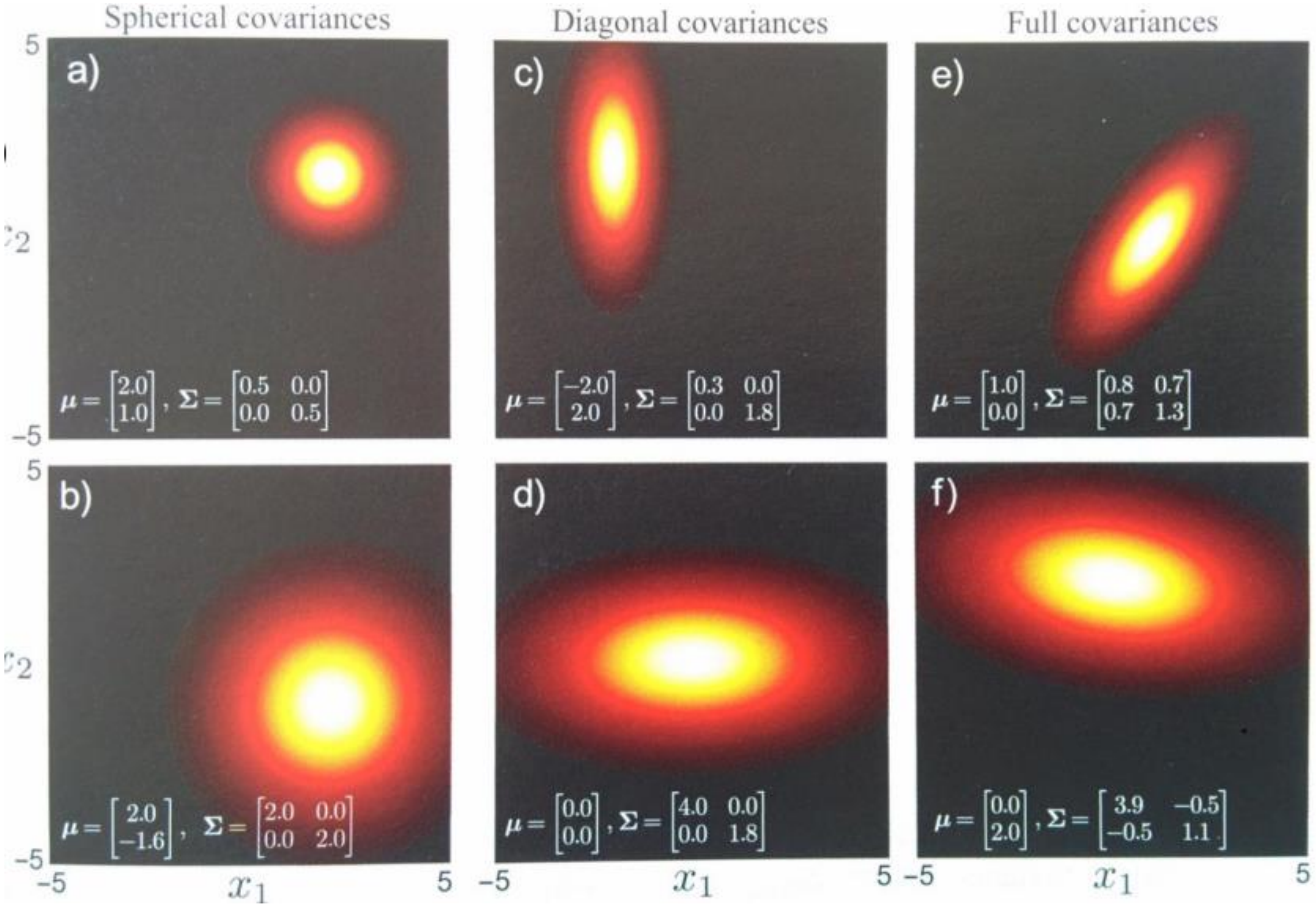
$$-\frac{1}{2}\left(\begin{bmatrix}1\\1\end{bmatrix}\right)^T \quad \begin{bmatrix}5 & 2\\2 & 4\end{bmatrix} \quad \begin{bmatrix}1\\1\end{bmatrix}$$



Shortcut notation: $||\mathbf{x}||^2_{\mathbf{\Sigma}} = \mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x}$

Multivariate Gaussian Distribution

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right)$$



From "Computer Vision: Models, Learning, and Inference" Simon Prince

Covariance

- Measures linear dependence between random variables X, Y. Does **not** measure independence.

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

- Variance of X

$$\text{Var}[X] = \text{Cov}[X] = \text{Cov}[X, X] = E[X^2] - E[X]^2$$

$$\text{Cov}[AX + b] = A\text{Cov}[X]A^T$$

$$\text{Cov}[X + Y] = \text{Cov}[X] + \text{Cov}[Y] - 2\text{Cov}[X, Y]$$

Covariance Matrix

$$\text{Var}[X] = \sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{Cov}[X,Y] = \sigma(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\begin{bmatrix} Var[X] & Cov[X,Y] \\ Cov[Y,X] & Var[Y] \end{bmatrix} = \Sigma = \begin{pmatrix} \sigma(x,x) & \sigma(x,y) \\ \sigma(y,x) & \sigma(y,y) \end{pmatrix}$$

Understanding the Covariance Matrix | DataScience+ (datascienceplus.com)

Covariance Matrix

- Measures linear dependence between random variables X, Y. Does **not** measure independence.

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

- Entry (i,j) of the covariance matrix measures whether changes in variable X_i co-occur with changes in variable Y_j

- It does not measure whether one causes the other.