# Probability Theory Review

STATS 415: Data Mining and Machine Learning

University of Michigan

Fall 2022

# The Multivariate Gaussian Distribution

One particularly important example of a probability distribution over random vectors $X$ is called the multivariate Gaussian or multivariate normal distribution. A random vector $X \in \mathbf{R}^d$ is said to have a multivariate normal (or Gaussian) distribution with mean $\mu \in \mathbf{R}^d$ and covariance matrix $\Sigma \in \mathbf{S}^d_{++}$ (where $\mathbf{S}^d_{++}$ refers to the space of symmetric positive definite $d \times d$ matrices)

$$f_{X_1, X_2, \ldots, X_d}(x_1, x_2, \ldots, x_d; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right). \tag{0.1}$$

We write this as $X \sim \mathcal{N}(\mu, \Sigma)$. In this section, we describe multivariate Gaussians and some of their basic properties.

# Outline

# Relationship to univariate Gaussians

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

The term $(x-\mu)^2$ is a quadratic function. Similarly, $-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)$ is a quadratic form in the vector variable $x$ since $\Sigma$ is positive definite, i.e.

$$(x-\mu)^T\Sigma^{-1}(x-\mu) > 0$$

## The covariance matrix

Proposition: For any random vector $X$ with mean $\mu$ and covariance matrix $\Sigma$,

$$\Sigma = E\left[(X-\mu)(X-\mu)^T\right] = E\left[XX^T\right] - \mu\mu^T$$

# The diagonal covariance matrix case

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

. . .

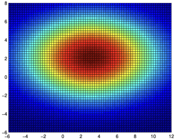
$$f(x;\mu,\Sigma) = \frac{1}{\sqrt{2\pi}\sigma_1} \mathsf{exp}\left(-\frac{1}{2\sigma_1^2}\left(x_1-\mu_1\right)^2\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \mathsf{exp}\left(-\frac{1}{2\sigma_2^2}\left(x_2-\mu_2\right)^2\right)$$

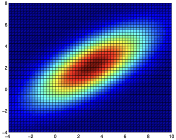
# Isocontours

Another way to understand a multivariate Gaussian conceptually is to understand the shape of its isocontours. For a function $f : \mathbf{R}^2 \to \mathbf{R}$, an isocontour is a set of the form

$$\left\{x \in \mathbf{R}^2 : f(x) = c\right\}.$$

for some $c \in \mathbf{R}^7$



# Linear transformation

Theorem: Let $X \sim \mathcal{N}(\mu, \Sigma)$ for some $\mu \in \mathbf{R}^d$ and $\Sigma \in \mathbf{S}^d_{++}$. Then, there exists a matrix $B \in \mathbf{R}^{d \times d}$ such that if we define $Z = B^{-1}(X - \mu)$, then $Z \sim \mathcal{N}(0, I)$

## Closure properties

A fancy feature of the multivariate Gaussian distribution is the following set of closure properties:

- The sum of independent Gaussian random variables is Gaussian.
- The marginal of a joint Gaussian distribution is Gaussian.
- The conditional of a joint Gaussian distribution is Gaussian.

Theorem 5.3.: Suppose that $y \sim \mathcal{N}(\mu, \Sigma)$ and $z \sim \mathcal{N}\left(\mu', \Sigma'\right)$ are independent Gaussian distributed random variables, where $\mu, \mu' \in \mathbf{R}^d$ and $\Sigma, \Sigma' \in \mathbf{S}^d_{++}$. Then, their sum is also Gaussian:

$$y + z \sim \mathcal{N}\left(\mu + \mu', \Sigma + \Sigma'\right).$$

## Closure properties

Suppose that

$$\begin{bmatrix} x_A \\ x_B \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \begin{bmatrix} \Sigma_{AA} & \Sigma_{AB} \\ \Sigma_{BA} & \Sigma_{BB} \end{bmatrix} \right)$$

where $x_A \in \mathbf{R}^n, x_B \in \mathbf{R}^d$, and the dimensions of the mean vectors and covariance matrix subblocks are chosen to match $x_A$ and $x_B$.

Theorem: Then, the marginal densities,

$$p(x_A) = \int_{x_B \in \mathbf{R}^d} p(x_A, x_B; \mu, \Sigma)\, dx_B$$
$$p(x_B) = \int_{x_A \in \mathbf{R}^n} p(x_A, x_B; \mu, \Sigma)\, dx_A$$

# Closure properties

are Gaussian:

$$x_A \sim \mathcal{N}(\mu_A, \Sigma_{AA})$$
$$x_B \sim \mathcal{N}(\mu_B, \Sigma_{BB}).$$

# Closure properties

Theorem: The conditional densities

$$p\left(x_A \mid x_B\right) = \frac{p\left(x_A, x_B; \mu, \Sigma\right)}{\int_{x_A \in \mathbf{R}^n} p\left(x_A, x_B; \mu, \Sigma\right) dx_A}$$

$$p\left(x_B \mid x_A\right) = \frac{p\left(x_A, x_B; \mu, \Sigma\right)}{\int_{x_B \in \mathbf{R}^d} p\left(x_A, x_B; \mu, \Sigma\right) dx_B}$$

are also Gaussian:

$$x_A \mid x_B \sim \mathcal{N}\left(\mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}\left(x_B - \mu_B\right), \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA}\right)$$

$$x_B \mid x_A \sim \mathcal{N}\left(\mu_B + \Sigma_{BA}\Sigma_{AA}^{-1}\left(x_A - \mu_A\right), \Sigma_{BB} - \Sigma_{BA}\Sigma_{AA}^{-1}\Sigma_{AB}\right)$$