02462 – Signals and data

Technical University of Denmark,
DTU Compute, Institut for Matematik og Computer Science.

Overview

1. Multivariate Normal Distributions

Multivariate Normal Distributions

Correlation

When we mix random variables, the result is often *correlated*,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ 2X_1 + 1X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

If X_1 changes, both Y_1 and Y_2 change!

Correlation

Correlation measures the influence two variables have on the other.

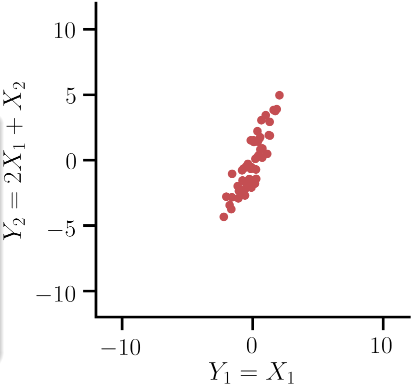
$$\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) / \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

where $\mathrm{Cov}(X, Y)$ is the *covariance*,

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

1. Correlation is in $[-1, 1]$.
2. If X and Y are independent, it is 0.
3. If $X = Y$ it is 1 and if $X = -Y$ it is -1.

Y_1 and Y_2 are *correlated*



Correlation

When we mix random variables, the result is often *correlated*,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ 2X_1 + 1X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

If X_1 changes, both Y_1 and Y_2 change!

Correlation

Correlation measures the influence two variables have on the other.

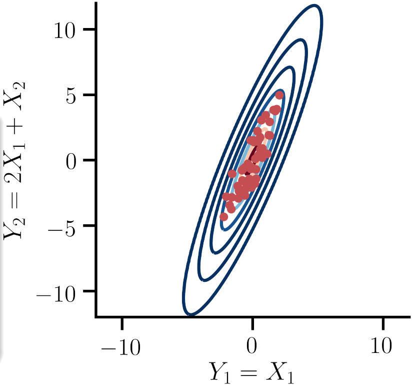
$$\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) / \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}.$$

where $\mathrm{Cov}(X, Y)$ is the *covariance*,

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

1. Correlation is in $[-1, 1]$.
2. If X and Y are independent, it is 0.
3. If $X = Y$ it is 1 and if $X = -Y$ it is -1.

Y_1 and Y_2 are *correlated*



we need to work with the *joint density* of Y_1 and Y_2.

Multivariate Normal Joint Density

If X_1 and X_2 are independent, their joint density is a *product*,

$$p(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

We can collect these factors using *vectors*,

$$\frac{1}{\sqrt{(2\pi)^2 \underbrace{\sigma_1^2\sigma_2^2}_{\det(\boldsymbol{\Sigma})}}} \exp\left(-\frac{1}{2}\Bigg(\underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\boldsymbol{x}} - \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}_{\boldsymbol{\mu}}\Bigg)^\top \underbrace{\begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}}_{\boldsymbol{\Sigma}^{-1}} \Bigg(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\Bigg)\right)$$

Multivariate Normal Distribution

A multivariate D-dimensional random variable X has a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if it has the joint probability density,

$$\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right)$$

with mean parameter $\boldsymbol{\mu} \in \mathbb{R}^D$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ a symmetric matrix with positive eigenvalues called the covariance matrix.

Multivariate Normal Joint Density

If X_1 and X_2 are independent, their joint density is a *product*,

$$p(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

We can collect these factors using *vectors*,

$$\frac{1}{\sqrt{(2\pi)^2 \underbrace{\sigma_1^2\sigma_2^2}_{\det(\boldsymbol{\Sigma})}}} \exp\left(-\frac{1}{2}\left(\underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_{\mathbf{x}} - \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}}_{\boldsymbol{\mu}}\right)^\top \underbrace{\begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}}_{\boldsymbol{\Sigma}^{-1}} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\right)\right)$$

Multivariate Normal Distribution

A multivariate D-dimensional random variable X has a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if it has the joint probability density,

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

with mean parameter $\boldsymbol{\mu} \in \mathbb{R}^D$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{D\times D}$ a symmetric matrix with positive eigenvalues called the covariance matrix.

The Covariance Matrix

Variables X_i have a variance, multivariate variables X have a *covariance matrix*,

$$\mathrm{Cov}(X) = \mathbb{E}\Big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\Big]. \qquad \text{(covariance matrix)}$$

where element (i, j) is the covariance between the i'th and the j'th element of X

$$\mathrm{Cov}(X)_{ij} = \mathrm{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])].$$

Remember that covariance is just *scaled correlation*,

$$\mathrm{Cov}(X_i, X_j) = \sqrt{\mathrm{Var}(X_i)\,\mathrm{Var}(X_j)}\,\mathrm{Corr}(X_i, X_j) \qquad (1)$$

In particular, note that the diagonal element $\mathrm{Cov}(X)_{ii} = \mathrm{Var}(X_i)$ is the *variance*!

- When $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{\Sigma} = \mathrm{Cov}(X)$.
- Before, X_1 and X_2 were independent, and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix},$$

If the dimensions of X are independent, $\mathrm{Cov}(X)$ is always *diagonal*.

Introducing Covariance

Mixing the elements of X with a matrix A introduces covariance,

$$\mathrm{Cov}(AX) = \mathbb{E}\left[A(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top A^\top\right] = A\,\mathrm{Cov}(X)A^\top \tag{2}$$

Example $\mathrm{Cov}(X) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is not covariant, but Y is,

$$Y = \begin{pmatrix} X_1 \\ 2X_1 + 1X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} X \Rightarrow \mathrm{Cov}(Y) = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$$

This is true no matter how X is distributed, but if X is normal then Y is also normal.

Linear Transforms of Multivariate Normal

For any linear transformation $Y = AX + b$ of any normal $X \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the result Y has distribution,

$$Y = AX + b \Rightarrow Y \sim \mathcal{N}(A\boldsymbol{\mu} + b, A\boldsymbol{\Sigma}A^\top) \tag{linearity}$$

Friendly Linear Algebra Reminder: Rotations

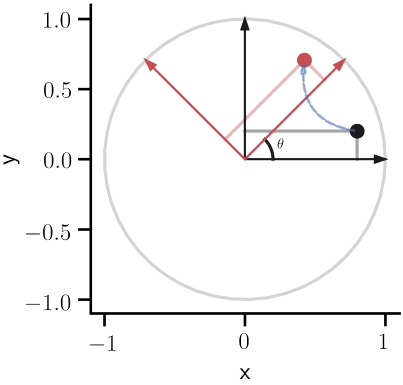
A rotation matrix is defined as,

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

A vector $\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2$ in the standard basis is rotated to a new basis,

$$R_\theta \boldsymbol{x} = x_1 \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} + x_2 \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}.$$

- $(\cos(\theta), \sin(\theta))$ is a unit vector at angle θ with the x-axis.
- The transpose rotates backwards, $R_\theta^\top R_\theta \boldsymbol{x} = \boldsymbol{x}$.
- In higher dimensions we have the *orthogonal matrices* U where $U^\top U = I$ and $\det(U) = 1$.

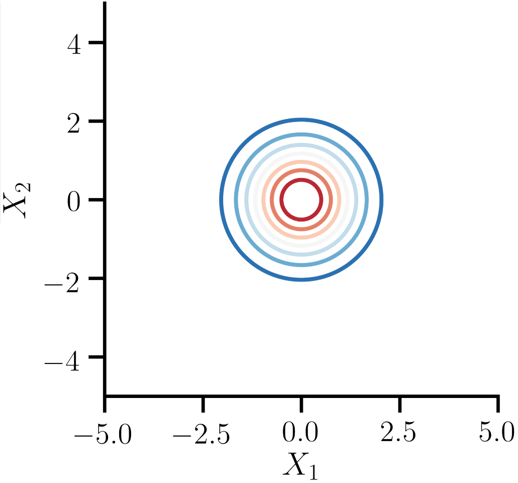


Multivariate Normals with Dependency

Standard Multivariate Normal

If we take the product of D independent univariate $X_d \sim \mathcal{N}(0, 1)$ normals we get the standard multivariate normal,

$$\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$



Multivariate Normals with Dependency

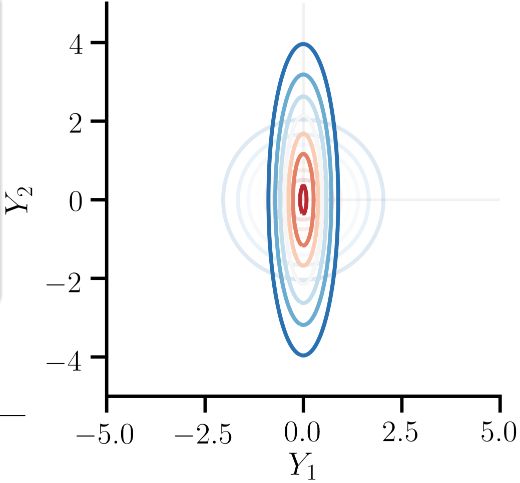
Scaled Diagonal Multivariate Normal

If $X \sim \mathcal{N}(\mathbf{0}, I)$ is scaled as $Y = SX$ by a diagonal matrix[1] S we get

$$X \sim \mathcal{N}(\mathbf{0}, S^2)$$

which is the product of independent normals $Y_d \sim \mathcal{N}(0, S_{dd}^2)$.

Note that the contour is only stretched along the two coordinate axes!



[1] When calculating, remember that the determinant of a diagonal matrix is the product of the diagonal elements!

Multivariate Normals with Dependency

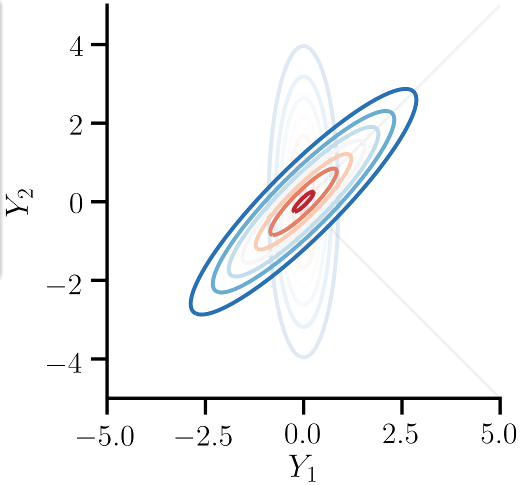
Full Multivariate Normal

If we first scale by S and then multiply by a *rotation matrix U* so that $Z = USx$ we get,

$$Z \sim \mathcal{N}(\mathbf{0}, US^2U^\top),$$

which is a product of independent Gaussians *in a different coordinate system*.

Instead of being stretched along the new coordinated axes, the contour is stretched along the rotated original axes.

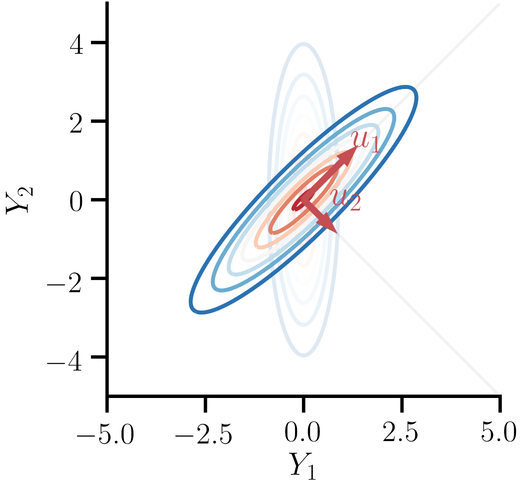


Multivariate Normals with Dependency

This should look familiar:

$$\boldsymbol{\Sigma} = US^2U^\top$$

- U and S^2 are the *eigenvectors* and *eigenvalues* of $\boldsymbol{\Sigma}$.
- $\boldsymbol{\Sigma}$ is symmetric and its eigenvectors form an *orthogonal basis*[2].
- If we rotate to the eigenbasis, the variables become independent.
- $\boldsymbol{\Sigma}$ has positive eigenvalues S^2 which makes it a *positive definite matrix*.



---

[2] as long as the eigenvalues are distinct.

Normal Marginals

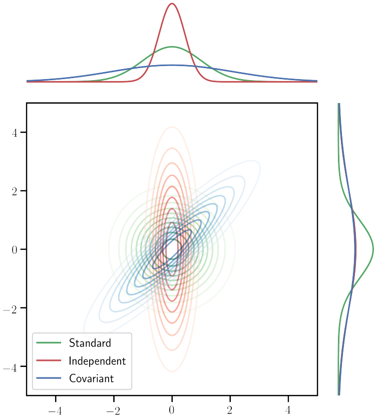
If $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

1. split the mean and covariance,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

2. *the marginal* of X_1 is then equal to

$$p(x_1) = \mathcal{N}(x_1; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}).$$



Normal Conditionals

If $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

1. split the mean and covariance,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

2. *the conditional* of X_1 given $X_2 = x_2$ is then equal to

$$p(x_1|x_2) = \mathcal{N}(x_1; \boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}),$$
$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(x_2 - \boldsymbol{\mu}_2),$$
$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}.$$