Machine learning - Probability; Multivariate

Yonghoon Dong

March 8, 2024

Covariance

The **covariance** between two random variables X and Y measures the degree to which X and Y are linearly related. It is defined as

$$\begin{aligned}\mathrm{Cov}[X, Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

Correlation coefficient (Pearson correlation coefficient)

The Pearson **correlation coefficient** between X and Y is defined as

$$\rho = \mathrm{Corr}[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}}$$

Equivalence class of random variables

We want to define a equivalence class as

$$[X] = [Y]$$

if there exists $b \in \mathbb{R}$ such that $X = Y + b$

Inner product space

Definition

An inner product space is a vector space V over the field F together with an inner product that satisfies the following three properties:

1. Symmetry : $\langle x, y\rangle = \langle y, x\rangle$
2. Linearity : $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$
3. Positive-definiteness : $\langle x, x\rangle > 0$, if $x \neq 0$

Inner product of random variables

Definition (Inner product of random variable)

Let $[X], [Y]$ are equivalence class of random variables. Then the inner product of $[X]$ and $[Y]$ can be defined as

$$\langle [X], [Y] \rangle = \mathrm{Cov}[X', Y']$$

where $\mathbb{E}[X'] = E[Y'] = 0$, $X' = X + b_1$ and $Y' = Y + b_2$

We have to check

1. Well-definedness of this function
2. Conditions of inner product

Well-definedness of inner product

Since every $[X], [Y]$, there exists X', Y' that satisfies the above conditions. Let X, X', Y, Y' such that $X' = X + b_1$ and $Y' = Y + b_2$. Then

$$\begin{aligned}\mathrm{Cov}[X', Y'] &= \mathbb{E}[(X' - \mathbb{E}[X'])(Y' - \mathbb{E}[Y'])] \\ &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathrm{Cov}[X, Y]\end{aligned}$$

Therefore, it is a well-defined function.

Conditions of inner product

1. Symmetry : Trivial
2. Linearity :

$$
\begin{aligned}
\langle aX + bY, Z\rangle &= \mathrm{Cov}[aX' + bY', Z'] \\
&= a\mathrm{Cov}[X', Z'] + b\mathrm{Cov}[Y', Z'] \\
&= a\langle X, Z\rangle + b\langle Y, Z\rangle
\end{aligned}
$$

3. Positive-definiteness : For all X, Cov[X', X'] $= \mathbb{V}[X'] \geq 0$. Since $\mathbb{E}[X'] = 0$,

$$\mathbb{V}[X'] = 0 \Leftrightarrow X' = 0$$

   Therefore, $\langle [X], [X]\rangle = 0 \Leftrightarrow [X] = 0$

Therefore, it satisfies the all conditions of inner product

Revisit the definitions of covariance and correlation

We can rewrite the definition of covariance and correlation differently

$$\mathrm{Cov}[X, Y] = \langle [X], [Y] \rangle$$

Similarly,

$$\begin{aligned}
\mathrm{Corr}[X, Y] &= \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}} \\
&= \frac{\langle [X], [Y] \rangle}{\sqrt{\langle [X], [X] \rangle}\sqrt{\langle [Y], [Y] \rangle}} = \frac{\langle [X], [Y] \rangle}{\|[X]\| \cdot \|[Y]\|} \\
&= \cos\theta
\end{aligned}$$

since $\mathbb{V}[X'] = \mathbb{V}[X]$ (i.e. $\langle [X], [X] \rangle = \mathrm{Cov}[X', X'] = \mathrm{Cov}[X, X] = \mathbb{V}[X]$)

Revisit the definitions of covariance and correlation

Therefore, if we **ignore** the effect of translation, we can understand covariance and correlation as an **inner product** and the **angle** between random variables respectively. This interpretation is highly intuitive, as what we seek to understand through covariance and correlation is the relationship between two random variables.

Uncorrelated does not imply independent

If X and Y are independent, then $\text{Cov}[X, Y] = 0$, and hence $\text{Corr}[X, Y] = 0$.
However, the converse it not true: uncorrelated **does not** imply independent.

Multivariate Gaussian distribution

Definition (Multivariate Gaussian distribution)

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

where $\mu = \mathbb{E}[y] \in \mathbb{R}^D, \Sigma = \text{Cov}[y] \in M^{D\times D}$

Therefore,

$$\mathbb{E}[yy^T] = \Sigma + \mu\mu^2$$

Bivariate Gaussian distribution

If $D = 2$, it is called **bivariate Gaussian** distribution. Its pdf can be represented as $y \sim (\mu, \Sigma)$, where $y \in \mathbb{R}^2, \mu \in \mathbb{R}^2$ and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

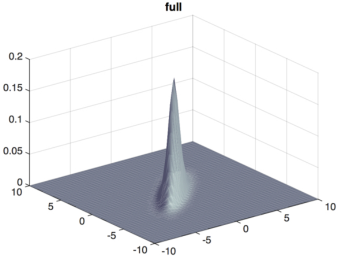
where ρ is the correlation coefficient, defined by $\text{corr}[Y_1, Y_2]$.

Bivariate Gaussian distribution

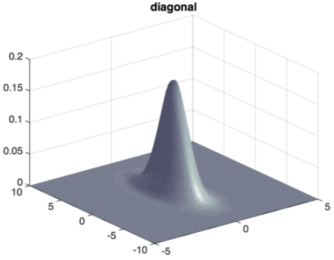
Expanding out the pdf gives the following equations

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(y_1-\mu_1)^2}{\sigma_1^2}+\frac{(y_2-\mu_2)^2}{\sigma_2^2}-2\rho\frac{(y_1-\mu_1)}{\sigma_1}\frac{(y_2-\mu_2)}{\sigma_2}\right)\right]$$

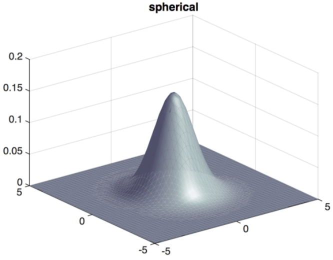
Bivariate Guassian distribution



(a)



(b)



(c)

Figure: Visualization of a 2d Gaussian density as a surface plot

Bivariate Guassian distribution

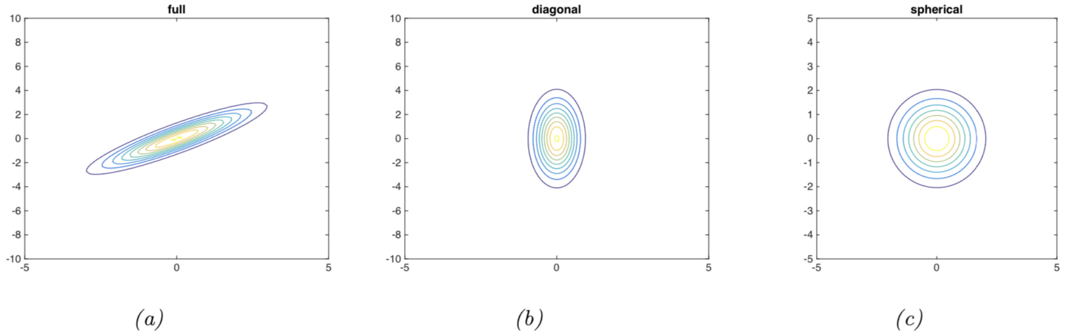


Figure: Visualization of a 2d Gaussian density in terms of level sets of constant probability density