Serie de Machine Learning
Revisión Gaussianas Multivariadas (en progreso)

SERGIO A. PERNICE[1]

Universidad del CEMA
Av. Córdoba 374, Buenos Aires, 1054, Argentina

7 de Julio de 2020

Abstract

En este documento presentamos una revision de distribuciones Gaussianas Multivariadas de una forma especialmente adaptada para sus eventuales aplicaciones en aprendizaje automático (machine learning). Es el ?? de una serie de documentos sobre machine learning en español. Es parte del contenido del curso "Métodos de Machine Learning para Economistas" de la Maestría en Economía de la UCEMA.

Keywords: Gaussianas Multivariadas, machine learning, aprendizaje automático.

Contents

[1]sp@ucema.edu.ar

1 Introducción

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \tag{1.1}$$

$$\int_{-\infty}^{\infty} e^{-x^2/(2\sigma^2)} dx = \sqrt{2}\sigma \int_{-\infty}^{\infty} e^{-\left(x/(\sqrt{2}\sigma)\right)^2} \frac{dx}{\sqrt{2}\sigma} = \sqrt{2}\sigma \int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{2\pi}\sigma \tag{1.2}$$

$$y = \frac{x}{\sqrt{2}\sigma}$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-x^2/(2\sigma^2)} dx = 1 \tag{1.3}$$

$$\frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} x \frac{1}{a} x\right) dx = 1, \quad a = \sigma^2 > 0 \tag{1.4}$$

$$\left[\frac{1}{\sqrt{2\pi}\sqrt{a_1}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} x_1 \frac{1}{a_1} x_1\right) dx_1\right]\left[\frac{1}{\sqrt{2\pi}\sqrt{a_2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} x_2 \frac{1}{a_2} x_2\right) dx_2\right] = 1 \tag{1.5}$$

$$\frac{1}{\left(\sqrt{2\pi}\right)^2 \sqrt{a_1 a_2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left[x_1 \frac{1}{a_1} x_1 + x_2 \frac{1}{a_2} x_2\right]\right) dx_1 dx_2 = 1 \tag{1.6}$$

$$\frac{1}{(2\pi)^{2/2}} \frac{1}{\det(\mathbf{\Sigma})^{\frac{1}{2}}} \int \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x}\right) d^2x \tag{1.7}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}, \quad \mathbf{\Sigma}^{-1} = \begin{pmatrix} 1/a_1 & 0 \\ 0 & 1/a_2 \end{pmatrix} \tag{1.8}$$

Todo esto extiende trivialmente a n variables:

$$\prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sqrt{a_i}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} x_i \frac{1}{a_i} x_i\right) dx_i\right] = \frac{1}{(2\pi)^{n/2}} \frac{1}{\det(\mathbf{\Sigma})^{\frac{1}{2}}} \int \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x}\right) d^n x \tag{1.9}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{\Sigma} = \mathrm{diag}(a_1, \ldots, a_n), \quad \mathbf{\Sigma}^{-1} = \mathrm{diag}(1/a_1, \ldots, 1/a_n) \tag{1.10}$$

en el exponente de una distribución normal hay una forma cuadrática diagonal (por ahora) con autovalores positivos (sino la integral no converge).

Volvamos a (1.3) y calculemos el valor esperado

$$\mathrm{E}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-x^2/(2a)} dx = 0 \tag{1.11}$$

por simetría (o por cálculo fuerza bruta). Para calcular la varianza y momentos superiores:

$$\mathrm{Var}(x) = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} x^2 e^{-x^2/(2a)} dx \tag{1.12}$$

hagamos el siguiente truco. Calculamos la integral:

$$F(j) = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{\infty} e^{-x^2/(2a)+xj} dx \tag{1.13}$$

Notemos que $F(j = 0) = 1$. Como

$$\text{exponente } = -\frac{x^2}{2a} + xj = -\frac{1}{2a}\left(x^2 - 2xaj + (aj)^2 - (aj)^2\right) = -\frac{1}{2a}(x - aj)^2 + \frac{1}{2}aj^2$$

haciendo el cambio de variable $y = x - aj$, integramos en (1.13) en y, obtenemos

$$F(j) = \exp\left(\frac{1}{2}aj^2\right) \tag{1.14}$$

y de (1.13) que

$$\mathrm{E}(x) = \left.\frac{dF(j)}{dj}\right|_{j=0} = ajF(j)|_{j=0} = 0 \tag{1.15}$$

$$\mathrm{Var}(x) = \left.\frac{d^2F(j)}{dj^2}\right|_{j=0} = \left.\frac{d\,(ajF(j))}{dj}\right|_{j=0} = \left.\left(aF(j) + (aj)^2\right)\right|_{j=0} = a = \sigma^2 \tag{1.16}$$

Si, en vez de la distribución (1.9), tenemos

$$\frac{1}{(2\pi)^{n/2}} \frac{1}{\det(\boldsymbol{\Sigma})^{\frac{1}{2}}} \int \exp\left(-\frac{1}{2}\left(\mathbf{x}^T - \boldsymbol{\mu}\right)\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) d^n x \tag{1.17}$$

haciendo el cambio de variables $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}$ volvemos a (1.9). Por lo que vemos que una distribución como el integrando de (1.17) es (1.9) trasladada en la manifold de las variables aleatorias. $\mathrm{E}(\mathbf{x}) = \boldsymbol{\mu}$, todo lo demás sigue igual.

Finalmente, si $\boldsymbol{\Sigma}$ no fuese diagonal, de todos modos sería simétrica, ya que $\boldsymbol{\Sigma}^{-1}$ aparece en el exponente de (1.9) dentro de una forma cuadrática. Esto implica que existe una base de autovectores de $\boldsymbol{\Sigma}$ (que también son autovectores de $\boldsymbol{\Sigma}^{-1}$) en la que $\boldsymbol{\Sigma}$ es diagonal y se reduce a (1.9).

Las variables correspondientes a los autovectores se distribuyen de manera normal, y son independiente de las otras variables, por lo que cada una de ellas tiene una distribución de la forma (1.3) con varianza igual al correspondiente autovalor de $\boldsymbol{\Sigma}$. En el fondo, entendiendo bien la distribución (1.3) y algebra lineal, todo lo demás relativo a distribuciones normales multivariadas se deriva de ellas.