机器学习–数学部分

RankFan *[1] and JIE †[2]

[1] 公众号：经济知识综合
[2] 公众号：经济知识综合

2021 年 2 月 28 日

*fanxiaolong98@gmail.com
†fanxiaolong98@gmail.com

目录

1 多维高斯分布

$\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{x}$ 是一个随机变量。

$$\boldsymbol{x} \sim N(\mu, \Omega)$$

$$p(\boldsymbol{x}) = \frac{1}{2\pi^{\frac{p}{2}}\,|\Omega|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Omega^{-1}}\,(\boldsymbol{x}-\boldsymbol{\mu})\}$$

$\boldsymbol{\Omega}$ 是正定阵，一般是半正定阵. 通常而言是半正定的对称阵。

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

$(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Omega^{-1}}\,(\boldsymbol{x}-\boldsymbol{\mu})$ 可以看做是一个关于 $\boldsymbol{x}, \boldsymbol{\mu}$ 之间的一个马氏距离。$\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ 为精度矩阵。

1.1 马氏距离

Example 1.1 (马氏距离的定义)**.**

$$\boldsymbol{z_1} = \begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix} \quad \boldsymbol{z_2} = \begin{pmatrix} z_{21} \\ z_{22} \end{pmatrix}$$

$$(\boldsymbol{z_1}-\boldsymbol{z_2})^T\Omega^{-1}(\boldsymbol{z_1}-\boldsymbol{z_2}) = (z_{11}-z_{21}, z_{12}-z_{22})\,\boldsymbol{\Omega^{-1}}\begin{pmatrix} z_{11}-z_{21} \\ z_{12}-z_{22} \end{pmatrix}$$

马氏距离与欧式距离的联系：$\boldsymbol{\Omega} = \boldsymbol{I}$ 时，马氏距离 = 欧式距离。

Definition 1.1 (方差矩阵)**.**

$$\boldsymbol{\Omega} = \boldsymbol{U\Lambda U^T}$$

$\boldsymbol{\Omega}$ 是对称正定阵，可进行矩阵分解。$diag(\lambda_i)$ 是对角矩阵。$\boldsymbol{\Omega}$ 可以对称正交化。

$$\begin{aligned}
\boldsymbol{UU^T} &= \boldsymbol{U^TU} = \boldsymbol{I} \\
\boldsymbol{\Lambda} &= diag(\lambda_i) \quad i = 1, 2, \ldots, p \\
\boldsymbol{U} &= (\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_p}); \quad \boldsymbol{u_i}\ \ is\ \ a\ \ column\ \ vector(p\times 1), \|\boldsymbol{u_i}\| = 1 \\
\implies \qquad \boldsymbol{\Omega} &= (\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_p})\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix}\begin{pmatrix} \boldsymbol{u_1^T} \\ \boldsymbol{u_2^T} \\ \vdots \\ \boldsymbol{u_p^T} \end{pmatrix} \\
&= (\boldsymbol{u_1}\lambda_1, \boldsymbol{u_2}\lambda_2, \cdots, \boldsymbol{u_p}\lambda_p)\begin{pmatrix} \boldsymbol{u_1^T} \\ \boldsymbol{u_2^T} \\ \vdots \\ \boldsymbol{u_p^T} \end{pmatrix} = \sum_{i=1}^{p} \boldsymbol{u_i}\lambda_i\boldsymbol{u_i^T}
\end{aligned}$$

$$\Longrightarrow \qquad \boldsymbol{\Omega^{-1}} = \left(\boldsymbol{U\Lambda U^T}\right)^{-1} = \boldsymbol{U\Lambda^{-1}U^T}$$
$$\boldsymbol{\Lambda^{-1}} = \ diag(\lambda_i^{-1}) \quad i = 1, 2, \ldots, p$$
$$\boldsymbol{\Omega^{-1}} = \sum_{i=1}^{p} \boldsymbol{u_i}\lambda_i^{-1}\boldsymbol{u_i^T}$$
$$\Longrightarrow \qquad (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Omega^{-1}}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}-\boldsymbol{\mu})^T\sum_{i=1}^{p}\boldsymbol{u_i}\lambda_i^{-1}\boldsymbol{u_i^T}(\boldsymbol{x}-\boldsymbol{\mu})$$
$$= \sum_{i=1}^{p}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{u_i}\lambda_i^{-1}\boldsymbol{u_i^T}(\boldsymbol{x}-\boldsymbol{\mu})$$

令 $y_i = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{u_i}$, y_i 是一个标量，几何意义是 $(\boldsymbol{x}-\boldsymbol{\mu})^T$ 在 $\boldsymbol{u_i}$ 上的投影。$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$

$$\Longrightarrow \qquad (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Omega^{-1}}(\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{i=1}^{p} y_i{\lambda_i}^{-1}y_i = \sum_{i=1}^{p} y_i^2{\lambda_i}^{-1}$$

令 $p = 2$，即二维。(1.1) 几何意义是二维平面中的一个椭圆

$$\therefore \qquad (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Omega^{-1}}(\boldsymbol{x}-\boldsymbol{\mu}) = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} = constant \tag{1.1}$$

假设 $\lambda_1 > \lambda_2$，椭圆的长半径为 $\lambda_1/2$ ，短半径为 $\lambda_2/2$。

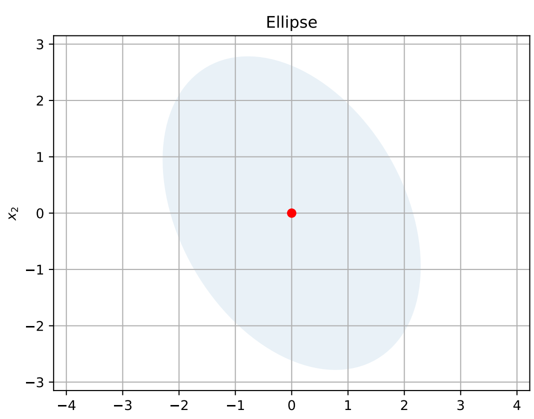

图 1.1: p = 2 时马氏距离的几何意义

如果是 $p = 3$，即三维。其几何意义就是切不同的等高线。

$$\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \frac{y_3^2}{\lambda_3} = constant$$

二维高斯分布经常用椭圆等高线表达。

Example 1.2 (问题)**.** 什么样的矩阵可以写成这种形式？

$$\boldsymbol{\Omega} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\boldsymbol{T}}$$

总结：

$$Summary: \quad \Delta = (\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Omega}^{\boldsymbol{-1}} (\boldsymbol{x}-\boldsymbol{\mu}) = \sum_{i=1}^{p} y_i^2 {\lambda_i}^{-1}$$

$$p(\boldsymbol{x}) = \frac{1}{2\pi^{\frac{p}{2}} |\Omega|^{\frac{1}{2}}} exp\{\Delta\}$$

其中 $y_i = (\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{u_i}$， $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}$

MATLAB Code

```
Euler method for the ODE model
u'(x)=x^2+x-u, x in [0,1]
Initial condition: u(0)=0 ;
Exact solution: u(x)=-exp(-x)+x^2-x+1.
clear all;  clf
```

Python Code

Python Code

```
#PythonDraw.py
import turtle as t
```