智能软件开发
方向基础

第七章 朴素贝叶斯分类
Naive Bayes Classifier
PART2-正态分布的概率密度函数

张朝晖

2022~2023学年第二学期

河北师范大学软件学院
Software College of Hebei Normal University

序号	内容
1	概述
2	机器学习的基本概念
3	模型的选择与性能评价
4	数据的获取、探索与准备
5	近邻模型------分类、回归
6	决策树模型------分类、回归
7	集成学习------分类、回归
8	(朴素)贝叶斯模型------分类
9	聚类
10	特征降维及低维可视化(PCA, t-SNE)
11	总复习

主要内容

1. 引言

2. 贝叶斯决策模型

2.1 最小错误率贝叶斯决策

连续情况；离散情况

2.2 最小风险的贝叶斯决策

连续情况；离散情况

3. 正态分布的概率密度函数
——“类条件概率密度函数为正态分布”

4. 朴素贝叶斯(Naive Bayes)分类实现流程

问题的引入：

- 贝叶斯规则中，涉及“连续随机变量或向量的类条件概率密度函数”。
- “正态分布”的概率密度函数，特点：

  物理上，合理性；

  数学上，比较简便。

➔ 简单、符合一些实际情况

主要内容

3.1. 单个随机变量的正态分布

3.2.随机向量的正态分布

连续随机变量X概率密度函数定义

$$p_X(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

记 $p_X(x) \sim N\left(\mu,\sigma^2\right)$

标准正态分布：$p_X(x) \sim N\left(0,1\right)$

其中

- $p_X(x)$ －随机变量(标量)在x处的密度值
- μ－随机变量X的期望，或均值

$$\mu \equiv E\left\{X\right\}=\int_{-\infty}^{\infty} x p_X(x)dx$$

- σ－x的标准差，描述x的分散程度
- σ^2－x的方差：$\sigma^2 \equiv E\left\{\left(X-\mu\right)^2\right\}=\int_{-\infty}^{\infty}\left(x-\mu\right)^2 p_X(x)dx$

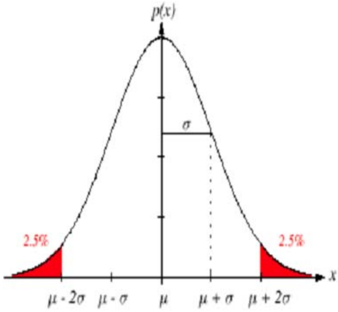
性质： $p_X(x) \geq 0, -\infty < x < +\infty$

$$\int_{-\infty}^{\infty} p_X(x)dx = 1$$

$$\int_{\mu-\sigma}^{\mu+\sigma} p_X(x)dx = 0.683$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} p_X(x)dx = 0.9544$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} p_X(x)dx = 0.9974$$



图：单变量正态分布。

其它：熵 $\boldsymbol{H}(p_X(x)) = -\int p_X(x) \ln p_X(x)dx > 0$

***Mahalanobis*距离** $\boldsymbol{r} = \dfrac{|\boldsymbol{x} - \mu|}{\sigma}$

$$\boldsymbol{p}(X = \mu) = \frac{1}{\sqrt{2\pi}\sigma}$$

$$\boldsymbol{E}\{f(\boldsymbol{X})\} = \int_{-\infty}^{\infty} f(x) p_X(x)dx$$

主要内容

3.1. 单个随机变量的正态分布

3.2.随机向量的正态分布

河北师范大学软件学院
Software College of Hebei Normal University

[1]多元正态分布的概率密度函数 $--p_X(x)$

定义： $$p_X(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

记为 $p_X(x)\sim \boldsymbol{N}(\boldsymbol{\mu},\Sigma)$

其中：

- $\boldsymbol{X}$ $-d$维列向量，$\boldsymbol{X}=[X_1,X_2,\dots,X_d]^T$
  X_i为随机变量，$i=1,2,\dots,\boldsymbol{d}$;
- $\boldsymbol{\mu}$ $-d$维均值向量，$\boldsymbol{\mu}=[\mu_1,\mu_2,\dots,\mu_d]^T$
  μ_i为第i维随机变量X_i的均值，$i=1,2,\dots,\boldsymbol{d}$;
- $\Sigma-d$阶**协方差矩阵**(*covariance matrix*)
- $\Sigma^{-1}-$矩阵Σ的逆矩阵，**精度矩阵**
- $|\Sigma|-$矩阵Σ的行列式

[1]多元正态分布概率密度函数定义(续) $--\boldsymbol{\mu}$

$\boldsymbol{\mu}$ $-d$维均值向量，$\boldsymbol{\mu}=[\mu_1,\mu_2,\dots,\mu_d]^T$

$$\boldsymbol{\mu}\equiv \boldsymbol{E}\{\boldsymbol{X}\}=\int_{\Re^d}\boldsymbol{x}p_X(x)d\boldsymbol{x}$$

$$=\left[\boldsymbol{E}\{x_1\},\boldsymbol{E}\{x_2\},\dots,\boldsymbol{E}\{x_d\}\right]^T=[\mu_1,\mu_2,\dots,\mu_d]^T$$

其中：

$$\mu_i=\boldsymbol{E}\{X_i\}=\int_{\Re^d}x_i p_X(x)d\boldsymbol{x}$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}x_i p_X(x)dx_1dx_2\dots dx_d=\int_{-\infty}^{+\infty}x_i\, p(x_i)dx_i$$

边缘密度函数 $$p_{X_i}(x_i)=\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\dots\int_{-\infty}^{+\infty}p_X(x)dx_1\dots dx_{i-1}dx_{i+1}\dots dx_d$$

[1]多元正态分布概率密度函数定义(续)--Σ

$\Sigma - d\times d$维协方差矩阵(*covariance matrix*):

$$\Sigma \equiv \boldsymbol{E}\left\{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right\} = \int_{\Re^d} (\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{p}_X(\boldsymbol{x})d\boldsymbol{x}$$

$$\Sigma = \boldsymbol{E}\left\{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right\} = \boldsymbol{E}\left\{\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_d-\mu_d \end{bmatrix}\begin{bmatrix} X_1-\mu_1 & X_2-\mu_2 & \cdots & X_d-\mu_d \end{bmatrix}\right\}$$

$$= \boldsymbol{E}\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_d-\mu_d) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \cdots & (X_2-\mu_2)(X_d-\mu_d) \\ \vdots & \vdots & \vdots & \vdots \\ (X_d-\mu_d)(X_1-\mu_1) & (X_d-\mu_d)(X_2-\mu_2) & \cdots & (X_d-\mu_d)^2 \end{bmatrix}$$

=(下页待续)

$$\Sigma \equiv \boldsymbol{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right]$$

$$= \begin{bmatrix} \boldsymbol{E}\left\{(X_1-\mu_1)^2\right\} & \boldsymbol{E}\left\{(X_1-\mu_1)(X_2-\mu_2)\right\} & \cdots & \boldsymbol{E}\left\{(X_1-\mu_1)(X_d-\mu_d)\right\} \\ \boldsymbol{E}\left\{(X_2-\mu_2)(X_1-\mu_1)\right\} & \boldsymbol{E}\left\{(X_2-\mu_2)^2\right\} & \cdots & \boldsymbol{E}\left\{(X_2-\mu_2)(X_d-\mu_d)\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{E}\left\{(X_d-\mu_d)(X_1-\mu_1)\right\} & \boldsymbol{E}\left\{(X_d-\mu_d)(X_2-\mu_2)\right\} & \cdots & \boldsymbol{E}\left\{(X_d-\mu_d)^2\right\} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2d} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \sigma_{d1} & \sigma_{d2} & \cdots\cdots & \sigma_{dd} \end{bmatrix} \qquad \Sigma\text{对称非负定}\begin{cases} \text{对角线元素}\sigma_{ii} = \sigma_i^2, \quad X_i\text{方差} \\ \text{非对角元素}\sigma_{ij}, \ X_i、X_j\text{协方差} \end{cases}$$

这里:只考虑**Σ对称正定**

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1\mathbf{d}} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2\boldsymbol{d}} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \sigma_{\boldsymbol{d}1} & \sigma_{\boldsymbol{d}2} & \cdots\cdots & \sigma_{\boldsymbol{dd}} \end{bmatrix}$$

$$\sigma_{\mathbf{ij}} = \boldsymbol{E}\left\{(X_i-\mu_i)(X_j-\mu_j)\right\} = \int_{\Re^d} (x_i-\mu_i)(x_j-\mu_j)\, p_X(\boldsymbol{x})d\boldsymbol{x}$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(x_i-\mu_i)(x_j-\mu_j)\,p_X(\boldsymbol{x})dx_1dx_2\ldots dx_d$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x_i-\mu_i)(x_j-\mu_j)\,p_{X_iX_j}(x_i,x_j)dx_idx_j$$

其中，边缘密度：

$$p_{X_iX_j}(x_i,x_j) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} p_X(\boldsymbol{x})dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_{j-1}dx_{j+1}\ldots dx_d$$

[2]多元正态分布概率密度函数性质

性质1 多元正态分布完全由μ、Σ决定

参数共计$\boldsymbol{d}+\dfrac{\boldsymbol{d}(\boldsymbol{d}+1)}{2}$个

$$\boldsymbol{\mu} = [\mu_1,\mu_2,\ldots,\mu_d]^{\boldsymbol{T}}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1\mathbf{d}} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2\boldsymbol{d}} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \sigma_{\boldsymbol{d}1} & \sigma_{\boldsymbol{d}2} & \cdots\cdots & \sigma_{\boldsymbol{dd}} \end{bmatrix}$$为对称矩阵

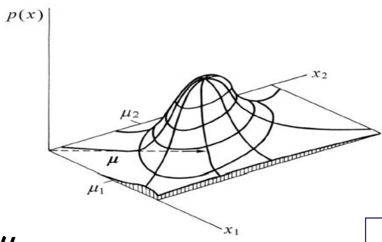
$$p_X(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\boldsymbol{T}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

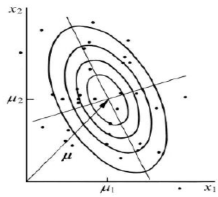
性质2 等概率密度点的轨迹为一超椭球面

$$p_X(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

x到μ的 ***Mahalanobis*距离**平方 $\quad r^2 = (x-\mu)^T \Sigma^{-1}(x-\mu)$

等概率密度点：$r^2 =$ 常数





超椭球面中心μ

超椭球面主轴 $\begin{cases}\text{方向取决于}\Sigma\text{本征向量} \\ \text{长度与}\Sigma\text{本征值的平方根成正比。}\end{cases}$

二维正态分布，等概率密度点的轨迹为椭圆周。

性质3 不相关性等价于独立性。

x多元正态分布 $\begin{cases}\text{"}X\text{任意两分量}X_i, X_j\text{间互不相关"} \\ \qquad \Leftrightarrow \text{"各分量间相互独立"} \\ \text{协方差矩阵}\Sigma\text{是对角矩阵} \Rightarrow X\text{各分量相互独立}\end{cases}$

定义：随机变量X_i, X_j之间

$$\begin{cases} X_i, X_j\text{间不相关：} & E\{X_i X_j\} = E\{X_i\} E\{X_j\} \\ X_i, X_j\text{相互独立：} & p_{X_i X_j}(x_i, x_j) = p_{X_i}(x_i) p_{X_j}(x_j) \end{cases}$$

并且

$$p_{X_i X_j}(x_i, x_j) = p_{X_i}(x_i) p_{X_j}(x_j) \Rightarrow E\{X_i X_j\} = E\{X_i\} E\{X_j\}$$

性质3 不相关性等价于独立性。

多元正态分布的任意随机变量间不相关性，等价于独立性。

$\forall X_i, X_j,\ i \neq j,\quad i,j = 1,\dots,d$ 若 X_i, X_j 间互不相关，则

$$\sigma_{ij} = E\left\{(X_i - \mu_i)(X_j - \mu_j)\right\} = E\left\{X_i X_j - X_i\mu_j - X_j\mu_i + \mu_i\mu_j\right\}$$

$$= E\{X_i\}E\{X_j\} - \mu_j E\{X_i\} - \mu_i E\{X_j\} + \mu_i\mu_j = 0$$

所以：$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2d} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ \sigma_{d1} & \sigma_{d2} & \cdots\cdots & \sigma_{dd} \end{bmatrix} = \begin{bmatrix} \sigma_{11} & & & \\ & \sigma_{22} & & \\ & & \cdots\cdots & \\ & & & \sigma_{dd} \end{bmatrix}$

因此：若 X_i, X_j 间互不相关，则协方差矩阵 Σ 是对角的

性质3 不相关性等价于独立性。

注意：$\sigma_{ii} = \sigma_i^2$

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_{11} & & & \\ & 1/\sigma_{22} & & \\ & & \cdots\cdots & \\ & & & 1/\sigma_{dd} \end{bmatrix}$$

$$|\Sigma| = \prod_{i=1}^{d} \sigma_{ii} \qquad |\Sigma|^{\frac{1}{2}} = \prod_{i=1}^{d} \sqrt{\sigma_{ii}} = \prod_{i=1}^{d} \sigma_i$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu) = [x_1-\mu_1, \dots, x_d-\mu_d] \begin{pmatrix} 1/\sigma_{11} & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_{dd} \end{pmatrix} \begin{bmatrix} x_1-\mu_1 \\ \vdots \\ x_d-\mu_d \end{bmatrix}$$

$$= \left[(x_1-\mu_1)/\sigma_{11} \quad \cdots \quad (x_d-\mu_d)/\sigma_{dd}\right] \begin{bmatrix} x_1-\mu_1 \\ \vdots \\ x_d-\mu_d \end{bmatrix} = \sum_{i=1}^{d} \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})=\sum_{i=1}^{d}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

$$p_X(\boldsymbol{x})=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

$$=\frac{1}{(2\pi)^{d/2}\prod_{i=1}^{d}\sigma_i}\exp\left[-\frac{1}{2}\sum_{i=1}^{d}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]$$

$$=\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right]=\prod_{i=1}^{d}p_{X_i}(x_i)$$

所以对于多元正态分布$\boldsymbol{X}$，

$$\begin{cases}\text{各分量}X_i,X_j\text{间互不相关}\Leftrightarrow\text{各分量相互独立。}\\ \text{协方差矩阵}\Sigma\text{是对角的}\Rightarrow \boldsymbol{X}\text{各分量相互独立，且正态分布}\end{cases}$$

性质4 边缘密度分布与条件密度分布仍为正态分布。

边缘概率密度：$p_{X_i}(x_i)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p_X(\boldsymbol{x})dx_1\ldots dx_{i\text{-}1}dx_{i+1}\ldots dx_d$

$$p_{X_i}(x_i)=\frac{1}{\sqrt{2\pi}\sigma_{ii}}\exp\left[-\frac{1}{2}\left(\frac{\boldsymbol{x}_i-\mu_i}{\sigma_i}\right)^2\right]$$

$\boldsymbol{i}=1,\ldots,\boldsymbol{d}$

条件概率密度：给定$\boldsymbol{X}_i=x_i$条件下，X_j的分布。

$$p_{X_i|X_j}(x_j\mid x_i)=\frac{p_{X_iX_j}(x_i,x_j)}{p_{X_i}(x_i)}$$

其中：

$$\begin{cases}p_{X_iX_j}(x_i,x_j)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p_X(\boldsymbol{x})dx_1\ldots dx_{i\text{-}1}dx_{i+1}\ldots dx_{j\text{-}1}dx_{j+1}\ldots dx_d\\ p_{X_i}(x_i)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}p_X(\boldsymbol{x})dx_1\ldots dx_{i\text{-}1}dx_{i+1}\ldots dx_d\end{cases}$$

性质5　**线性变换的正态性**

正态分布的随机向量X经线性变换后，仍为正态分布。

若 $\begin{cases} \text{随机向量} \boldsymbol{X} = \left[X_1, \ldots, X_d\right]^T, \boldsymbol{p_X}(\boldsymbol{x}) \sim N(\boldsymbol{\mu}, \Sigma) \\ \textbf{变换矩阵} \boldsymbol{A} = \left[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_k\right] \quad d\text{行} \times k\text{列}; \\ \text{线性变换} \quad \boldsymbol{Y} = \boldsymbol{A}^T \boldsymbol{X} \end{cases}$

则 $\boldsymbol{p_Y}(\boldsymbol{y}) \sim \boldsymbol{N}\left(\boldsymbol{A}^T \boldsymbol{\mu}, \boldsymbol{A}^T \Sigma \boldsymbol{A}\right)$

性质6 线性组合的正态性

正态分布的X各分量的线性组合仍为正态分布。

若：随机向量 $\boldsymbol{X} = \left[x_1, \ldots, x_d\right]^T, p_X(\boldsymbol{x}) \sim \boldsymbol{N}(\boldsymbol{\mu}, \Sigma)$

d维向量 $\boldsymbol{a} = \left[a_1, \ldots, a_d\right]^T$;

随机变量 $Y = \boldsymbol{a}^T \boldsymbol{X}$

则：$\boldsymbol{p_Y}(y) \sim \boldsymbol{N}\left(\mu_y, \sigma_y^2\right) = \boldsymbol{N}\left(\boldsymbol{a}^T \boldsymbol{\mu}, \boldsymbol{a}^T \Sigma \boldsymbol{a}\right)$

$\Rightarrow$ 通过线性组合可获知观测数据在a方向的整体分散程度。

多元正态分布特征空间的线性变换