投飞镖

- 事件：只有中靶和不中靶两类
  - 两种可能： Bernoulli Random Variable
  - 投一次结果：Bernoulli Distribution
  - 投N次结果： Binomial Distribution
- 事件：投中情况下，按中靶区域细分为K类
  - K种可能： Categorical Random Variable
  - 投一次结果：Categorical Distribution
  - 投N次结果： Multinomial Distribution

多项分布

- K个区域，投中(没投中)第k个区域定义为：$x_k = 1(0)$
- 投中第k个区域的概率定义为：$p(x_k = 1) = \mu_k$
- 利用向量$\boldsymbol{x} = (x_1, x_2, ..., x_K)^\top$代表某一次投飞镖的结果
- 投N次后在第k个区域投中m_k次的概率即为多项分布：

$$p(\boldsymbol{m}|\boldsymbol{\mu}, N) = \binom{N}{\boldsymbol{m}} \prod_{k=1}^{K} \mu_k^{m_k}$$

- 多项式展开：

$$\left(\sum_{k=1}^{K} \mu_k\right)^N \equiv \sum_{\boldsymbol{m}} p(\boldsymbol{m}|\boldsymbol{\mu}, N) \equiv 1$$

均值，方差，特例

- 均值

$$E[\boldsymbol{m}] = \sum_{\boldsymbol{m}} p(\boldsymbol{m}|\mu, N)\boldsymbol{m} = N\boldsymbol{\mu}$$

- 方差

$$Cov[\boldsymbol{m}] = E[\boldsymbol{m}\boldsymbol{m}^\top] - E[\boldsymbol{m}](E[\boldsymbol{m}])^\top = \begin{cases} N\mu_i(1-\mu_i) & i = j \\ -N\mu_i\mu_j & i \neq j \end{cases}$$

- 特例
  - $K = 2, N = 1$： Bernoulli Distribution
  - $K = 2, N > 1$： Binomial Distribution
  - $K > 2, N = 1$： Categorical Distribution
  - $K > 2, N > 1$： Multinomial Distribution

高斯分布(正态分布)

- 因具有很多完美的性质而无处不在
  - 在某些制约条件下能使熵最大的分布
  - 中心极限定理
- D维向量$\boldsymbol{x}$的高斯分布

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

- 二次型(Quadratic Form)

$$\Delta^2 = (\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

Δ是马氏距离(Mahalanobis Distance)
$\boldsymbol{\Sigma}$是单位阵时，Δ为欧式距离(Euclidean Distance)

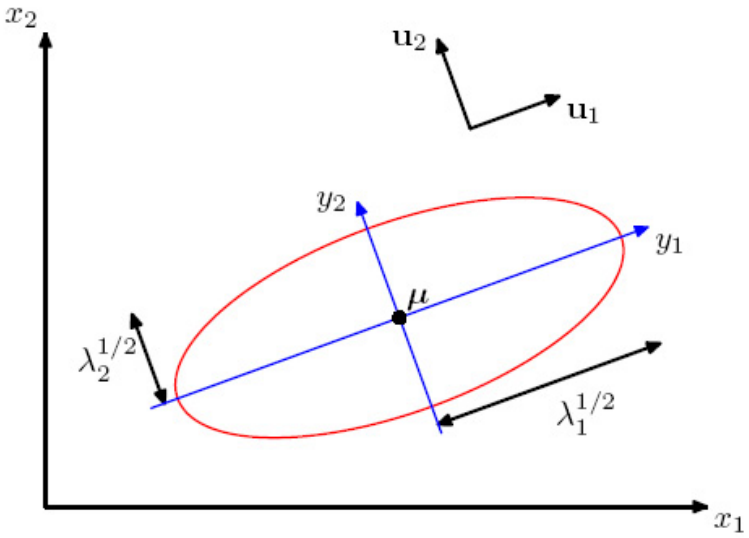
二次型的几何解释

- $\boldsymbol{\Sigma}$对应的特征向量和特征值分别为$\boldsymbol{u}_i$和λ_i

$$y_i = \boldsymbol{u}_i^\top (\boldsymbol{x} - \boldsymbol{\mu})$$

$$\Delta^2 = (\boldsymbol{x} - \boldsymbol{\mu})^\top \left[\sum_{i=1}^{D} \frac{1}{\lambda_i} \boldsymbol{u}_i \boldsymbol{u}_i^\top \right] (\boldsymbol{x} - \boldsymbol{\mu}) = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

The red curve shows the elliptical surface of constant probability density for a Gaussian in a two-dimensional space $\mathrm{x} = (x_1, x_2)$ on which the density is $\exp(-1/2)$ of its value at $\mathrm{x} = \boldsymbol{\mu}$. The major axes of the ellipse are defined by the eigenvectors u_i of the covariance matrix, with corresponding eigenvalues λ_i.



高阶矩

- 一阶矩

$$E[\boldsymbol{x}] = \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\boldsymbol{x}d\boldsymbol{x} = \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma})(\boldsymbol{x} + \boldsymbol{\mu})d\boldsymbol{x} = \boldsymbol{\mu}$$

- 二阶矩

$$E[\boldsymbol{x}\boldsymbol{x}^\top] = \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\boldsymbol{x}\boldsymbol{x}^\top d\boldsymbol{x} = \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma})(\boldsymbol{x}+\boldsymbol{\mu})(\boldsymbol{x}+\boldsymbol{\mu})^\top d\boldsymbol{x}$$

$$= \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma})\boldsymbol{x}\boldsymbol{x}^\top d\boldsymbol{x} + \boldsymbol{\mu}\boldsymbol{\mu}^\top = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top$$

- 三阶矩是？

3M

- 点估计(Point Estimation)中非常重要的三个概念
  - Mean(平均数/均值)

$$Mean = \int p(\boldsymbol{x})\boldsymbol{x}d\boldsymbol{x}$$

  - Mode(众数/最大值)

$$Mode = argmax\ p(\boldsymbol{x})$$

  - Median(中位数)

$$\int_{-\infty}^{Median} p(\boldsymbol{x})d\boldsymbol{x} = \int_{Median}^{\infty} p(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{2}$$

- 高斯分布：“三为一体”

协方差矩阵

- $\mathbf{\Sigma} = \{\sigma_{ij}\}$： General Covariance Matrix
- $\mathbf{\Sigma} = diag(\sigma_i^2)$： Diagonal Covariance Matrix
- $\mathbf{\Sigma} = \sigma^2\mathbf{I}$： Isotropic Covariance Matrix

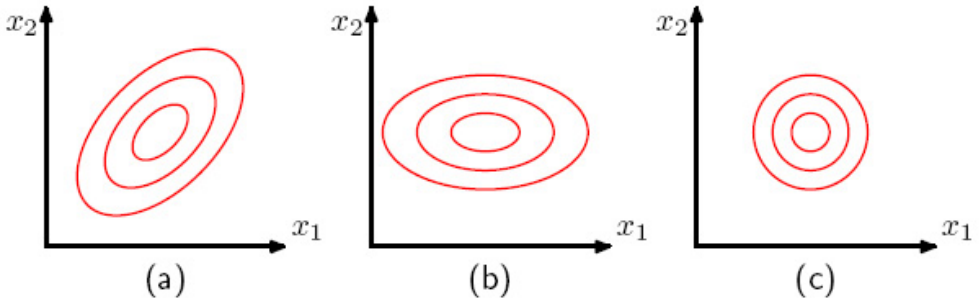


Figure 2.8 Contours of constant probability density for a Gaussian distribution in two dimensions in which the covariance matrix is (a) of general form, (b) diagonal, in which the elliptical contours are aligned with the coordinate axes, and (c) proportional to the identity matrix, in which the contours are concentric circles.

- $\mathbf{\Lambda} \equiv \mathbf{\Sigma}^{-1}$： Precision Matrix

条件分布和边缘分布

- 联合分布，条件分布，边缘分布关系

$$p(\boldsymbol{x},\boldsymbol{y}) = p(\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})$$

- 联合两次型Δ_{xy}^2

$$\begin{pmatrix} \boldsymbol{x}-\mu_x \\ \boldsymbol{y}-\mu_y \end{pmatrix}^\top \begin{pmatrix} \boldsymbol{\Lambda}_{xx} & \boldsymbol{\Lambda}_{xy} \\ \boldsymbol{\Lambda}_{yx} & \boldsymbol{\Lambda}_{yy} \end{pmatrix} \begin{pmatrix} \boldsymbol{x}-\mu_x \\ \boldsymbol{y}-\mu_y \end{pmatrix}$$

$$\begin{aligned} = & (\boldsymbol{x}-\mu_x)^\top \boldsymbol{\Lambda}_{xx}(\boldsymbol{x}-\mu_x) + (\boldsymbol{x}-\mu_x)^\top \boldsymbol{\Lambda}_{xy}(\boldsymbol{y}-\mu_y) + \\ & (\boldsymbol{y}-\mu_y)^\top \boldsymbol{\Lambda}_{yx}(\boldsymbol{x}-\mu_x) + (\boldsymbol{y}-\mu_y)^\top \boldsymbol{\Lambda}_{yy}(\boldsymbol{y}-\mu_y) \\ = & (\boldsymbol{x}-\mu_x)^\top \boldsymbol{\Lambda}'_{xx}(\boldsymbol{x}-\mu_x) + (\boldsymbol{y}-\mu'_y)^\top \boldsymbol{\Lambda}_{yy}(\boldsymbol{y}-\mu'_y) \end{aligned}$$

$$\begin{aligned} \boldsymbol{\Lambda}'_{xx} &= \boldsymbol{\Lambda}_{xx} - \boldsymbol{\Lambda}_{xy}\boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx} = \boldsymbol{\Sigma}_{xx}^{-1}(Schur\ \ Complement) \\ \mu'_y &= \mu_y - \boldsymbol{\Lambda}_{yy}^{-1}\boldsymbol{\Lambda}_{yx}(\boldsymbol{x}-\mu_x) = \mu_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{x}-\mu_x) \end{aligned}$$

条件分布和边缘分布

- 边缘分布

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{y} = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$$

- 条件分布(线性高斯模型)

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})} = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x})$$

$$\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)$$

$$\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Lambda}_{yy} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$$

- 两者均为高斯分布，反之亦然(或者两高斯乘积仍是高斯)

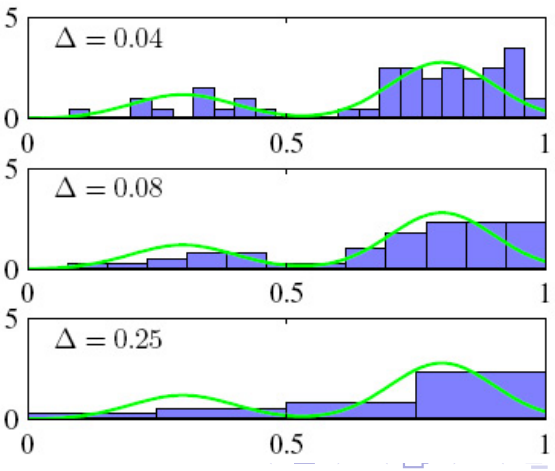
直方图

- 直方图定义下的PDF

$$p_i = \frac{n_i}{N\Delta_i}$$

- 数据量和区间的关系 区间太大：平；区间太小：尖
- 优点：一旦建立，容易计算
- 缺点：
  - 对数据量及存储量要求高
  - “区间”数随维数指数级增长

An illustration of the histogram approach to density estimation, in which a data set of 50 data points is generated from the distribution shown by the green curve. Histogram density estimates, based on (2.241), with a common bin width Δ are shown for various values of Δ.



多维情况下的概率密度

- 对于D维向量$\boldsymbol{x}$, 考虑包含$\boldsymbol{x}$的小区域$\mathcal{R}$上的概率

$$P = \int_{\mathcal{R}} p(\boldsymbol{x}) d\boldsymbol{x}$$

- 现有N个样本，落入$\mathcal{R}$的有K个，并假设
  - $\mathcal{R}$足够小，使得$p(\boldsymbol{x})$为常数
  - $\mathcal{R}$足够大，使得有点落入
- 则根据二项分布有

$$P = K/N$$

- 另外假设区域$\mathcal{R}$的容量为V，则

$$P = p(\boldsymbol{x})V$$

- 所以

$$p(\boldsymbol{x}) = \frac{K}{NV}$$

固定V，估计K

- Kernel Density Estimator (Parzen Estimator)

$$\begin{cases} V = h^D \\ K = \sum_{n=1}^{N} k\left(\frac{\boldsymbol{x}-\boldsymbol{x_n}}{h}\right) \end{cases} \Longrightarrow p(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{h^D}k\left(\frac{\boldsymbol{x}-\boldsymbol{x_n}}{h}\right)$$

- 核函数(Parzen窗函数)几何意义

$$k(\boldsymbol{u}) = \begin{cases} 1 & |\boldsymbol{u}| \leq 1/2 \\ 0 & otherwise \end{cases}$$

$$k(\boldsymbol{u}) = \frac{1}{(2\pi)^{D/2}}\exp\left\{-\frac{\| \boldsymbol{u} \|^2}{2}\right\}$$

固定K，估计V

- 最近邻法(KNN)

$$\begin{cases} K(constant) \\ V = h^D \end{cases} \Longrightarrow p(\boldsymbol{x}) = \frac{1}{Nh^D} K$$

- h不断增大，直到区域能够包含K个训练样本
- KNN思想用于分类问题(分类错误上限是两倍理论分类错误)

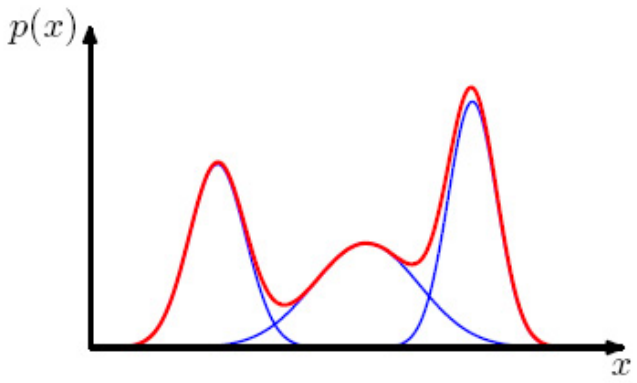
$$\begin{cases} p(\boldsymbol{x}|\mathcal{C}_k) & = & \frac{K_k}{N_k V} \\ p(\boldsymbol{x}) & = & \frac{K}{NV} \\ p(\mathcal{C}_k) & = & \frac{N_k}{N} \end{cases} \Longrightarrow p(\mathcal{C}_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\mathcal{C}_k)p(\mathcal{C}_k)}{p(\boldsymbol{x})} = \frac{K_k}{K}$$

- 核函数和最近邻法的最大问题都是要保留整个训练集

高斯混合模型(Gaussian Mixture Model)

- 高斯分布由于单峰的局限性，不能很好拟合一些实际数据
- GMM定义

$$p(\boldsymbol{x}) = \sum_{m=1}^{M} w_m \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$$

- M足够大时，可以拟合任意复杂分布
- 无法直接使用MLE进行参数估计，需要引入EM算法

Example of a Gaussian mixture distribution in one dimension showing three Gaussians (each scaled by a coefficient) in blue and their sum in red.