二、数理统计与参数估计

1、协方差

协方差

- 定义 $Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$
- 性质：

$$Cov(X,Y)=Cov(Y,X)$$

$$Cov(aX+b,cY+d)=acCov(X,Y)$$

$$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)$$

协方差的意义，当xy相互独立的时候E(X,Y)=E(X)E(Y)，协方差等于0，当xy线性相关的时候，协方差=1

协方差的意义

- 协方差是两个随机变量具有相同方向变化趋势的度量；
  - 若Cov(X,Y)>0，它们的变化趋势相同；
  - 若Cov(X,Y)<0，它们的变化趋势相反；
  - 若Cov(X,Y)=0，称X和Y不相关。

协方差矩阵，假设数据特征包括X1 X2 X3...

协方差矩阵

$$C=\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

- 对于n个随机向量$(X_1,X_2\ldots X_n)$，任意两个元素Xi和Xj都可以得到一个协方差，从而形成n*n的矩阵；协方差矩阵是对称阵。

$$c_{ij}=E\{[X_i-E(X_i)][X_j-E(X_j)]\}=Cov(X_i,X_j)$$

- 将随机向量X_i写成列向量，则$X=(X_1,X_2\ldots X_n)$为n列矩阵，将X的列分别去均值后，得到矩阵$\widetilde{X}$，则协方差矩阵为：

$$C=\frac{1}{n}\left(\widetilde{X}^T\cdot\widetilde{X}\right)$$

二、大数定理

大数定理

- 设随机变量$X_1,X_2\ldots X_n\ldots$互相独立，并且具有相同的期望μ和方差σ^2。作前n个随机变量的平均$Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i$，则对于任意正数ε，有

$$\lim_{n\to\infty}P\{|Y_n-\mu|<\varepsilon\}=1$$

大数定理的意义在于，当N无穷大的时候，均值无限逼近于期望

三、极大似然估计

贝叶斯公式带来的思考 $P(A|D)=\frac{P(D|A)P(A)}{P(D)}$

- 给定某些样本D，在这些样本中计算某结论A_1、A_2……A_n出现的概率，即$P(A_i|D)$

$$\max P(A_i|D)=\max\frac{P(D|A_i)P(A_i)}{P(D)}=\max\left(P(D|A_i)P(A_i)\right)\xrightarrow{P(A_i)\text{近似相等}}\max P(D|A_i)$$
$$\Rightarrow\max P(A_i|D)\to\max P(D|A_i)$$

A1、A2、An可以看成是分布或者参数的取值，极大似然估计就是求得在已知样本下的最大分布的概率

极大似然估计

- 设总体分布为f(x, θ)，$X_1, X_2 \ldots X_n$为该总体采样得到的样本。因为$X_1,X_2\ldots X_n$独立同分布，于是，它们的联合密度函数为：

$$L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots \theta_k) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2, \cdots \theta_k)$$

- 这里，θ被看做固定但未知的参数；反过来，因为样本已经存在，可以看成$x_1,x_2\ldots x_n$是固定的，L(x, θ)是关于θ的函数，即似然函数。
- 求参数θ的值，使得似然函数取极大值，这种方法就是极大似然估计。