Ch 5.6 条件分布与条件期望

回顾前一次课

X与Y的相关系数 $\rho_{XY}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$、性质

二维正态分布$\begin{pmatrix}X\\Y\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_1\\\mu_2\end{pmatrix}\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\\rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$， ρ为X与Y的相关系数、$\boldsymbol{X}$**与**$\boldsymbol{Y}$**独立** $\Leftrightarrow$ $\boldsymbol{X}$**与**$\boldsymbol{Y}$**不相关**

随机向量$\boldsymbol{X}$的协方差矩阵

$$Cov(X)=\Sigma=\begin{pmatrix}Cov(X_1,X_1) & \cdots & Cov(X_1,X_n)\\Cov(X_2,X_1) & \cdots & Cov(X_2,X_n)\\\vdots & & \vdots\\Cov(X_n,X_1) & \cdots & Cov(X_n,X_n)\end{pmatrix}$$

半正定

正态分布相关的结论$\boldsymbol{X}=(X_1,X_2,\cdots,X_n)^{\mathsf{T}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$

离散型随机变量的条件分布

设二维离散型随机变量(X,Y)的分布列为$\{p_{ij}\}$，若Y的边缘分布$P\left(Y=y_j\right)=p_{\cdot j}>0$，称

$$P\left(X=x_i \middle| Y=y_j\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(Y=y_j\right)}=\frac{p_{ij}}{p_{\cdot j}}$$

在$Y=y_j$条件下随机变量X的**条件分布列**.

类似定义在$X=x_i$条件下随机变量Y的条件分布列

连续型随机变量

对于连续型随机变量(X, Y), 对任意$x, y \in (-\infty, +\infty)$, 有

$$P(X = x) = 0 \text{和} P(Y = y) = 0$$

因此不能利用离散随机变量的条件概率推导连续随机变量的条件分布.

但可借鉴离散的思想

分布函数的极限情况

分布函数

$$F_{X|Y}(x|y) = \lim_{\epsilon \to 0^+} P[X \leq x|\, y \leq Y \leq y + \epsilon]$$

$$= \lim_{\epsilon \to 0^+} \frac{P[X \leq x, y \leq Y \leq y + \epsilon]}{P[y \leq Y \leq y + \epsilon]}$$

根据积分中值定理展开

连续随机变量的条件概率

定义：设连续随机变量(X,Y)的联合概率密度为$f(x,y)$，以及Y的边缘概率密度为$f_Y(y)$，对任意$f_Y(y)>0$，称

$$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$$

在$\boldsymbol{Y=y}$**条件下随机变量**$\boldsymbol{X}$**的条件概率密度**

$$F_{X|Y}(x|y)=P[X\le x|Y=y]=\int_{-\infty}^{x} f_{X|Y}(v|y)dv$$

为$\boldsymbol{Y=y}$**条件下**$\boldsymbol{X}$**的条件分布函数**

乘法公式

对于随机变量X和Y, 有

$$f(x,y) = f_X(x) f_{Y|X}(y|x), \qquad (f_X(x) > 0)$$

$$f(x,y) = f_Y(y) f_{X|Y}(x|y) \qquad (f_Y(y) > 0)$$

对独立的随机变量X和Y, 有

$$f_{Y|X}(y|x) = f_Y(y) \qquad f_{X|Y}(x|y) = f_X(x)$$

例题

设二维随机变量(X,Y)的概率密度

$$f(x,y)=\begin{cases}\dfrac{Ae^{-\frac{x}{y}}e^{-y}}{y} & 0<x<+\infty, 0<y<+\infty\\ 0 & \text{其它}\end{cases}$$

求$P(X>1|Y=y)$.

例题

已知随机变量$X \sim U(0,1)$, 当观察到$X = x$的条件下, 随机变量$Y \sim U(x, 1)$. 求Y和$Z = X - Y$的概率密度.

课堂练习

设随机变量(X,Y)的概率密度为

$$f(x,y)=\begin{cases}Ae^{-y} & y>x>0\\ 0 & \text{其它}\end{cases}$$

求1) 常数A

2)求X与Y的边缘概率密度或条件概率密度$f_{X|Y}(x|y)$

3)求X与Y是否独立?

4)求$Z=X+Y$的概率密度

多维正态分布的条件分布

定理： 多维正太分布的条件分布是正太分布.

条件期望

离散随机变量(X,Y)在$Y=y$条件下随机变量X的期望为

$$E[X|Y=y]=\sum_{i=1}^{+\infty}x_iP(X=x_i|Y=y)$$

连续随机变量(X,Y)在$Y=y$条件下随机变量X的期望为

$$E[X|Y=y]=\int_{-\infty}^{+\infty}x\,f(x|y)dx$$

注：$E[X|Y=y]$是y的函数

条件期望的性质

定理： 对随机变量$X_1, X_2, \cdots, X_n$和常数$c_1, c_2, \cdots, c_n$有

$$E\left[\sum_{i=1}^{n} c_i X_i | Y = y\right] = \sum_{i=1}^{n} c_i E[X_i | Y = y].$$

条件期望的性质

全期望公式(The law of total expectation)

对随机变量X和事件A有

$$E[X]=E[X|A]P(A)+E[X|\bar{A}]\big(1-P(A)\big)$$

其中事件$\bar{A}$为事件A的补.

条件期望的性质

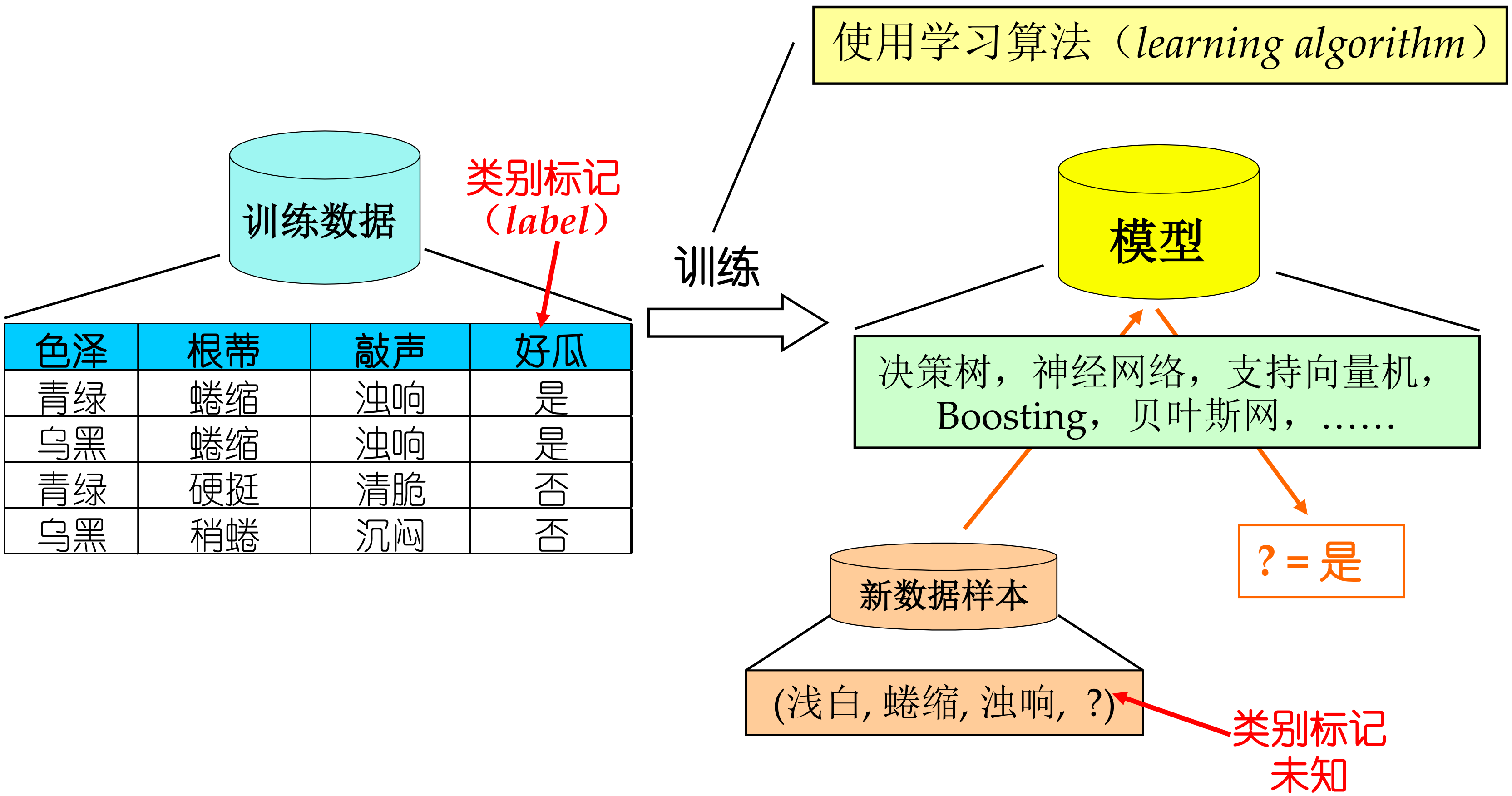
定理： 对二维随机变量(X, Y)有

$$E[X] = E_Y[E(X|Y)].$$

特别地, 对二维离散随机变量有

$$E[X] = E[E(X|Y)] = \sum_{Y=y_j} P\left[Y = y_j\right] E\left[X \middle| Y = y_j\right]$$

典型的机器学习过程



色泽	根蒂	敲声	好瓜
青绿	蜷缩	浊响	是
乌黑	蜷缩	浊响	是
青绿	硬挺	清脆	否
乌黑	稍蜷	沉闷	否