1. 计算$\frac{\partial \ln \det(\boldsymbol{A})}{\partial x}$

$$\frac{\partial \ln \det(A)}{\partial x} = \frac{\partial \ln \det(A)}{\partial \det(A)} \cdot \frac{\partial \det(A)}{\partial x}$$

其中 $$\frac{\partial \det(A)}{\partial x} = \sum_{i,j} \frac{\partial \det(A)}{\partial a_{ij}} \cdot \frac{\partial a_{ij}}{\partial x} = \sum_{i,j} A_{ij} \cdot \frac{\partial a_{ij}}{\partial x}$$

$$= \sum_{i,j} (A^*)_{ji} \frac{\partial a_{ij}}{\partial x} = \sum_{i} \left(\frac{\partial A}{\partial x} \cdot A^*\right)_{ii}$$

$$\therefore \frac{\partial \ln \det(A)}{\partial x} = \frac{1}{\det(A)} \cdot tr\left(\frac{\partial A}{\partial x} \cdot A^*\right) = tr\left(\frac{\partial A}{\partial x} \cdot \frac{A^*}{\det(A)}\right)$$

$$= tr\left(\frac{\partial A}{\partial x} \cdot A^{-1}\right)$$

1.2 与使用单个合取式来进行假设表示相比, 使用“析合范式”将使得假设空间具有更强的表示能力. 例如

析合范式即多个合取式的析取.

$$\text{好瓜} \leftrightarrow \Big((\text{色泽}=*) \wedge (\text{根蒂}=\text{蜷缩}) \wedge (\text{敲声}=*)\Big)$$
$$\vee \Big((\text{色泽}=\text{乌黑}) \wedge (\text{根蒂}=*) \wedge (\text{敲声}=\text{沉闷})\Big),$$

会把“(色泽=青绿) ∧ (根蒂=蜷缩) ∧ (敲声=清脆)”以及“(色泽=乌黑) ∧ (根蒂=硬挺) ∧ (敲声=沉闷)”都分类为“好瓜”. 若使用最多包含 k 个合取式的析合范式来表达表 1.1 西瓜分类问题的假设空间, 试估算共有多少种可能的假设.

提示: 注意冗余情况, 如 $(A=a) \vee (A=*)$ 与 $(A=*)$ 等价.

表 1.1 西瓜数据集

编号	色泽	根蒂	敲声	好瓜
1	青绿	蜷缩	浊响	是
2	乌黑	蜷缩	浊响	是
3	青绿	硬挺	清脆	否
4	乌黑	稍蜷	沉闷	否

该问题包含3种属性，其中属性色泽有2个取值青绿、乌黑，属性根蒂有3个取值蜷缩、硬挺、稍蜷，属性敲声有3个取值浊响、清脆、沉闷.

假设空间的大小为 $(2+1)\times(3+1)\times(3+1)+1 = 49$ 种

具体假设有 $2\times3\times3 = 18$ 种

1个属性泛化的假设有 $2\times3+2\times3+3\times3 = 21$ 种

2个属性泛化的假设有 $2+3+3 = 8$ 种

3个属性泛化的假设有 1 种

不考虑空集与冗余的情况下 K 的最大取值为48

若使用最多包含 k个合取式的析合范式来表达，共有 $\sum_{i=1}^{k} C_{48}^{i}$ 种可能的假设

若考虑冗余，k的最大值为18

k=1时，共48种，若使用最多包含1个合取式的析合范式来表达，共有1种可能的假设

k=18时，共1种，若使用最多包含18个合取式的析合范式来表达，共有$(2^{18}-1)$种可能的假设

k为其他值时较复杂.

3. 已知随机变量 $\boldsymbol{x}=[\boldsymbol{x}_1,\boldsymbol{x}_2]\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 计算 $P(\boldsymbol{x}_1), P(\boldsymbol{x}_1|\boldsymbol{x}_2)$

$\because x=[x_1,x_2]\sim N(\mu,\Sigma)$,

二维时有 $X=[x_1,x_2]\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,

其中 $\mu=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$, $\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$

$$\therefore f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$$

$$\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]=\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_2-\mu_2}{\sigma_2}-\rho\frac{x_1-\mu_1}{\sigma_1}\right)^2+\frac{(x_1-\mu_1)^2}{\sigma_1^2}(1-\rho^2)\right]$$

设 $t=\dfrac{\frac{x_2-\mu_2}{\sigma_2}-\rho\frac{x_1-\mu_1}{\sigma_1}}{\sqrt{2}\sqrt{1-\rho^2}}$ 则 $f(x_1,x_2)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}e^{-t^2}$

$$P(x_1)=\int_{-\infty}^{\infty}f(x_1,x_2)\,dx_2=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{\infty}e^{-t^2}(\sigma_2\sqrt{2}\sqrt{1-\rho^2})\,dt=\frac{1}{\sqrt{2\pi}\,\sigma_1}e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}}$$

同理, $P(X_2)=\dfrac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}$

$$P(x_1|x_2)=\frac{f(x_1,x_2)}{f(x_2)}$$

$$=\frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}}{\dfrac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}}$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho^2}}e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left\{x_1-\left[\mu_1+\frac{\sigma_1\rho(x_2-\mu_2)}{\sigma_2}\right]\right\}^2}$$

4. 证明范数 $\|\boldsymbol{x}\|_p$ 是凸函数

$f(x) = \|x\|_p$，显然 $dom(f) = R^n$，是凸集

对范数 $\|x\|_p$ 有 $1 \le p \le +\infty$

由 Minkowski 不等式，$\forall t \in [0,1]$，有 $\|tx + (1-t)y\|_p \le \|tx\|_p + \|(1-t)y\|_p = t\|x\|_p + (1-t)\|y\|_p$

$$\text{即 } \|tx + (1-t)y\|_p \le t\|x\|_p + (1-t)\|y\|_p$$

$\therefore$ $\|x\|_p$ 是凸函数

若不使用 Minkowski 不等式，

$f(x) = \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$，不妨设 $x_i \ge 0$，$i = 1, \cdots, n$

$$\nabla f(x) = \frac{1}{p}(\sum_{j=1}^n x_j^p)^{\frac{1}{p}-1} \cdot p \cdot [x_1^{p-1}, x_2^{p-1}, \cdots, x_n^{p-1}]^T = (\sum_{j=1}^n x_j^p)^{\frac{1}{p}-1} [x_1^{p-1}, x_2^{p-1}, \cdots, x_n^{p-1}]^T$$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial (\sum_{k=1}^n x_k^p)^{\frac{1}{p}-1} x_i^{p-1}}{\partial x_i} = \frac{1-p}{p}(\sum_{k=1}^n x_k^p)^{\frac{1}{p}-2} \cdot p x_i^{p-1} \cdot x_i^{p-1} + (\sum_{k=1}^n x_k^p)^{\frac{1}{p}-1} \cdot (p-1) \cdot x_i^{p-2}$$

$$= (1-p)(\sum_{k=1}^n x_k^p)^{\frac{1}{p}} \frac{1}{x_i^2} \left(\frac{x_i^p}{\sum_{k=1}^n x_k^p} - 1\right) \frac{x_i^p}{\sum_{k=1}^n x_k^p}$$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial (\sum_{k=1}^n x_k^p)^{\frac{1}{p}-1} x_i^{p-1}}{\partial x_j} = (1-p)(\sum_{k=1}^n x_k^p)^{\frac{1}{p}-2} \cdot x_i^{p-1} x_j^{p-1}$$

$$= (1-p)(\sum_{k=1}^n x_k^p)^{\frac{1}{p}} \frac{1}{x_i x_j} \left(\frac{x_i^p}{\sum_{k=1}^n x_k^p}\right)\left(\frac{x_j^p}{\sum_{k=1}^n x_k^p}\right)$$

令 $z = (x_1^p, x_2^p, \cdots, x_n^p)^T$　　$A = diag(\frac{1}{x_1}, \frac{1}{x_2}, \cdots, \frac{1}{x_n})$

则 $\nabla^2 f = (1-p) f A^T (zz^T - (1^T z) diag(z)) A \frac{1}{(1^T z)^2}$

$$\forall y \in R^n,\ y^T \nabla^2 f y = \frac{(1-p)f}{(1^T z)^2} y^T A^T (zz^T - (1^T z) diag(z)) Ay \text{，令 } u = Ay\text{，}$$

$$= \frac{(1-p)f}{(1^T z)^2}[u^T z z^T u - (1^T z) u^T diag(z) u]$$

$$= \frac{(1-p)f}{(1^T z)^2}[(\sum_{i=1}^n u_i z_i)^2 - (\sum_{i=1}^n z_i)(\sum_{i=1}^n u_i^2 z_i)] \text{，令 } a_i = u_i\sqrt{z_i},\ b_i = \sqrt{z_i}$$

$$= \frac{(1-p)f}{(1^T z)^2}[(a^T b)^2 - (b^T b)(a^T a)]$$

其中 $p \ge 1$，$1-p \le 0$，$f \ge 0$，$(1^T z)^2 \ge 0$，由 Cauchy 不等式，$(a^T b)^2 - (b^T b)(a^T a) \le 0$

$\therefore$ $\forall y \in R^n$，$y^T \nabla^2 f y \ge 0$，Hessian 矩阵半正定，是凸函数

5. 证明判定凸函数的0阶和1阶条件相互等价

$$\forall x,y \in dom(f), \forall t \in [0,1], f(tx+(1-t)y) \le tf(x)+(1-t)f(y)$$

$$\Updownarrow$$

$$\forall x,y \in dom(f),\ f(y) \ge f(x) + \nabla f(x)^T(y-x)$$

"$\Rightarrow$"

$\forall x, y \in dom(f)$, $\forall t \in [0,1]$, $f(tx+(1-t)y) \le tf(x)+(1-t)f(y)$

令 $\theta = 1-t$, $\Leftrightarrow f((1-\theta)x+\theta y) \le (1-\theta)f(x)+\theta f(y)$

$\Leftrightarrow \theta f(y) \ge f((1-\theta)x+\theta y) + f(x) - \theta f(x)$

$\Leftrightarrow f(y) \ge f(x) + \dfrac{f(x+\theta(y-x)) - f(x)}{\theta}$

两边同时对θ取极限，有

$$f(y) \ge f(x) + \lim_{\theta \to 0} \frac{f(x+\theta(y-x)) - f(x)}{\theta\,(y-x)} \cdot (y-x)$$

$\Leftrightarrow f(y) \ge f(x) + \nabla f(x)^T (y-x)$

"$\Leftarrow$"

$\forall x, y \in dom f$，令 $z = tx+(1-t)y$, $t \in [0,1]$, $x \ne y$

有 $f(x) \ge f(z) + \nabla f(z)^T (x-z)$ ①

$f(y) \ge f(z) + \nabla f(z)^T (y-z)$ ②

t① $+ (1-t)$②：

$$tf(x) + (1-t)f(y) \ge f(z) + \nabla f(z)^T \underbrace{(tx - tz + y - ty - z + tz)}_{0}$$

即 $tf(x) + (1-t)f(y) \ge f(tx+(1-t)y)$