HW1

PB20111689 蓝俊玮

T1

$\frac{\partial \ln \det(\mathbf{A})}{\partial x} = \frac{\partial \ln \det(\mathbf{A})}{\partial \det(\mathbf{A})} \frac{\partial \det(\mathbf{A})}{\partial x} = \frac{1}{\det(\mathbf{A})} \frac{\partial \det(\mathbf{A})}{\partial x}$

由链式法则 $\frac{\partial \det(\mathbf{A})}{\partial x} = \sum_i \sum_j \frac{\partial \det(\mathbf{A})}{\partial a_{ij}} \frac{\partial a_{ij}}{\partial x}$ 而
$\frac{\partial \det(\mathbf{A})}{\partial a_{ij}} = \det(\mathbf{A}) \mathbf{A}_{ji}^{-1}, \frac{\partial \det(\mathbf{A})}{\partial x} = \sum_i \sum_j \det(\mathbf{A}) \mathbf{A}_{ji}^{-1} (\frac{\partial \mathbf{A}}{\partial x})_{ij} = \det(\mathbf{A}) \sum_i \sum_j \mathbf{A}_{ji}^{-1} (\frac{\partial \mathbf{A}}{\partial x})_{ij} = \det(\mathbf{A}) \operatorname{tr}\left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x}\right)$

所以结果为 $\frac{\partial \ln \det(\mathbf{A})}{\partial x} = \operatorname{tr}\left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x}\right)$

T2

根据给定的西瓜数据，西瓜的色泽共有 2 种，根蒂和敲声有 3 种。所以对于单个合取式来说共有 $2 \times 3 \times 3 = 18$ 种，那么对于所有的析取范式来说，其上限为 $\binom{18}{0} + \binom{18}{1} + \binom{18}{2} + \ldots + \binom{18}{18} = 2^{18}$ 种。当 $k = 9$ 的时候，去重所有可能的结果后就可以使假设空间达到上限。

T3

因为 $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2] \sim \mathcal{N}(\mu, \Sigma)$，则 $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$，则可以计算得到条件分布 $(\mathbf{x}_1 | \mathbf{x}_2) \sim \mathcal{N}(\overline{\mu}, \overline{\Sigma})$ 其中 $\overline{\mu} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (\mathbf{x}_2 - \mu_2), \overline{\Sigma} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$。而对于边缘分布，由其性质易知边缘分布的参数就是与自己本身参数所相关的一维正态分布，所以其分布为 $\mathbf{x}_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$。因此得到了这些分布后，其概率也就显然可得

$$P(\mathbf{x}_1) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma_{11})}} \exp\left(-\frac{1}{2} (\mathbf{x}_1 - \mu_1)^T \Sigma_{11}^{-1} (\mathbf{x}_1 - \mu_1)\right)$$
$$P(\mathbf{x}_1 | \mathbf{x}_2) = \frac{1}{\sqrt{(2\pi)^n \det\left(\overline{\Sigma}\right)}} \exp\left(-\frac{1}{2} (\mathbf{x}_1 - \overline{\mu})^T \overline{\Sigma}^{-1} (\mathbf{x}_1 - \overline{\mu})\right)$$

T4

首先范数 $\|\mathbf{x}\|_p$ 的定义式为 $\|\mathbf{x}\|_p = \sqrt[p]{(\sum_{i=1}^{i=n} |x_i|^p)}$，那么根据凸函数的定义 $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \lambda \in [0,1]$ 有 $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$。那么可以根据范数的性质 $\lambda \|\mathbf{x}\|_p = \lambda \|\mathbf{x}\|_p$ 和三角性质 $\|\mathbf{x} + \mathbf{y}\|_p \le \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$ 可以得到 $\|\lambda \mathbf{x} + (1-\lambda)\mathbf{y}\|_p \le \|\lambda \mathbf{x}\|_p + \|(1-\lambda)\mathbf{y}\|_p = \lambda \|\mathbf{x}\|_p + (1-\lambda)\|\mathbf{y}\|_p$，所以范数 $\|\mathbf{x}\|_p$ 满足凸函数的性质，因此范数 $\|\mathbf{x}\|_p$ 是凸函数。

T5

充分性证明：

因为有 $f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \le \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \forall \lambda \in [0,1]$，所以可得

$$f(\lambda \mathbf{y} + (1-\lambda)\mathbf{x}) \le \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{x})$$
$$f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) \le \lambda f(\mathbf{y}) + (1-\lambda) f(\mathbf{x})$$
$$f(\mathbf{y}) \ge f(\mathbf{x}) + \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda}$$
$$\lim_{\lambda \to 0} f(\mathbf{y}) \ge \lim_{\lambda \to 0} (f(\mathbf{x}) + \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda})$$
$$f(\mathbf{y}) \ge f(\mathbf{x}) + \lim_{\lambda \to 0} \frac{f(\mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})) - f(\mathbf{x})}{\lambda}$$
$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x})$$

必要性证明：

因为有 $f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$，所以假设并得到：

$$\mathbf{z} = t\mathbf{y} + (1-t)\mathbf{x}$$
$$f(\mathbf{y}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{y}-\mathbf{z})$$
$$f(\mathbf{x}) \geq f(\mathbf{z}) + \nabla f(\mathbf{z})^T(\mathbf{x}-\mathbf{z})$$

接着对 1 式和 2 式分别乘 t 和 $1-t$ 并相加得到

$$\begin{aligned} tf(\mathbf{y}) + (1-t)f(\mathbf{x}) &\geq tf(\mathbf{z}) + t\nabla f(\mathbf{z})^T(\mathbf{y}-\mathbf{z}) + (1-t)f(\mathbf{z}) + (1-t)\nabla f(\mathbf{z})^T(\mathbf{x}-\mathbf{z}) \\ &= f(\mathbf{z}) + \nabla f(\mathbf{z})^T(t(\mathbf{y}-\mathbf{x}) + (\mathbf{x}-\mathbf{z})) \\ &= f(\mathbf{z}) + \nabla f(\mathbf{z})^T(t(\mathbf{y}-\mathbf{x}) + (\mathbf{x}-t\mathbf{y}-(1-t)\mathbf{x})) = f(\mathbf{z}) \end{aligned}$$

因此得到了 $tf(\mathbf{y}) + (1-t)f(\mathbf{x}) \geq f(t\mathbf{y} + (1-t)\mathbf{x}) = f(\mathbf{z})$。

综上所述，可以得到凸函数的 0 阶和 1 阶条件相互等价。