3.1 多元随机变量联合分布
3.2 多元高斯分布

纪一仑

Harbin Institute of Technology,ShenZhen

Group Meeting, November 2022

Table of Contents

Table of Contents

3.1.1 协方差

协方差 (covariance): 用于度量 rv's X、Y 的 (线性) 相关性

$$\mathrm{Cov}[X, Y] \triangleq \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

协方差矩阵 (covariance matrix)(x 为 D 维随机向量)

$$\mathrm{Cov}[\boldsymbol{x}] \triangleq \mathbb{E}\left[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])^\top\right] \triangleq \boldsymbol{\Sigma}$$

$$= \begin{pmatrix} \mathbb{V}[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_D] \\ \mathrm{Cov}[X_2, X_1] & \mathbb{V}[X_2] & \cdots & \mathrm{Cov}[X_2, X_D] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[X_D, X_1] & \mathrm{Cov}[X_D, X_2] & \cdots & \mathbb{V}[X_D] \end{pmatrix}$$

协方差相关性质

- $\mathrm{Cov}[\boldsymbol{x}] = \Sigma = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top$
- $$\begin{aligned}\mathrm{Cov}[\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}] = \mathrm{Cov}[\boldsymbol{A}\boldsymbol{x}] &= \mathbb{E}[\boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^\top\boldsymbol{A}^\top] - \mathbb{E}[\boldsymbol{A}\boldsymbol{x}]\mathbb{E}[\boldsymbol{x}^\top\boldsymbol{A}^\top] \\ &= \boldsymbol{A}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]\boldsymbol{A}^\top - \boldsymbol{A}\mathbb{E}[\boldsymbol{x}]\mathbb{E}[\boldsymbol{x}^\top]\boldsymbol{A}^\top \\ &= \boldsymbol{A}\,\mathrm{Cov}[\boldsymbol{x}]\boldsymbol{A}^\top\end{aligned}$$

互协方差 (cross-covariance)

- $\mathrm{Cov}[\boldsymbol{x},\boldsymbol{y}] = \mathbb{E}\left[(\boldsymbol{x}-\mathbb{E}[\boldsymbol{x}])(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])^\top\right]$
- $\mathrm{Cov}[\boldsymbol{x}] = \mathrm{Cov}[\boldsymbol{x},\boldsymbol{x}]$

3.1.2 相关性

协方差的值可以是任意的，有时候使用相关系数 (correlation coefficient) 更方便。

相关系数定义

$$\rho \triangleq \mathrm{corr}[X, Y] \triangleq \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}}$$

性质

- $-1 \leq \rho \leq 1$
- $\mathrm{corr}[X, Y] = 1$ iff $Y = aX + b (a > 0)$
- $a = \mathrm{Cov}[X, Y] / \mathbb{V}[X]$

3.1.2 相关性

性质的解释

$$\rho = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}} = \sqrt{\frac{[\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]]^2}{\mathbb{E}[(X - \mathbb{E}[X])^2]\mathbb{E}[(Y - \mathbb{E}[Y])^2]}}$$

We first construct an function $g(t)$

$$\begin{aligned} g(t) &= t^2E(X^2) + 2tE(XY) + E(Y^2) \\ &= E(t^2X^2 + 2tXY + Y^2) \\ &= E(tX + Y)^2 \\ &\geqslant 0 \end{aligned}$$

This function has at most one crossing point with t axis line $t = 0$. Therefore, we know $\Delta \leqslant 0$. That is

$$\begin{aligned} \Delta &= (2E(XY))^2 - 4E(X^2)E(Y^2) \\ &= 4[E(XY)]^2 - 4E(X^2)E(Y^2) \\ &\leqslant 0 \end{aligned}$$

So that $[E(XY)]^2 \leqslant E(X^2)E(Y^2)$ is proved.

Figure: 柯西不等式

3.1.2 相关性

$$\rho = 1 \iff Y - \mathbb{E}[Y] = a[X - \mathbb{E}[X]], a > 0$$

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[a(X - \mathbb{E}[X])^2] = a\mathbb{V}[X]$$

3.1.2 相关性

相关矩阵 (correlation matrix)

$$\mathrm{corr}(\boldsymbol{x}) = \begin{pmatrix} 1 & \frac{\mathbb{E}[(X_1-\mu_1)(X_2-\mu_2)]}{\sigma_1\sigma_2} & \cdots & \frac{\mathbb{E}[(X_1-\mu_1)(X_D-\mu_D)]}{\sigma_1\sigma_D} \\ \frac{\mathbb{E}[(X_2-\mu_2)(X_1-\mu_1)]}{\sigma_2\sigma_1} & 1 & \cdots & \frac{\mathbb{E}[(X_2-\mu_2)(X_D-\mu_D)]}{\sigma_2\sigma_D} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbb{E}[(X_D-\mu_D)(X_1-\mu_1)]}{\sigma_D\sigma_1} & \frac{\mathbb{E}[(X_D-\mu_D)(X_2-\mu_2)]}{\sigma_D\sigma_2} & \cdots & 1 \end{pmatrix}$$

X、Y 的相关系数为 0 不代表相互独立。
例如 $X \sim U(-1,1)$ 和 $Y = X^2$，明显 $\mathbb{E}[X] = 0, \mathbb{E}[XY] = \mathbb{E}[X^3] = 0$，所以 $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$。但是这里 Y 是完全由 X 决定的。

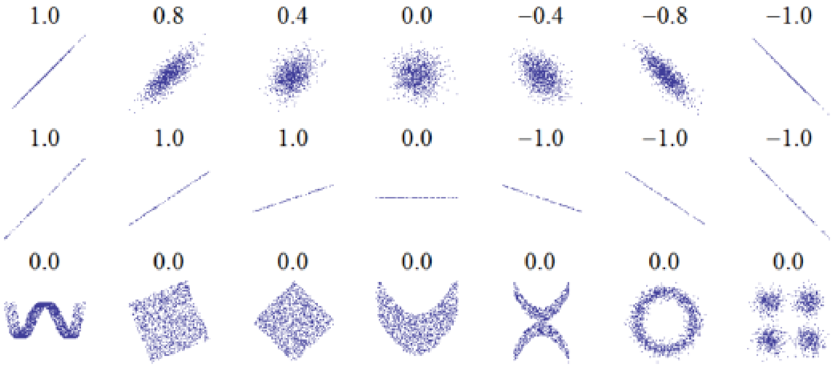


Figure: 数字为相关系数

相关性不代表因果性

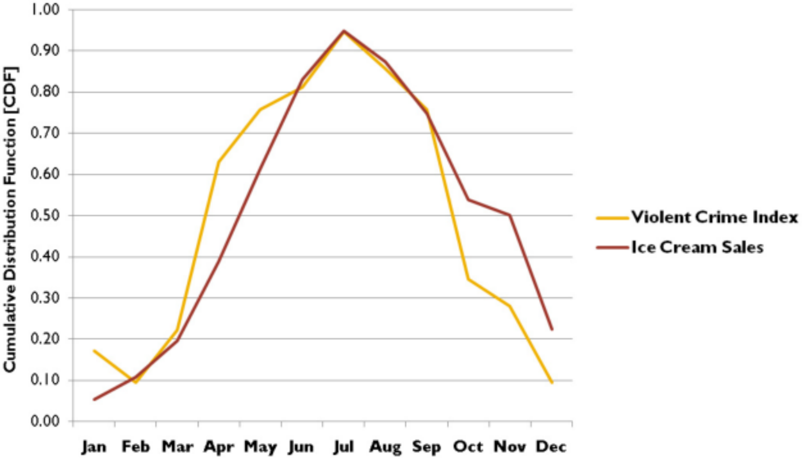


Figure: 冰淇淋与谋杀

在不同组数据中出现的相同的关系，在考察所有数据的时候，可能会消失，甚至是相反。

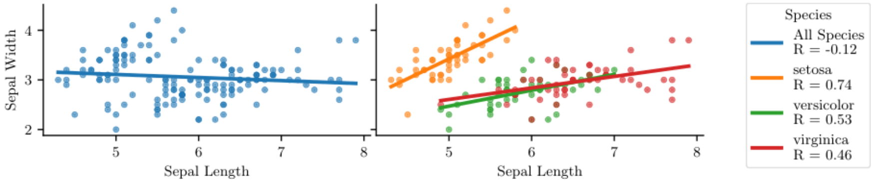


Figure: 鸢尾花: 花萼的长度和宽度

不能忽略组和组之间的差异

Table of Contents

MVN 的 pdf(概率密度函数)

$$\mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]$$

$$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{y}] \in \mathbb{R}^D$$

$$\boldsymbol{\Sigma} = \mathrm{Cov}[\boldsymbol{y}] \triangleq \mathbb{E}\left[(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])^\top\right]$$

$$= \begin{pmatrix} \mathbb{V}[Y_1] & \mathrm{Cov}[Y_1, Y_2] & \cdots & \mathrm{Cov}[Y_1, Y_D] \\ \mathrm{Cov}[Y_2, Y_1] & \mathbb{V}[Y_2] & \cdots & \mathrm{Cov}[Y_2, Y_D] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_D, Y_1] & \mathrm{Cov}[Y_D, Y_2] & \cdots & \mathbb{V}[Y_D] \end{pmatrix}$$

3.2.1 定义

在二维下，MVN 也被叫做二元高斯分布 (bivariate Gaussian distribution)，它的 pdf 可以表示为 $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

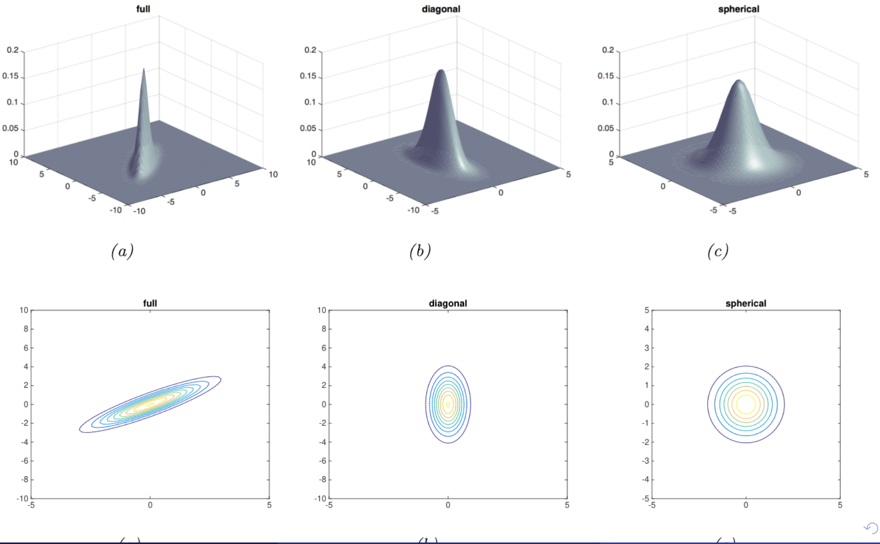
$$\rho = \operatorname{corr}[Y_1, Y_2] \triangleq \frac{\operatorname{Cov}[Y_1, Y_2]}{\sqrt{\mathbb{V}[Y_1]\,\mathbb{V}[Y_2]}} = \frac{\sigma_{12}^2}{\sigma_1\sigma_2}$$

$$\begin{aligned} p(y_1, y_2) =& \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Bigg(-\frac{1}{2(1-\rho^2)} \times \\ & \left[\frac{(y_1-\mu_1)^2}{\sigma_1^2} + \frac{(y_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(y_1-\mu_1)}{\sigma_1}\frac{(y_2-\mu_2)}{\sigma_2}\right]\Bigg) \end{aligned}$$

三种协方差矩阵

- full covariance matrix，有所有的参数
- diagonal covariance matrix，只有对角线上的参数
- spherical covariance matrix，只有一个参数

$$
\begin{aligned}
\boldsymbol{\Sigma} = \mathrm{Cov}[\boldsymbol{y}] &\triangleq \mathbb{E}\left[(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])(\boldsymbol{y} - \mathbb{E}[\boldsymbol{y}])^\top\right] \\
&= \begin{pmatrix}
\mathbb{V}[Y_1] & \mathrm{Cov}[Y_1, Y_2] & \cdots & \mathrm{Cov}[Y_1, Y_D] \\
\mathrm{Cov}[Y_2, Y_1] & \mathbb{V}[Y_2] & \cdots & \mathrm{Cov}[Y_2, Y_D] \\
\vdots & \vdots & \ddots & \vdots \\
\mathrm{Cov}[Y_D, Y_1] & \mathrm{Cov}[Y_D, Y_2] & \cdots & \mathbb{V}[Y_D]
\end{pmatrix}
\end{aligned}
$$



full
diagonal
spherical
(a)
(b)
(c)
full
diagonal
spherical

3.2.2 马氏距离 (Mahalanobis distance)

$$\mathcal{N}(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]$$

$$\log p(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) + \mathrm{const}$$

定义

$$\Delta^2 \triangleq (\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})$$

深入了解

假设 Σ 是正定矩阵

$$\boldsymbol{\Sigma} = \sum_{d=1}^{D} \lambda_d \boldsymbol{u}_d \boldsymbol{u}_d^\top$$

$$\boldsymbol{\Sigma}^{-1} = \sum_{d=1}^{D} \frac{1}{\lambda_d} \boldsymbol{u}_d \boldsymbol{u}_d^\top \quad z_d \triangleq \boldsymbol{u}_d^\top (\boldsymbol{y} - \boldsymbol{\mu})$$

$$\begin{aligned}\Delta^2 &= (\boldsymbol{y} - \boldsymbol{\mu})^\top \left(\sum_{d=1}^{D} \frac{1}{\lambda_d} \boldsymbol{u}_d \boldsymbol{u}_d^\top \right) (\boldsymbol{y} - \boldsymbol{\mu}) \\ &= \sum_{d=1}^{D} \frac{1}{\lambda_d} (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{u}_d \boldsymbol{u}_d^\top (\boldsymbol{y} - \boldsymbol{\mu}) = \sum_{d=1}^{D} \frac{z_d^2}{\lambda_d}\end{aligned}$$

3.2.2 马氏距离 (Mahalanobis distance)

二维空间下

$$\Delta^2 = \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} = \frac{(\boldsymbol{u}_1^\top(\boldsymbol{y}-\boldsymbol{\mu}))^2}{\lambda_1} + \frac{(\boldsymbol{u}_2^\top(\boldsymbol{y}-\boldsymbol{\mu}))^2}{\lambda_2}$$

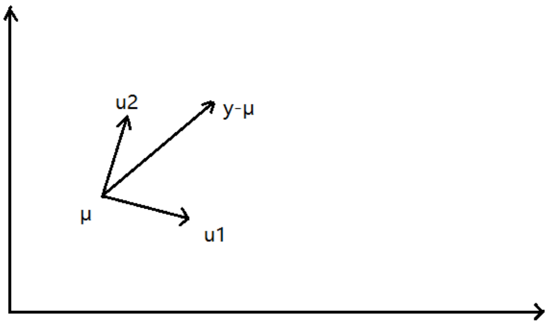


Figure: 二维下的新坐标系

3.2.3MVN 的边缘和条件概率分布

考虑一个联合高斯分布 $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)$，它的参数为

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix}$$

边缘概率密度 (marginal)

$$p(\boldsymbol{y}_1) = \mathcal{N}(\boldsymbol{y}_1 | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$$

$$p(\boldsymbol{y}_2) = \mathcal{N}(\boldsymbol{y}_2 | \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

3.2.3MVN 的边缘和条件概率分布

后验概率密度 (posterior conditional)

$$
\begin{aligned}
p(\boldsymbol{y}_1|\boldsymbol{y}_2) &= \mathcal{N}(\boldsymbol{y}_1|\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}) \\
\boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2) \\
&= \boldsymbol{\Sigma}_{1|2}(\boldsymbol{\Lambda}_{11}\boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{12}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2)) \\
\boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Lambda}_{11}^{-1}
\end{aligned}
$$

考察二元高斯分布。

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

边缘概率密度为 $p(y_1) = \mathcal{N}(y_1|\mu_1, \sigma_1^2)$

条件概率密度为 $p(y_1|y_2) = \mathcal{N}(y_1|\mu_1 + \frac{\rho\sigma_1\sigma_2}{\sigma_2^2}(y_2 - \mu_2), \sigma_1^2 - \frac{(\rho\sigma_1\sigma_2)^2}{\sigma_2^2})$

相关系数和斜率

$$\rho = 1 \iff Y - \mathbb{E}[Y] = a[X - \mathbb{E}[X]], a > 0$$

$$\mathrm{Cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[a(X - \mathbb{E}[X])^2] = a\mathbb{V}[X]$$

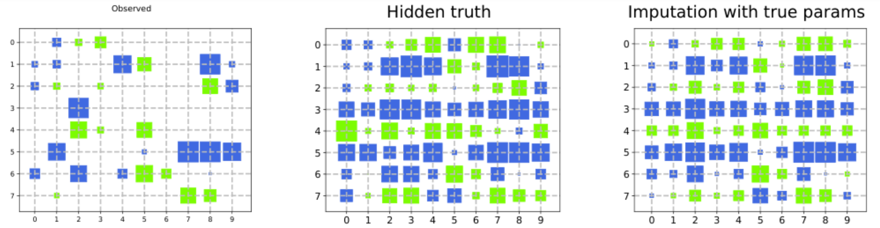
当 $\sigma_1 = \sigma_2 = \sigma$ 时，$p(y_1|y_2) = \mathcal{N}(y_1|\mu_1 + \rho(y_2 - \mu_2), \sigma^2(1 - \rho^2))$

更具体一点

假设 $\rho = 0.8$，$\sigma_1 = \sigma_2 = 1$，$\mu_1 = \mu_2 = 0$，在 $y_2 = 1$ 时，$\mathbb{E}[y_1|y_2 = 1] = 0.8$，这代表我们认为如果 y_2 相对于 μ_2 增加 1，y_1 相对于 μ_1 会随之增加 0.8。$\mathbb{V}[y_1|y_2 = 1] = 1 - 0.8^2 = 0.36$，这代表固定 y_2 之后，y_1 随之受到了限制，可以看到选择哪个 y_2 对 $\mathbb{V}[y_1|y_2]$ 没有影响。

3.2.5 实例：预测缺省值

如果我们得到了一个向量 $\boldsymbol{y}$ 的一部分 (维度) 值，我们可以利用不同维度的相互关系，推测缺省的维度值。



(a) (b) (c)

Figure: imputing missing values

$$p(\boldsymbol{y}_{n,\boldsymbol{h}}|\boldsymbol{y}_{n,\boldsymbol{v}},\boldsymbol{\theta}) \rightarrow p(y_{n,i}|\boldsymbol{y}_{n,\boldsymbol{v}},\boldsymbol{\theta}) \rightarrow \bar{y}_{n,i} = \mathbb{E}[y_{n,i}|\boldsymbol{y}_{n,\boldsymbol{v}},\boldsymbol{\theta}]$$