2016 Spring Notes

Yue Yu

February 20, 2016

Contents

1 各种相关背景知识 **2**
1.1 概率论 2
1.1.1 条件概率期望: 2
1.1.2 Correlation coefficient(相关系数) 2
1.1.3 协方差矩阵 2
1.1.4 Multivariate normal distribution 2
1.1.5 Laplace Distribution: 3
1.2 矩阵求导法则： 3

2 Information Theory **4**
2.1 Differential Entropy 4
2.1.1 常见 Differential Entropy 4
2.1.2 Properties 4

1 各种相关背景知识

1.1 概率论

1.1.1 条件概率期望:

$$E(X) = E(E(X|Y)) \tag{1}$$

1.1.2 Correlation coefficient(相关系数)

$$\begin{aligned}
\rho_{X,Y} &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} && (2)\\
&= \frac{E(X-E(X))(Y-E(Y))}{\sigma_X \sigma_Y} && (3)\\
&= \frac{E(XY)-E(X)E(Y)}{\sigma_X \sigma_Y} && (4)
\end{aligned}$$

1.1.3 协方差矩阵

$$\begin{aligned}
\mathrm{X} &= [X_1, \ldots, X_n]^T && (5)\\
\Sigma &= \mathrm{E}\Big[(\mathrm{X}-\mathrm{E}(\mathrm{X}))(\mathrm{X}-\mathrm{E}(\mathrm{X}))^T\Big] && (6)\\
\Sigma_{i,j} &= Cov(X_i, X_j) && (7)\\
&= \mathrm{E}\Big[(X_i-\mu_i)(X_j-\mu_j)\Big] && (8)
\end{aligned}$$

1.1.4 Multivariate normal distribution

- 多维高斯联合分布:

$$f_X(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

- 当为二维高斯分布时（ ρ 为相关系数）:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\Big(-\frac{1}{2(1-\rho)^2}\Big[\frac{(x-\mu_x)}{\sigma_x^2}+\frac{(y-\mu_y)}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\Big]\Big)$$

  条件概率服从:

$$P(X_1\big|X_2 = x_2) \sim N(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2), (1-\rho^2)\sigma_1^2)$$

1.1.5 Laplace Distribution:

- 概率密度函数

$$f(x|\mu,b)=\frac{1}{2b}\exp\big(-\frac{|x-\mu|}{b}\big)$$

- 期望

$$\mu$$

- 方差

$$2b^2$$

1.2 矩阵求导法则：

$$\begin{aligned}
\frac{\partial \mathbf{u^T A v}}{\partial \mathbf{x}} &= \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{Av}+\frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A^T u} && (9)\\
\frac{\partial(\mathbf{u(x)}+\mathbf{v(x)})}{\partial \mathbf{x}} &= \frac{\partial \mathbf{u(x)}}{\partial \mathbf{x}}+\frac{\partial \mathbf{v(x)}}{\partial \mathbf{x}} && (10)\\
\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} &= \mathbf{A^T} && (11)\\
\frac{\partial \mathbf{a}}{\partial \mathbf{x}} &= \mathbf{0} && (12)\\
\frac{\partial \mathbf{x^T A b}}{\partial \mathbf{x}} &= \mathbf{Ab} && (13)\\
\frac{\partial \mathbf{x^T A x}}{\partial \mathbf{x}} &= (\mathbf{A}+\mathbf{A^T})\mathbf{x} && (14)\\
&= \mathbf{2Ax} \qquad \text{如果 A 为对称阵} && (15)\\
\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} &= |\mathbf{X}|(\mathbf{X^{-1}})^{\mathbf{T}} && (16)\\
\frac{\partial \ln|\mathbf{X}|}{\partial \mathbf{X}} &= (\mathbf{X^{-1}})^{\mathbf{T}} && (17)
\end{aligned}$$

2 Information Theory

2.1 Differential Entropy

2.1.1 常见 Differential Entropy

- *Uniform distribution(from 0 to a)*
$$h(X) = \log(a)$$

- *Normal Distribution*
$$X \sim N(0, \sigma^2)$$
$$h(X) = \frac{1}{2}\log 2\pi e\sigma^2$$

- *Multivariate Normal Distribution*
$$N_n \sim (\mu, K)$$
μ is mean and K is covariance matrix
$$h(X_1, \ldots, X_n) = \frac{1}{2}\log(2\pi e)^n |K|$$

2.1.2 Properties

- $h(X + c) = h(X)$

- $h(aX) = h(X) + \log|a|$

- $h(AX) = h(X) + \log\left|\det(A)\right|$