极大似然估计与充分统计量.

$D=\{x_1, x_2, \cdots, x_N\}$

$\eta_{MLE} = \arg\max \log P(D|\eta)$

$= \arg\max \log \prod_{i=1}^{N} P(x_i|\eta)$

$= \arg\max \sum_{i=1}^{N} (\eta^T \phi(x_i) - A(\eta))$

$\frac{\partial}{\partial \eta}\left(\sum_{i=1}^{N}(\eta^T\phi(x_i) - A(\eta))\right)$

$= \sum_{i=1}^{N} \phi(x_i) - A'(\eta) = 0 \quad A'(\eta_{MLE}) = \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)$

最大熵:

信息量: $-\log P$

熵: $E_{p(x)}[-\log P] = \int -p(x)\cdot\log p(x)\,dx$

$= -\sum_x p(x)\cdot\log p(x)$

$\begin{cases} \max H[P] = \max(-\sum_{i=1}^{K} P_i \log P_i \\ s.t. \sum_{i=1}^{K} P_i = 1 \end{cases} \Rightarrow \begin{cases} \min \sum_{i=1}^{K} P_i \log P_i \\ s.t. \sum_{i=1}^{K} P_i = 1 \end{cases}$

$L(P,\lambda) = \sum_{i=1}^{K}(P_i \log P_i) + \lambda(1 - \sum_{i=1}^{K} P_i)$

$\hat{P}_i = \exp(\lambda - 1)$

$\because \hat{P}_1 = \hat{P}_2 \cdots = \hat{P}_K$.

$\therefore P(x)$ 为均匀分布

当经验分布已知:

$f(x)$ 是关于 x 的函数

$E_{\hat{P}}[f(x)] = \Delta$ (已知)

$\begin{cases} \min \sum_x P(x) \log P(x) \\ s.t. \sum_x P(x) = 1 \\ \quad E_{\hat{P}}[f(x)] = \Delta \end{cases}$

(九) 概率图.

Vazyme

高维随机变量 $P(x_1, x_2, \cdots x_p)$ $\begin{cases} \text{边缘概率}: P(x_i) \\ \text{条件概率}: P(x_j|x_i) \end{cases}$

Sum Rule: $P(x_1) = \int P(x_1, x_2)\,dx_2$

Product Rule: $P(x_1, x_2) = P(x_1)P(x_2|x_1) = P(x_2)\cdot P(x_1|x_2)$

Chain Rule: $P(x_1, x_2, \cdots x_p) = \prod_{i=1}^{p} P(x_i|x_1, x_2, x_{i-1})$

Bayesian Rule: $P(x_2|x_1) = \frac{P(x_1, x_2)}{P(x_1)} = \frac{P(x_1,x_2)}{\int p(x_1,x_2)dx_2} = \frac{P(x_2)P(x_1|x_2)}{\int p(x_2)P(x_1|x_2)dx_2}$

维度高,计算复杂

↓

简化 $\xrightarrow{\text{相互独立}}$ $P(x_1, x_2, \cdots x_p) = \prod_{i=1}^{p} P(x_i)$

↳ Naive Bayes: $P(x|y) = \prod_{i=1}^{p} P(x_i|y)$

Markov Property → $x_j \perp x_{i+1} | x_i, j < i$ (x_i 与 x_{i+1} 齐次,同分布)

条件独立性 → $x_A \perp x_B | x_C$

概率图 $\begin{cases} \text{表示} - \begin{cases} \text{有向图 Bayesian Network} \\ \text{无向图 Markov Network} \end{cases} \\ \text{推断} - \begin{cases} \text{精确推断} \\ \text{近似推断} - \begin{cases} \text{确定性近似(变分推断)} \\ \text{随机近似(MCMC)} \end{cases} \end{cases} \\ \text{学习} \quad \text{参数学习} - \text{完备数据} \end{cases}$