频率派 vs 贝叶斯派

(集合)

X : data $\longrightarrow$ $X = (x_1, x_2, \cdots x_N)^T_{N\times p}$

θ : parameter $= \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix}_{N\times p}$

$x \sim p(x|\theta)$ 服从分布

$\boxed{\text{频率派}}$: θ: 未知的常量　　X: r.v (随机变量)

估计出来　　　　　　　　更关心

累乘 → 累加 → 方便运算

$$\theta_{MLE} = \arg\max_{\theta} \underbrace{\log P(x|\theta)}_{L(\theta)}$$

(极大似然估计)

$\boxed{\text{贝叶斯派}}$: θ : r.v , $\theta \sim p(\theta)$ ← 先验

似然　　先验

$$P(\theta|X) = \frac{P(X|\theta)\cdot P(\theta)}{P(X)} \propto P(X|\theta)P(\theta)$$

后验

$$P(X) = \int_\theta P(X|\theta)P(\theta)\,d\theta$$

MAP(最大后验概率估计):

$$\theta_{MAP} = \arg\max_{\theta} P(\theta|x) = \arg\max_{\theta} P(x|\theta)\cdot P(\theta)$$

贝叶斯估计: $P(\theta|x) = \dfrac{P(x|\theta)\cdot P(\theta)}{\int_\theta P(x|\theta)P(\theta)d\theta}$

贝叶斯预测: X , $\tilde{x}$ (新数据).

$X \to \underbrace{\theta}_{\text{作为关系连接的桥梁}} \to \tilde{x}$

$$P(\tilde{x}|X) = \int_\theta P(\tilde{x},\theta|X)\,d\theta$$

$$= \int_\theta P(\tilde{x}|\theta)\underbrace{P(\theta|X)}_{\text{后验}}\,d\theta$$

< 求后验是做预测的前提 >

贝叶斯 → 概率图模型

↓

(本质) 求积分

频率派 → 统计机器学习 (本质"优化")

① 模型

② loss function

③ 梯度下降 + S 求