

# 频率派 VS 贝叶斯派

(集合)

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

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

$$x \sim P(x|\theta) \text{ 服从分布}$$

频率派:

$\theta$ : 未知的常量

$X$ : r.v (随机变量)

估计出来

更关心

累乘  $\rightarrow$  累加  $\rightarrow$  方便运算



$$\theta_{MLE} = \arg \max_{\theta} \underbrace{\log P(x|\theta)}_{L(\theta)} \quad (\text{极大似然估计})$$

贝叶斯派:

$\theta$ : r.v,  $\theta \sim P(\theta)$   $\leftarrow$  先验

偶然

必然

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

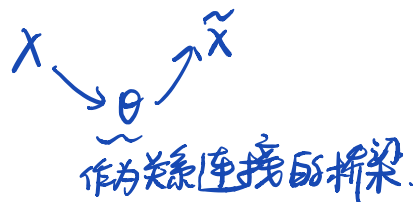
$\int_{\theta} P(x|\theta) P(\theta) d\theta$

MAP (最大后验概率估计):

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

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

贝叶斯预测:  $x, \tilde{x}$  (新数据),



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

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

< 后验是做预测的前提 >

贝叶斯  $\rightarrow$  概率图模型

$\downarrow$   
(本质) 求积分

机器学习  $\rightarrow$  统计机器学习 (本质“优化”)

# 线性回归

① 模型

② loss function ↵

③ 梯度下降+求解