频率派 vs 贝叶斯　　△默认都是

X表示数据　$X=(x_1\ x_2 \cdots x_N)^T_{N\times p}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix}_{N\times p}$

θ表示参数

$x \sim p(x|\theta)$

统计机器学习 → 优化
① 模型
② loss function
③ 优化算法

频率派：θ是未知的常量，数据是 r.v.

极大似然估计　$\hat{\theta}=\underset{\theta}{\arg\max}\ \underbrace{\log p(X|\theta)}_{L(\theta)}$　$x_i \overset{iid}{\sim} p(x|\theta)$ 独立同分布 $\Rightarrow p(X|\theta)=\prod_{i=1}^{N} p(x_i|\theta)$

$\Rightarrow \log p(X|\theta)=\sum_{i=1}^{N}\log p(x_i|\theta)$

贝叶斯派：θ, r.v.　$\theta \sim p(\theta)$ 这个称为先验

贝叶斯定理：将先验后验和似然联系起来

$p(\theta|X)=\dfrac{\overbrace{p(X|\theta)}^{似然}\cdot \overbrace{p(\theta)}^{先验}}{\underbrace{p(X)}_{积分=\int_\theta p(X|\theta)p(\theta)d\theta}} \propto p(X|\theta)\cdot p(\theta)$

$p(\theta|X)$ → 后验

引入参数估计方法 MAP：最大后验估计

MAP vs MLE：让后验概率取到最大，认为θ是服从一个分布的

在分布中找到可以让后验概率最大的点来做为估计

$\theta_{MAP}=\arg\max p(\theta|X)=\arg\max p(X|\theta)\cdot p(\theta)$

在根据θ的分布去估计后验概率时完全和　$p(X)$ 无关

严格上来讲 MAP不是标准的贝叶斯估计

↳ 就是要求 $\dfrac{p(X|\theta)\cdot p(\theta)}{\int_\theta p(X|\theta)p(\theta)d\theta}$

求这个 $p(\theta|X)$ 的目的是：贝叶斯预测

现在样本X，如果现在有了新的数据 $\tilde{x}$　$p(\tilde{x}|X)=\int_\theta p(\tilde{x},\theta|X)d\theta$

$X \to \theta \to \tilde{x}$

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

要在整个积分空间上积分，积不出。　$p(\theta|X)$：后验

⇒ 有新的计算方法

贝叶斯 → 概率图模型
↳ 求积分 → MCMC
↳ Monte Carlo Method

书

学习资料的介绍　　李航：统计学习方法

频率派 → 统计机器学习　　周志华：机器学习"西瓜书" 西瓜

贝叶斯 → 概率图模型 ← PRML：模式识别与机器学习

MLAPP：以概率的视角看机器学习

ESL：统计学习的元素

Deep learning 花书

中译版 张志华

李航：统计学习方法：感K朴决逻

支提EM隐条

PRML：模式识别与机器学习

回分神核稀　　图混近采连

顺组 boost，随机森林