白板推导1 绪论

$\begin{cases} \text{频率派} \to \text{统计机器学习} \\ \text{贝叶斯派} \to \text{概率图模型} \end{cases}$

简介：机器学习算法可以按思想不同分成概率派和贝叶斯派。概率派使用MLE估计参数，应用MLE时会使用一些优化方法；贝叶斯派对θ的分布进行建模，并将θ的后验用于预测，预测时使用积分，因此引入采样积分等方法。

X: data $\quad X = (x_1 \cdots x_N)^T_{N\times p} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Np} \end{pmatrix}_{N\times p}$

θ: parameter

假设概率模型：$x \sim p(x|\theta)$

- 频率派：θ为未知常量，x为r.v.

$x_i \overset{iid}{\sim} p(x|\theta)$

$\theta_{MLE} = \arg\max\limits_{\theta} \underbrace{\log p(X|\theta)}_{L(\theta)}$ ↖ 最大似然函数

$P(X|\theta) = \prod\limits_{i}^{N} p(x_i|\theta)$

- 贝叶斯派：θ为r.v. $\quad \theta \sim p(\theta)$

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

↑ 最大后验概率

贝叶斯定理：$p(\theta|X) = \dfrac{p(\theta)p(X|\theta)}{p(X)} \propto p(\theta)p(X|\theta)$

$p(X) = \int_\theta p(X|\theta)p(\theta)\,d\theta$（和θ无关）

贝叶斯预测：$X \to \theta \to \tilde{x}$. 求$p(\tilde{x}|X)$

已知（数据） 新样本

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

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

将X和$\tilde{x}$的关系解构成θ和X、$\tilde{x}$和θ的关系。X存在的作用为求出θ，求好θ后X失去了它的意义。

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

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

∵给定θ，X和$\tilde{x}$相互独立，即 ✓

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

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

机器学习framework：

① 设计概率模型（x的分布）。

② loss function $\begin{cases} \text{MLE} \\ \text{MAP（需作假设θ分布）。} \end{cases}$

③ 求解。

得到θ以后相当于得到完整模型

条件联合分布的分解：

$P(AB) = P(A|B)P(B)$ 条件概率公式

$P(AB|C) = P(A|BC)P(B|C)$

$= P(B|AC)P(A|C)$