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

θ: parameter

$$= \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & & & \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix}_{N\times p}.$$

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

- 频率派: θ 未知的常量, x. 随机变量.

  MLE: $\theta_{MLE} = \arg\max_{\theta} \log p(x|\theta)$.

- 贝叶斯派 x. 随机变量, $\theta \sim p(\theta)$ 先验

- $p(\theta|x) = \dfrac{p(x|\theta)p(\theta)}{p(X)}$, $p(X) \longrightarrow \int_\theta p(x|\theta)p(\theta)d\theta$.

- MAP $= \arg\max_{\theta} p(\theta|x) = \arg\max_{\theta} p(x|\theta)\cdot p(\theta)$.

- 贝叶斯预测: X $\tilde{x}$

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

- 贝叶斯 $\longrightarrow$ 概率图问题.

  ↳ 求积分. $\longrightarrow$ MCMC.

频率派 $\longrightarrow$ 统计机器学习

↓

优化.

① 模型

② loss function

③ 优化算法.

- 频率派：
  - 把 model 的参数看成未知的常数，通过训练data和准则来估计参数
  - 最大似然估计
  - 是一个优化问题
- 贝叶斯 model
  - 参数有先验 $P(\theta)$
  - 有似然 $P(x|\theta)$
  - 求后验 $P(\theta|x)$
  - $$P(\theta|x) = \frac{P(\theta)\cdot P(x|\theta)}{\underset{\substack{\uparrow \\ \text{通过蒙特卡洛去抽样}}}{P(x)}}$$ 贝叶斯

- 两者区别

$$L(y, f(x)) + \lambda P(\omega) \quad \text{利用优化}$$

$$\underbrace{\exp(-L(y, f(x)))}_{\text{似然}} \cdot \underbrace{\exp(\lambda P(\omega))}_{\text{先验}}$$

这样是等价的

但若用抽样方法确定 $P(x)$，则不同