最大似然估计　　似然：likelihood

$X = \{x_1, x_2, x_3 \cdots x_n\}$

独立同分布假设

X出现的联合概率（似然函数）

$l(\theta) = p(X|\theta) = p(x_1, x_2 \ldots x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$

$p(x_i|\theta)$：θ相对于每一个样本的似然函数

使 $l(\theta)$ 值最大的 $\hat{\theta} = d(x_1, x_2 \ldots x_n)$ 即 θ 的最大似然估计量，$\hat{\theta} \in \Theta$

$\hat{\theta} = \arg\max_{\theta} l(\theta)$

对数似然函数 $H(\theta) = \ln l(\theta)$

最大似然估计求解

$$\frac{dH(\theta)}{d\theta} = \sum_{i=1}^{n} \frac{d}{d\theta} \ln p(x_i|\theta) = 0$$

$$\vec{\theta} = [\theta_1, \theta_2 \cdots \theta_s]^T$$

$$\nabla_{\vec{\theta}} H(\vec{\theta}) = \sum_{i=1}^{n} \nabla_{\vec{\theta}} \ln P(x_i|\vec{\theta}) = 0$$

$$\nabla_{\vec{\theta}} = [\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, \cdots \frac{\partial}{\partial\theta_s}]^T$$

对于不可导的 $p(x_i|\theta)$ 则采用观察法、迭图法求 $\hat{\theta} = \arg\max_{\theta} l(\theta)$

正态分布下的最大似然估计

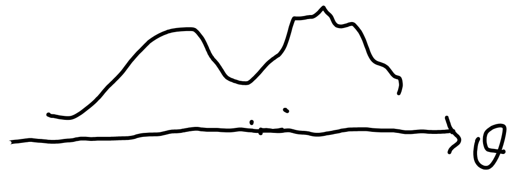
*（3-17）~（3-23）

对于多元正态分布 $x \sim N(\vec{\mu}, \Sigma)$

$$\frac{\partial (x-\mu)^T \cdot \Sigma^{-1} (x-\mu)}{\partial \mu} = \frac{\partial R^{1\times n} \cdot R^{n\times n} \cdot R^{n\times 1}}{\partial \mu} = R^{1\times 1}$$

分子先展开，~~整体~~再合并成向量表达式

贝叶斯估计



$p(x|\theta)$ 中的 θ 看作一个随机变量 $p(\theta)$.

θ 不是一个固定的值、

MLE 中的 θ 是一个固定的值.

我们目标是一样的求 θ^*

$$R = \int_E d \int_\Theta \lambda(\hat{\theta}, \theta) p(x, \theta) \, d\theta \, dx$$

很像

$$R = \sum_i \int_{R_i} \lambda(\alpha_i, \omega_i) \, p(x, \theta) \, d\theta \, dx$$

(你看书, 书上不会告诉你知识点之间的联系, 也不是所有老师都告诉你, 这就是视频的重要性)

$$R = \int_E d \int_\Theta \lambda(\hat{\theta}, \theta) p(x, \theta) \, d\theta \, dx$$

$$R(\hat{\theta}|x) = \int_\Theta \lambda(\hat{\theta}, \theta) \, p(\theta|x) \, d\theta$$

$$\theta^* = \arg\min_\theta R(\hat{\theta}|x)$$

决策时需要定义损失表, 连续情况下需定义损失函数

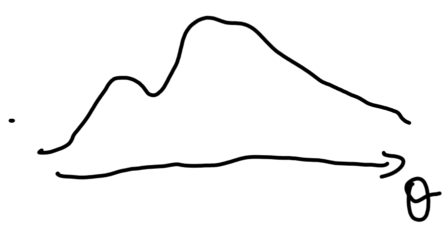
$$L(\hat{\theta},\theta) = (\theta-\hat{\theta})^2$$

$$R(\hat{\theta}|x) = \int_\theta (\hat{\theta}^2+\theta^2-2\theta\hat{\theta})\cdot p(\theta|x)\,d\theta$$

$$\frac{\partial R}{\partial \hat{\theta}} = \underbrace{\int_\theta p(\theta|x)\,d\theta}_{\text{对}\hat{\theta}\text{是个constant}} \cdot (2\hat{\theta}-2\theta) = 0$$

$$\theta^* = \frac{\int \theta\, p(\theta|x)\,d\theta}{\int p(\theta|x)\,d\theta} = \int \theta\, p(\theta|x)\,d\theta$$

也就是 $\theta \sim p(\theta|x)$ 在不同概率下加权求和即期望



$$p(x|X) = \int_\theta p(x|\theta)\,p(\theta|X)\,d\theta$$ 为什么呢?

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

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

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

$$p(\theta|X) \sim p(X|\theta)\cdot p(\theta)$$