Ch2 Probability Distribution

常用概率分布

共轭分布(conjugate prior)：
先验分布 $q(\theta)$ 是分布 $p(x|\theta)$ 的共轭分布 iff 后验分布 $p(\theta) \propto p(x|\theta)q(\theta)$ 和 $q(\theta)$ 是同一概率分布

	PDF	说明
Bernoulli	$Bern(x\|\mu) = \mu^x(1-\mu)^{1-x}$	一次硬币
Binomial	$Bin(n\|N, \mu) = C(N,\ n)\mu^n(1-\mu)^{N-n}$	多次硬币(二项分布)
Beta	$Beta(\mu\|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ \mu^{a-1}(1-\mu)^{b-1}$	conj. prior for Bern
Multinomial	$Multi(x\|\mu) = \prod_{k=1}^{K}\mu_k^{x_k}$	多项分布
Dirichlet	$Dir(\mu\|\alpha) = \frac{\Gamma(\alpha_1+\alpha_2+...\alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2)...\Gamma(\alpha_K)}\prod_{k=1}^{K}\mu_k^{\alpha_k-1}$	conj prior for multi
Gaussian	$N(x\|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}\|\Sigma\|^{1/2}}exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)^T\}$	
Gamma	$Gam(\lambda\|a, b) = \frac{1}{\Gamma(a)}b^a\lambda^{a-1}exp(-b\lambda)$	conj prior for 1D Gaussian
Wishart	$W(\Lambda\|M, \nu) \propto \|\Lambda\|^{(\nu-D-1)/2}exp\{-\frac{1}{2}Tr(M^{-1}\Lambda)\}$	conj prior for N-D Gaussian
Student-t	$St(x\|\mu, \Lambda, \nu) \propto [1+\frac{(x-\mu)^T\Lambda(x-\mu)}{\nu}]^{-\nu/2-D/2}$	假设Gaussian的precision matrix是随机变量, 基于Wishart先验的两个参数的条件概率
Mixture of Gaussian	$MoG(x) = \sum_{k=1}^{K}\pi_k N(x\|\mu_k, \Sigma_k)$	

指数族

定义：满足 $p(x|\eta) = f(x)g(\eta)exp\{-\eta^T u(x)\}$ 的概率分布 $p(x|\eta)$
例子：上面的概率分布除了MoG别的都是指数族分布
最大似然: $-\nabla ln\ g(\mu_{ML}) = \frac{1}{N}\sum_{n=1}^{N}u(x_n)$
共轭分布: $p(\eta|\chi, \nu) = \frac{1}{Z(\chi,\nu)}g(\eta)^\nu exp\{\nu\eta^T\chi\}$

多元Gaussian的性质

联合分布 $N(x|\mu, \Sigma)$, 其中 $\Lambda = \Sigma^{-1}$, $x = (x_a, x_b),\ \mu = (\mu_a, \mu_b),\ \Sigma = (\Sigma_{aa}, \Sigma_{ab}; \Sigma_{ab}, \Sigma_{ab}),\ \ \Lambda = (\Lambda_{aa}, \Lambda_{ab}; \Lambda_{ab}, \Lambda_{ab})$

条件分布 $p(x_a|x_b) = N(\mu_a - \Lambda_{aa}^{-1}\Lambda_{ab}(x_b-\mu_b),\ \Sigma_{aa}^{-1})$
边际分布 $p(x_a) = N(x_a|\mu_a, \Sigma_{aa})$
Bayesian rule 定义Gaussian 边际概率 $p(x) = N(x|\mu, \Lambda^{-1})$ 和条件概率 $p(y|x) = N(y|Ax+b,\ L^{-1})$
则有

$$p(y) = N(y|A\mu + b,\ L^{-1} + A\Lambda^{-1}A^T)$$
$$p(x|y) = N(x|S\{A^T L(y-b) + \Lambda\mu\},\ S),\ S = (\Lambda + A^T L A)^{-1}$$

最大似然估计 $\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$, $\Sigma_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})(x_n - \mu_{ML})^T$

注意其中Σ_{ML}的数学期望是$\frac{N-1}{N}\Sigma$

Full Bayesian估计 考虑正态分布$p(x) = N(\mu, \sigma^2)$,

μ 的先验概率$p(\mu) = N(\mu|\mu_0, \sigma_0^2)$

后验概率$p(\mu|X) = N(\alpha\mu_0 + (1-\alpha)\mu_{ML})$, 其中 $\alpha = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}$

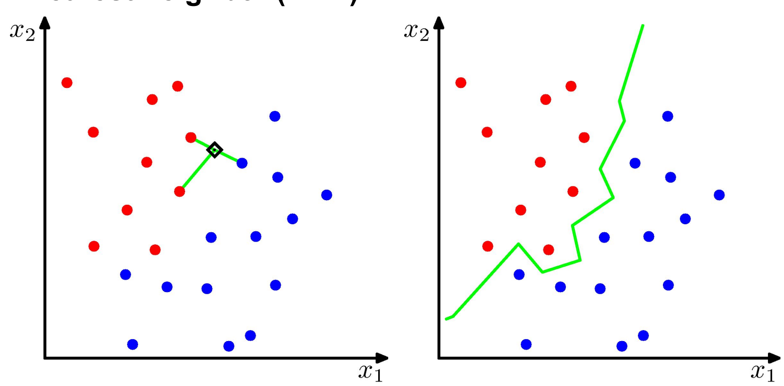
$\lambda = \sigma^{-2}$ 的先验概率$p(\lambda|a, b) = Gam(\lambda|a, b) \propto \lambda^{a-1} exp(-b\lambda)$

后验概率$p(\lambda|X) = \lambda^{a_0 - 1 + N/2} exp\{-b_0\lambda - \frac{\lambda}{2}\sum_{n=1}^{N}(x_n - \mu)^2\} = Gam(\lambda|a_0 + \frac{N}{2},\ b_0 + \frac{N}{2}\sigma_{ML}^2)$

非参估计

parzen-window $p(x) = \frac{1}{Z}\sum_{n=1}^{N} k(\frac{x - x_n}{h})$, 其中$k(\cdot)$是kernel function

K nearest neighbor (KNN)



(a) (b)