线性动态系统 (Linear Dynamical Systems)

① 背景

② 推断

③ 学习

背景（Background）

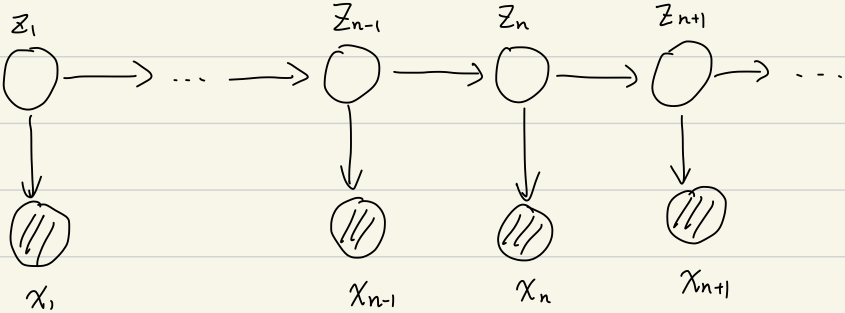
HMM：潜在变量是离散的

Linear Dynamical Systems：潜在变量连续且满足线性与高斯

Particle Filters：连续、非线性、非高斯

我们将 LDS 看作是连续潜在变量模型（例如 P-PCA 和因子分析）的推广

线性动态系统（LDS）：



$$p(z_1) = N(z_1 \mid \mu_0, P_0)$$

$$p(z_n \mid z_{n-1}) = N(z_n \mid A z_{n-1}, Q)$$

$$p(x_n \mid z_n) = N(x_n \mid C z_n, R)$$

这里省略了可加性常数，如果加上，结果为

$$N(z_n \mid A z_{n-1} + B, Q)$$

$$N(x_n \mid C z_n + D, R)$$

$$z_1 = \mu_0 + u, \quad u \sim N(u \mid 0, P_0)$$

$$z_n = A z_{n-1} + w_n, \quad w \sim N(w \mid 0, Q)$$

$$x_n = C z_n + v_n, \quad v \sim N(v \mid 0, R)$$

$$z_n = A z_{n-1} + B + w_n$$

$$x_n = C z_n + D + v_n$$

由于 LDS 是一个线性高斯模型，因此在所有变量上的联合概率分布以及边缘概率分布和条件概率分布都是高斯分布。

推断（Inference）

在LDS中的推断一般就是filter和smooth两个方面，
分别称之为Kalman filter方程和Kalman smoother方程。
我们可以利用之前推导得到的结果来解这两个问题。

Kalman Filter → 重要的应用是跟踪。

记 $\hat{\alpha}(z_n) = p(z_n | x_1, \cdots, x_n) = N(z_n | \mu_n, V_n)$

则递推式为： $c_n \hat{\alpha}(z_n) = p(x_n | z_n) \int \hat{\alpha}(z_{n-1}) p(z_n | z_{n-1}) dz_{n-1}$

$$c_n N(z_n | \mu_n, V_n) = N(x_n | C z_n, R) \int N(z_{n-1} | \mu_{n-1}, V_{n-1}) N(z_n | A z_{n-1}, Q) dz_{n-1}$$

公式：给定边缘高斯分布和条件高斯分布：

$$p(x) = N(x | \mu, \Lambda^{-1}) \quad ①$$

$$p(y|x) = N(y | Ax + b, L^{-1}) \quad ②$$

则y的边缘高斯分布和x的条件高斯分布为

$$p(y) = N(y | A\mu + b, L^{-1} + A\Lambda^{-1}A^T) \quad ③$$

$$p(x|y) = N(x | \Sigma[A^T L(y-b) + \Lambda\mu], \Sigma), \quad \Sigma = (\Lambda + A^T L A)^{-1} \quad ④$$

对于积分号，我们可以看成给定边缘与条件概率，求边缘概率，应用上述公式：

$$\int N(z_{n-1} | \mu_{n-1}, V_{n-1}) N(z_n | A z_{n-1}, Q) dz_{n-1} = N(z_n | A\mu_{n-1}, P_{n-1}), \quad P_{n-1} = Q + A V_{n-1} A^T$$

则 $c_n N(z_n | \mu_n, V_n) = N(x_n | C z_n, R) N(z_n | A\mu_{n-1}, P_{n-1})$

再次应用公式，得：

将 x_n 看作 y，z_n 看作 x，如下：

$$\underbrace{p(x_n|x_1,\cdots,x_{n-1})}_{p(y)}\underbrace{p(z_n|x_1,\cdots,x_n)}_{p(x|y)}=\underbrace{p(x_n|z_n)}_{p(y|x)}\underbrace{p(z_n|x_1,\cdots,x_{n-1})}_{p(x)}$$

$$\mu_n=V_n(C^TR^{-1}x_n+P_{n-1}^{-1}A\mu_{n-1})$$

$$V_n=(P_{n-1}^{-1}+C^TR^{-1}C)^{-1}$$

$$C_n=N(x_n|CA\mu_{n-1},CP_{n-1}C^T+R)$$

公式：

$$(P^{-1}+B^TR^{-1}B)^{-1}B^TR^{-1}=PB^T(BPB^T+R)^{-1} \quad ⑤$$

$$(A+BD^{-1}C)^{-1}=A^{-1}-A^{-1}B(D+CA^{-1}B)^{-1}CA^{-1} \quad ⑥$$

对 V_n 应用公式⑥，得：$V_n=P_{n-1}-P_{n-1}C^T(R+CP_{n-1}C^T)^{-1}CP_{n-1}$

令 $K_n=P_{n-1}C^T(R+CP_{n-1}C^T)^{-1}$ → Kalman增益矩阵(Kalman gain matrix)

得 $V_n=(I-K_nC)P_{n-1}$

对 K_n 应用公式⑤．得：$K_n=(C^TR^{-1}C+P_{n-1}^{-1})^{-1}C^TR^{-1}=V_nC^TR^{-1}$

则 $V_n=K_nR(C^T)^{-1}$，代入 μ_n.

得 $\mu_n=K_nR(C^T)^{-1}C^TR^{-1}x_n+(I-K_nC)P_{n-1}P_{n-1}^{-1}A\mu_{n-1}$

$$=A\mu_{n-1}+K_n(x_n-CA\mu_{n-1}).$$

最终
$$\begin{cases}\mu_n=A\mu_{n-1}+K_n(x_n-CA\mu_{n-1}), \quad K_n=P_{n-1}C^T(R+CP_{n-1}C^T)^{-1}\\ V_n=(I-K_nC)P_{n-1}\\ C_n=N(x_n|CA\mu_{n-1},CP_{n-1}C^T+R)\end{cases}$$

初始条件：$c_1\hat{\alpha}(z_1) = p(z_1)p(x_1|z_1)$

其中. $p(z_1) = N(z_1|\mu_0, P_0)$

$p(x_1|z_1) = N(x_1|Cz_1, R)$

再次应用公式①，②，③，④，⑤，⑥，得：

$$\boxed{\begin{aligned} \mu_1 &= \mu_0 + K_1(x_1 - C\mu_0) \\ V_1 &= (I - K_1C)P_0 \\ c_1 &= N(x_1|C\mu_0, CP_0C^T + R) \\ K_1 &= P_0C^T(CP_0C^T + R)^{-1} \end{aligned}}$$

初始条件.

因此，给定上一时刻的 μ_{n-1}，V_{n-1} 和当前时刻的观测 x_n，就可以计算当前时刻 z_n 的均值 μ_n，方差 V_n 以及归一化系数 c_n

LDS 的似然函数可以由 $p(X) = \prod_{n=1}^{N} c_n$ 求得.

Kalman Filter 的直观感受：先用一个转移概率矩阵 A 对上一时刻的 μ_{n-1} 进行投影，得到最初的 μ_n 的预测 $A\mu_{n-1}$，这个预测也会给出 x_n 的预测 $CA\mu_{n-1}$。当获得当前时刻的 x_n 后，将 $x_n - CA\mu_{n-1}$ 作为修正项，乘上一个修正系数（K_n），修正对 μ_n 的预测。总的来说，我们先对后续进行预测，然后用后续的观测值来修正预测

Kalman Filter的工作流程：（$t=n$代表接收到x_n，工作流程为先update，再predict）

$t=1$: $\begin{cases} p(z_1|x_1) & \hat{\alpha}(z_1) & \text{update} \\ p(z_2|x_1) & A\mu_1 & \\ p(x_2|x_1) & c_2 & \end{cases}$ predict

$\Downarrow$

$t=2$: $\begin{cases} p(z_2|x_1,x_2) & \hat{\alpha}(z_2) & \text{update} \\ p(z_3|x_1,x_2) & A\mu_2 & \\ p(x_3|x_1,x_2) & c_3 & \end{cases}$ predict

$\Downarrow$

$\vdots$

$\Downarrow$

先用update修正上一次的预测

再用predict对下一时刻进行预测

$t=n$: $\begin{cases} p(z_n|x_1,\cdots,x_n) & \hat{\alpha}(z_n) & \text{update} \\ p(z_{n+1}|x_1,\cdots,x_n) & A\mu_n & \\ p(x_{n+1}|x_1,\cdots,x_n) & c_{n+1} & \end{cases}$ predict

$\Downarrow$

$\vdots$

Kalman Smoother

解决的问题：给定$x_1,\cdots,x_N$，寻找z_n的边缘概率，即$p(z_n|x_1,\cdots,x_N)$

类比HMM模型，$p(z_n|x_1,\cdots,x_N)=\gamma(z_n)=\hat{\alpha}(z_n)\hat{\beta}(z_n)=\mathcal{N}(z_n|\hat{\mu}_n,\hat{V}_n)$

借用$\hat{\beta}(z_n)$的更新公式：$c_{n+1}\hat{\beta}(z_n)=\int\hat{\beta}(z_{n+1})p(x_{n+1}|z_{n+1})p(z_{n+1}|z_n)dz_{n+1}$

两侧乘上 $\hat{\alpha}(z_n)$ 后，经过计算，得：

https://www.bilibili.com/video/BV1eb4y167wF/?spm_id_from=333.999.0.0&vd_source=e3780c93bbfab1295672c1a3f1be54d5

具体过程不写了，可以参考b站视频.

$$\hat{\mu}_n = \mu_n + J_n(\hat{\mu}_{n+1} + A\mu_n)$$

$$\hat{V}_n = V_n + J_n(\hat{V}_{n+1} - P_n)J_n^T$$

$$J_n = V_n A^T (P_n)^{-1}$$

从而，可以由 $\gamma(z_{n+1})$ 向前递归得到 $\gamma(z_n)$

注意：该递归方程要求先完成前向过程，得到 μ_n 和 V_n.

此外，为了下一节的学习，我们还需要求 $\xi(z_{n-1}, z_n)$

$$\xi(z_{n-1}, z_n) = (C_n)^{-1}\hat{\alpha}(z_{n-1})p(x_n|z_n)p(z_n|z_{n-1})\hat{\beta}(z_n) \quad \rightarrow \frac{\gamma(z_n)}{\hat{\alpha}(z_n)}$$

$$= \frac{N(z_{n-1}|\mu_{n-1}, V_{n-1})N(x_n|Cz_n, R)N(z_n|Az_{n-1}, Q)N(z_n|\hat{\mu}_n, \hat{V}_n)}{N(x_n|CA\mu_{n-1}, CP_{n-1}C^T + R)N(z_n|\mu_n, V_n)}$$

整理后，得 $\xi(z_{n-1}, z_n)$ 均值为 $(\hat{\mu}_{n-1}, \hat{\mu}_n)^T$，协方差为

$$\text{cov}[z_{n-1}, z_n] = J_{n-1}\hat{V}_n$$

学习(Learning)

使用EM算法来求解LDS参数的极大似然估计问题，即LDS的学习。

根据EM算法，我们要在$p(Z|X,\theta^{(t)})$这个分布上求期望，
通过前面的计算，我们可以得到：

$$E[z_n] = \hat{\mu}_n$$

$$E[z_n z_{n-1}^T] = \hat{V}_n J_{n-1}^T + \hat{\mu}_n \hat{\mu}_{n-1}^T$$

$$E[z_n z_n^T] = \hat{V}_n + \hat{\mu}_n \hat{\mu}_n^T$$

对数似然函数：

$$\log p(X,Z|\theta) = \log p(z_1|\mu_0,P_0) + \sum_{n=2}^{N} \log p(z_n|z_{n-1},A,Q) + \sum_{n=1}^{N} p(x_n|z_n,C,R)$$

EM的Q函数为：

$$Q(\theta,\theta^{(t)}) = E_{Z\sim p(Z|X,\theta^{(t)})}[\log p(X,Z|\theta)]$$

当我们只考虑μ_0和P_0时，代入$p(z_1|\mu_0,P_0) = N(z_1|\mu_0,P_0)$

$$Q(\theta,\theta^{(t)}) = -\frac{1}{2}\log|P_0| - E[\frac{1}{2}(z_1-\mu_0)^T P_0^{-1}(z_1-\mu_0)] + C_1$$

所有不依赖μ_0和P_0的项全部放入C_1中。

则关于μ_0和P_0最大化$Q(\theta,\theta^{(t)})$，就可以得到：

$$\mu_0^{(t+1)} = E[z_1]$$

$$V_0^{(t+1)} = E[z_1 z_1^T] - E[z_1]E[z_1^T]$$

当我们只考虑A和Q时，代入 $p(z_n|z_{n-1},A,Q)=\mathcal{N}(z_n|Az_{n-1},Q)$

$$Q(\theta,\theta^{(t)}) = -\frac{N-1}{2}\log|Q| - E\left[\frac{1}{2}\sum_{n=2}^{N}(z_n-Az_{n-1})^TQ^{-1}(z_n-Az_{n-1})\right]+C_2$$

所有不依赖A和Q的项全部放入C_2中.

解得:

$$A^{(t+1)} = \left(\sum_{n=2}^{N}E[z_nz_{n-1}^T]\right)\left(\sum_{n=2}^{N}E[z_{n-1}z_{n-1}^T]\right)^{-1}$$

$$Q^{(t+1)} = \frac{1}{N-1}\sum_{n=2}^{N}\Big\{E[z_nz_n^T]-A^{(t+1)}E[z_{n-1}z_n^T] - E[z_nz_{n-1}^T](A^{(t+1)})^T + A^{(t+1)}E[z_{n-1}z_{n-1}^T](A^{(t+1)})^T\Big\}$$

当我们只考虑C和R时，代入 $p(x_n|z_n,C,R)=\mathcal{N}(x_n|Cz_n,R)$

$$Q(\theta,\theta^{(t)}) = -\frac{N}{2}\log|R| - E\left[\frac{1}{2}\sum_{n=1}^{N}(x_n-Cz_n)^TR(x_n-Cz_n)\right]+C_3$$

所有不依赖C和R的项全部放入C_3中

解得:

$$C^{(t+1)} = \left(\sum_{n=1}^{N}x_nE[z_n^T]\right)\left(\sum_{n=1}^{N}E[z_nz_n^T]\right)^{-1}$$

$$R^{(t+1)} = \frac{1}{N}\sum_{n=1}^{N}\Big\{x_nx_n^T - C^{(t+1)}E[z_n]x_n^T - x_nE[z_n^T](C^{(t+1)})^T + C^{(t+1)}E[z_nz_n^T](C^{(t+1)})^T\Big\}$$

上面的结果直接出自PRML，我没有推导过，可以参考PRML中第十三章的内容

或者b站视频：https://www.bilibili.com/video/BV1a64y1z7y5/?spm_id_from=333.999.0.0&vd_source=e3780c93bbfab1295672c1a3f1be54d5

有兴趣可以看一下PRML中的LDS的推导