Machine Learning
Gaussian models

华中科技大学计算机学院
王天江

高斯模型

第3章：高斯模型

- 概述
- 高斯判别式分析
- 用高斯联合分布进行推理
- 线性高斯系统
- Inferring the parameters of an MVN

概述

高斯分布的重要性

- 多元高斯分布（正态分布 MVN）
  - 是最广泛使用的联合概率密度分布
  - 它用于连续随机变量
  - 它是构成许多模型的基础

多元高斯分布

$$N(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2} \mid \boldsymbol{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

- **x** = [x_1, . . . , x_D] **:** ***随机向量，每个分量都是随机变量***
- **μ**：均值向量
- **Σ**: 协方差矩阵

马氏距离（Mahalanobis distance）

- 高斯分布的指数部分，具有距离特征
  - **我们把它称为马氏距离:**

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\textstyle\sum^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

  - **马式距离具有一定程度的尺度不变性，数据经过缩放，距离值也不会改变。**
- 常用的欧氏距离：
  - **欧式距离会受尺度的影响**

$$(\boldsymbol{x}-\boldsymbol{\mu})^T\,(\boldsymbol{x}-\boldsymbol{\mu})$$

马氏距离分析

- 协方差矩阵可以分解成：**$\Sigma = U\Lambda U^T$,**
  - **U ：特征向量组成的正交矩阵，$U^TU = I$,**
    - ⑩ **u_i ： *U的第 i列，是第i个特征向量***
  - **Λ ：特征值λ_i组成的对角矩阵.**

$$\boldsymbol{\Sigma}^{-1} = \mathbf{U}^{-T}\Lambda^{-1}\mathbf{U}^{-1} = \mathbf{U}\Lambda^{-1}\mathbf{U}^{T} = \sum_{i=1}^{D}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$$

- 所以，马氏距离可变成：

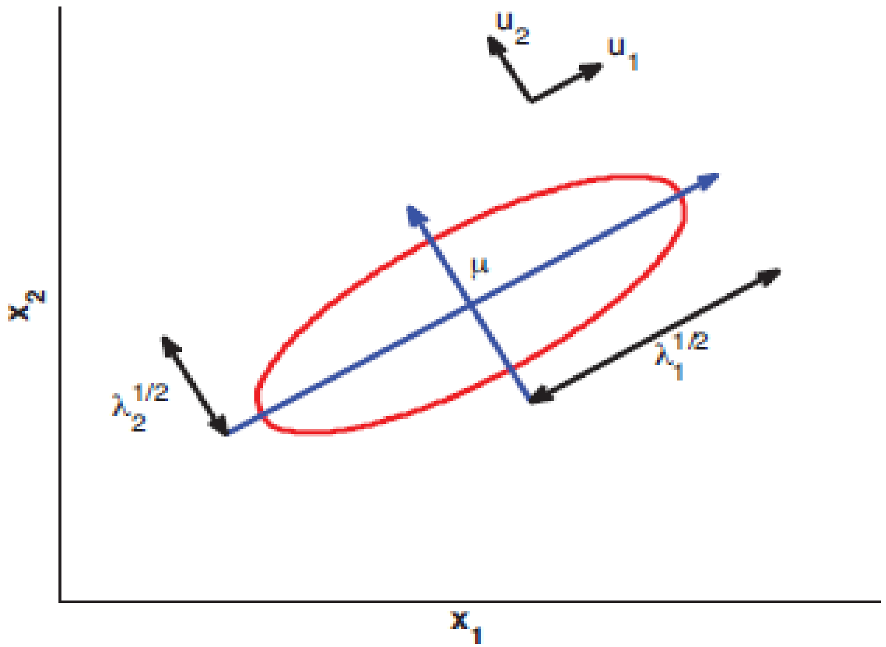
$$(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = (\mathbf{x}-\boldsymbol{\mu})^T\left(\sum_{i=1}^{D}\frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T\right)(\mathbf{x}-\boldsymbol{\mu}) \qquad \textit{where}\ \ y_i = \mathbf{u}_i^T(\mathbf{x}-\boldsymbol{\mu})$$

$$= \sum_{i=1}^{D}\frac{1}{\lambda_i}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{u}_i\mathbf{u}_i^T(\mathbf{x}-\boldsymbol{\mu}) = \sum_{i=1}^{D}\frac{y_i^2}{\lambda_i}$$

多元高斯分布的特点

- 不失一般性，考虑2维空间，马氏距离可写成：

$$(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} \qquad where \;\; y_i = \mathbf{u}_i^T(\mathbf{x}-\boldsymbol{\mu})$$



- 沿椭圆轮廓都是等概率密度
- 特征向量确定椭圆的方向
- 特征值决定了它的伸长程度

基于最大似然，估计多元高斯分布的参数

- 定理：N个独立同分布的样本$\mathbf{x_i} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, 则最大似然估计的参数结果为

$$\hat{\boldsymbol{\mu}}_{mle} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i = \bar{\mathbf{x}}$$

$$\hat{\boldsymbol{\Sigma}}_{mle} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{1}{N}(\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T) - \bar{\mathbf{x}}\bar{\mathbf{x}}^T$$

- 在单变量情形下，变成我们熟悉的形式:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}x_i = \bar{x} \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \left(\frac{1}{N}\sum_{i=1}^{N}x_i^2\right) - (\bar{x})^2$$

高斯分布具有最大熵

- 在给定了均值和方差的所有分布中，高斯分布的熵最大
- 为了简化符号，假设均值为0
- 定理：**设**$q(x)$ 是任意密度分布，满足 $\int q(\mathbf{x})x_i x_j = \boldsymbol{\Sigma}_{ij}$ 。

  如果 $p = N(0,\Sigma)$，则 $h(q) \le h(p)$
- 高斯分布的熵：

$$h(N(\mu,\Sigma)) = \frac{1}{2}\ln[(2\pi e)^D|\Sigma|]$$

高斯判别分析(GDA)

产生式分类器回顾

- 我们知道产生式分类器：

$$p(y=c|x,\theta)=\frac{p(y=c|\theta)p(x|y=c,\theta)}{\sum_{c'}p(y=c'|\theta)p(x|y=c',\theta)}$$

- 我们可以用下面规则进行决策：

$$\hat{y}(\mathbf{x})=\arg\max_{c}[\log p(y=c\mid\pi_{\mathrm{c}})+\log p(x\mid\theta_c)]$$

高斯判别分析

- 假设类条件密度服从高斯分布： $$p(\mathbf{x} \mid y=c, \theta) = N(x \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$
- 后验决策规则就变成： $$\hat{y}(\mathbf{x}) = \arg\max_c [\log \pi_c + \log N(x \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)]$$
  - 用这个式子进行推理，称为（高斯）判别式分析 **(GDA)**
  - 尽管它是产生式而不是判别式分类器
  - 如果 Σ_c 是对角阵，则等价于朴素贝叶斯.
- 为什么称它为判别式分类器呢？

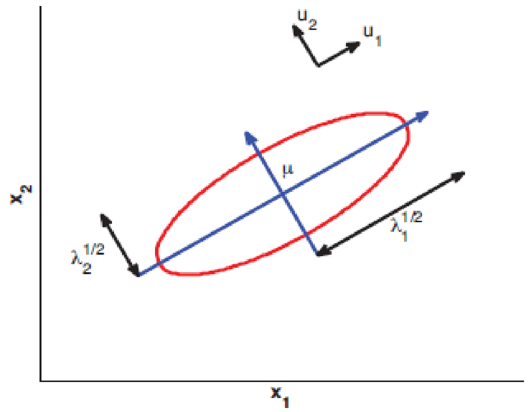
产生式模型被称为判别式模型的原因

- 决策规则:

$$\hat{y}(\mathbf{x}) = \arg\max_{c}[\log \pi_c + \log N(x \mid \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)]$$

$$\hat{y}(\mathbf{x}) = \arg\max_{c}[\log \pi_c - \frac{1}{2}\log|2\pi\Sigma_c| - \frac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)]$$

  - 决策规则就变成了多项式



- 因此，才将它称为判别式模型
  - 如果决策规则是二次多项式，称为二次判别模型
  - 如果决策规则是一次多项式，称为线性判别模型

最近中心分类器

- 假定类先验为均匀分布，分类器就从:

$$\hat{y}(\mathbf{x}) = \arg\max_{c}[\log \pi_c - \frac{1}{2}\log|2\pi\Sigma_c| - \frac{1}{2}(x-\mu_c)^T\Sigma_c^{-1}(x-\mu_c)]$$

- 变成为:

$$\hat{y}(x) = \underset{c}{\operatorname{argmin}}[log|2\pi\Sigma_c| + (x-\mu_c)^T\textstyle\sum_c^{-1}(x-\mu_c)]$$

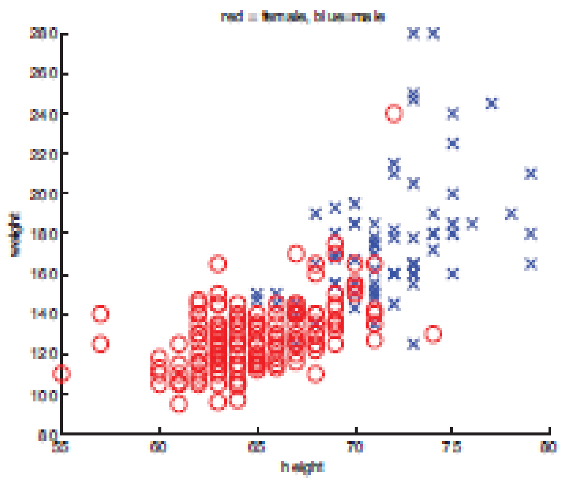
其中，$log|2\pi\Sigma_c|$ 描述类别的散布程度，即该类别的数据点分散程度

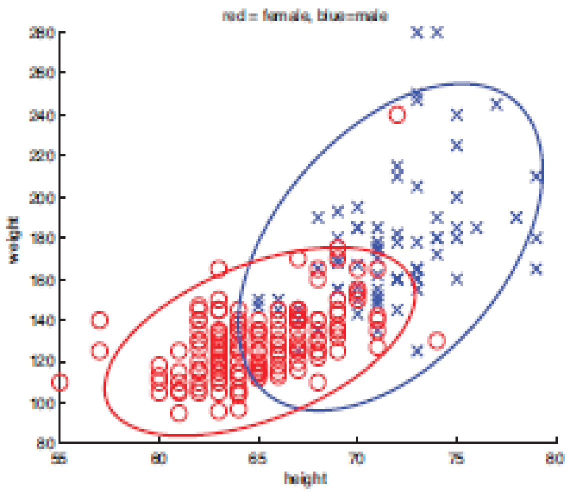
$(x-\mu_c)^T\sum_c^{-1}(x-\mu_c)$ 描述样本点离中心点的马氏距离

- 如果不考虑散布情形，则变成为: $\hat{y}(x) = \underset{c}{\operatorname{argmin}}[(x-\mu_c)^T\sum_c^{-1}(x-\mu_c)]$

最近中心分类器应用

- 样本数据取自人群的身高体重
- 每个类的2d高斯拟合图示
  - 概率分布的95% 都在椭圆的内部.





二次判别分析(QDA)

- 对于类c1和c2，如果x属于c1，则应有： $p(y=c_1\mid x)>p(y=c_2\mid x)$
- 令，

$$\delta(x)=p(y=c_1\mid x)-p(y=c_2\mid x)$$

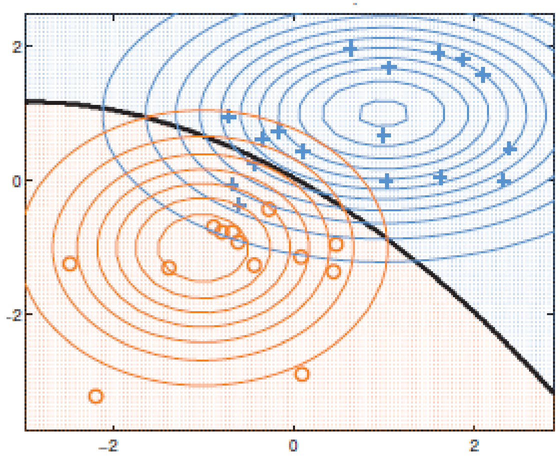
$$=-\frac{1}{2}(x-\mu_{c1})^T\Sigma_{c1}^{-1}(x-\mu_{c1})-\frac{1}{2}\log|\Sigma_{c1}|$$

$$+\frac{1}{2}(x-\mu_{c2})^T\Sigma_{c2}^{-1}(x-\mu_{c2})+\frac{1}{2}\log|\Sigma_{c2}|+\log\frac{\pi_{c1}}{\pi_{c2}}$$

- δ(x) 成为了判别函数，因为是二次多项式，特称为二次判别分析
  - δ(x)>0 则 x∈c1,
  - δ(x)<0 则 x∈c2

二次判别分析的边界

- 二次判别分析中，判别函数δ(x) =0，表示x在c1和c2类的边界上

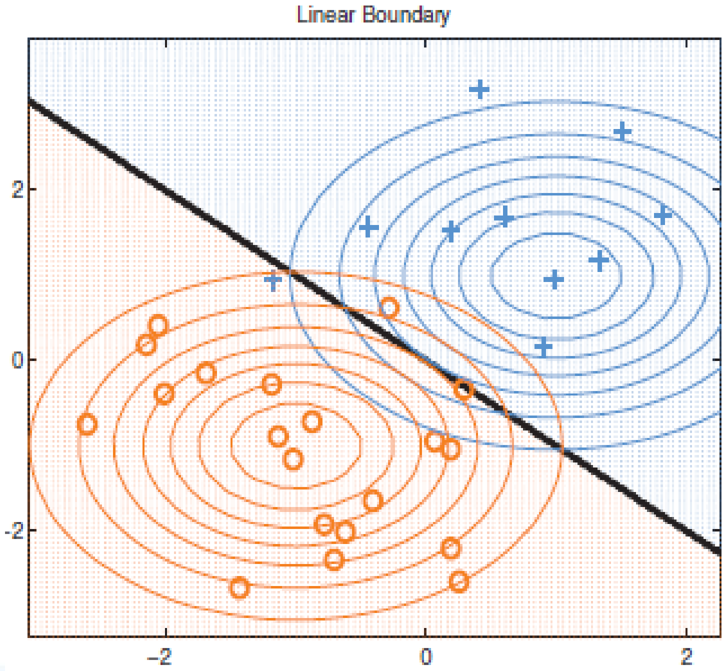


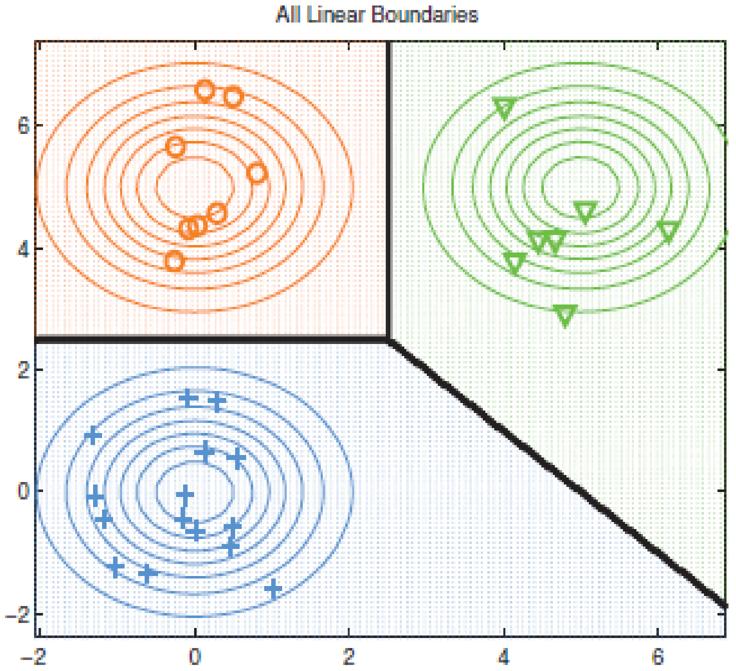
线性判别分析(LDA)

- 在二次判别分析中，如果协方差矩阵$\Sigma_{c1}=\Sigma_{c2}=\Sigma$，则判别函数δ(x)变成了一次多项式:

$$\delta(x)=-\frac{1}{2}\mu_{c1}^T\Sigma^T x-\frac{1}{2}x^T\Sigma^T\mu_{c1}-\frac{1}{2}\mu_{c1}^T\Sigma^T\mu_{c1}+\frac{1}{2}\mu_{c2}^T\Sigma^T x+\frac{1}{2}x^T\Sigma^T\mu_{c2}+\frac{1}{2}\mu_{c2}^T\Sigma^T\mu_{c2}+\log\frac{\pi_{c1}}{\pi_{c2}}$$

- **因此，称为线性判别分析**
- 其类边界δ(x) =0 为一条直线





一般由概率构造的判别分析

- 在两类分类条件下，生成式分类器为：

$$p(c_1 \mid x) = \frac{p(c_1)p(x \mid c_1)}{p(c_1)p(x \mid c_1) + p(c_2)p(x \mid c_2)} = \frac{1}{1 + \dfrac{p(c_2)p(x \mid c_2)}{p(c_1)p(x \mid c_1)}}$$

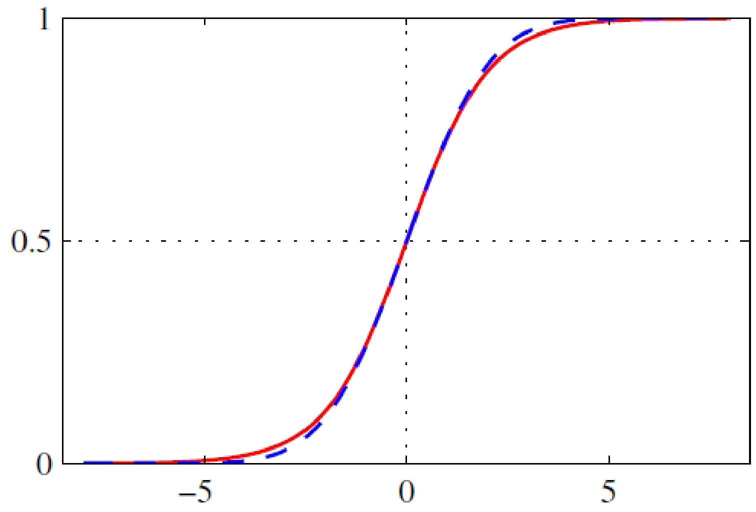
- 设

$$a = \ln \frac{p(c_1)p(x \mid c_1)}{p(c_2)p(x \mid c_2)}$$

- 则

$$p(c_1 \mid x) = \frac{1}{1 + \exp(-a)} = \sigma(a)$$

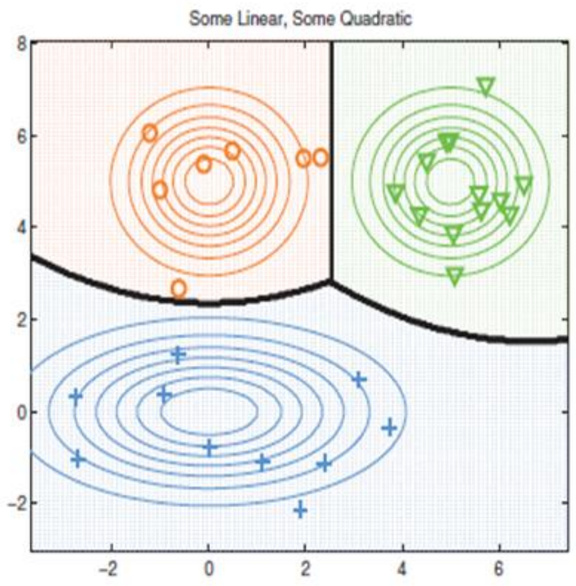
  - 这个是sigmoid函数



- 判别函数： $\delta(x) = a$

多类分类器

- 一种表示多类分类器的方法
  - 运用一组判别函数 $\delta_{i,j}(x)$, i,j=1,2,…,K
  - $x \in C_i$ ， $\text{argmax}_j\ (\delta_j(x))$



运用高斯联合分布进行推理

问题的提出

- 给定联合概率分布, $p(x1, x2)$,
- 计算这些有用的分布
  - ***边缘分布：*** $\boldsymbol{p(x1)}$
  - ***条件分布：*** $\boldsymbol{p(x1|x2)}$.

多元高斯分布的边缘分布

- 定理
  - *设* **x = (x1, x2)** *服从联合高斯分布，参数为：*

$$\mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix},\quad \Lambda=\Sigma^{-1}=\begin{pmatrix}\Lambda_{11} & \Lambda_{12}\\ \Lambda_{21} & \Sigma_{22}\end{pmatrix}$$

  - *则有边缘分布：*

$$p(x_1)=N(x_1\,|\,\mu_1,\Sigma_{11}),\quad p(x_2)=N(x_2\,|\,\mu_2,\Sigma_{22})$$

多元高斯分布的条件分布

❖ 定理

- *设 **x = (x1, x2)** 服从联合高斯分布，参数为：*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Sigma_{22} \end{pmatrix}$$

- *则有后验条件分布：*

$$p(x_1 \mid x_2) = N(x_1 \mid \mu_{1|2}, \Sigma_{1|2})$$

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(x_2 - \mu_2)$$

$$= \Sigma_{12}(\Lambda_{11}\mu_1 - \Lambda_{12}(x_2 - \mu_2))$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Lambda_{11}^{-1}$$

高斯推理举例

边缘分布与条件分布

2元高斯分布

- 二元联合高斯分布*p(x1, x2)*:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

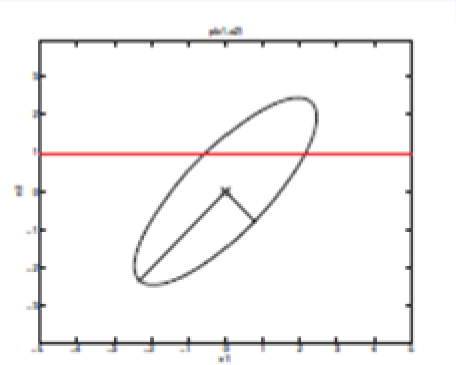
- 边缘分布 *p(x1) 是一元高斯分布*

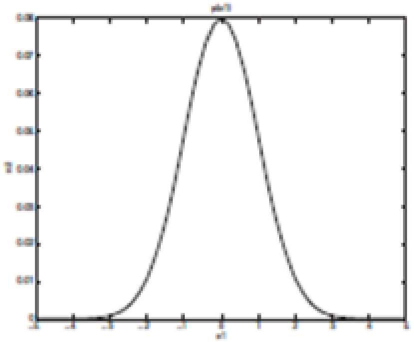
$$p(x_1) = N(x_1 \mid \mu_1, \sigma_1^2)$$

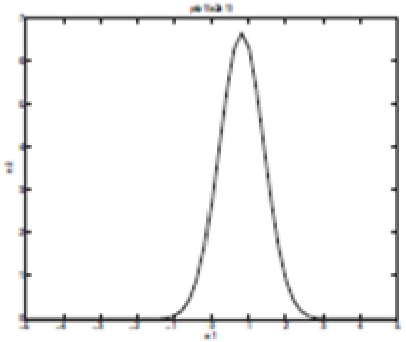
- 条件分布 *p(x1) 是一元高斯分布*

$$p(x_1 \mid x_2) = N(x_1 \mid \mu_1 + \frac{\rho\sigma_1\sigma_2}{\sigma_2^2}(x_2 - \mu_2), \sigma_1^2 - \frac{(\rho\sigma_1\sigma_2)^2}{\sigma_2^2})$$

  - *当 ρ = 0.8, σ1 = σ2 = 1, μ = 0 和 x2 = 1. p(x1|x2) = N(x1|0.8, 0.36),*







高斯推理举例

无噪声数据的函数拟合

基于无噪声数据的函数拟合

- 估计一个函数$f()$
  - 定义在 [0, T]区间，有N个观测值 $y_i = f(x_i)$
  - 因为数据无噪声， 拟合的函数会通过观测点.
- 基于概率观点看无噪声数据拟合函数
  - 函数的观测数据集： $\mathbf{f}_D = \{f_1, f_2, \ldots, f_N\}$，
  - 函数需要拟合的数据： $\mathbf{f}_* = \{f_{*1}, f_{*2}, \ldots, f_{*m}\}$
  - 拟合方法：求最大后验概率 $p(\mathbf{f}_* \mid \mathbf{f}_D)$，得到要插的值
- 这就需要将函数$f()$ 的值表达成随机变量

用随机变量表达函数值

- 将 *f()* 的定义域分成n等分：$f_j = f(s_j),\quad s_j = jh,\quad h = \frac{T}{n},\quad 0 \le j \le n$
  - 假设f() *是光滑的，则：* $f_j = \frac{1}{2}(f_{j-1} + f_{j+1}) + \varepsilon_j,\quad 0 \le j \le n-1$　　这里，$\varepsilon_j \sim N(0, 1/\lambda^2)$

$$\frac{1}{2}(-f_{j-1} + 2f_j - f_{j+1}) = \varepsilon_j,\quad 0 \le j \le n-1$$

  - 写成矩阵形式：**L*f* = *ε***

$$\mathrm{L} = \frac{1}{2}\begin{pmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & & & \\ & & & -1 & 2 & -1 \end{pmatrix} \in R^{(n-1)\times(n+1)}\qquad f = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{pmatrix}$$

随机变量函数的分布

- 设随机变量 $x \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $f()$ 为线性函数:
  - $\mathbf{y} = f(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$
- 则有: $\mathbf{y} \sim N(E[\mathbf{y}], \operatorname{cov}[\mathbf{y}])$
  - $E[\mathbf{y}] = E[\mathbf{Ax} + \mathbf{b}] = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$
    - ❿其中: $\boldsymbol{\mu} = E[\mathbf{x}]$.
  - $\operatorname{cov}[\mathbf{y}] = \operatorname{cov}[\mathbf{Ax} + \mathbf{b}] = \mathbf{A\Sigma A}^T$
    - ❿其中: $\boldsymbol{\Sigma} = \operatorname{cov}[\mathbf{x}]$.

f的先验分布

- **f**的先验分布为：

$$p(\boldsymbol{f}) = N\left(\boldsymbol{f}|0,(\lambda^2\boldsymbol{L}^T\boldsymbol{L})^{-1}\right) \propto exp\left(-\frac{\lambda^2}{2}\|\boldsymbol{L}\boldsymbol{f}\|_2^2\right)$$

  - 这是因为，$Lf = \varepsilon$，所以有，$\Sigma_\varepsilon = L\Sigma_f L^T = \frac{1}{\lambda^2}I$，所以有，$\Sigma_f =$（$\lambda^2 L^T L$）$^{-1}$

- *参数λ可看成是调节L 尺度，函数变化大小*
  - *参数λ变大：函数变光滑*
  - *参数λ变小：函数变得“摆动”。*

f的后验分布

- 将f看成两部分：

$$f = \begin{pmatrix} f_N \\ f_* \end{pmatrix}$$

$$L = (L_1, L_2)$$

$$\mathbf{\Lambda} = \mathbf{L}^T\mathbf{L} = \begin{pmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_1^T\mathbf{L}_1 & \mathbf{L}_1^T\mathbf{L}_2 \\ \mathbf{L}_2^T\mathbf{L}_1 & \mathbf{L}_2^T\mathbf{L}_2 \end{pmatrix}$$

- 所以有后验分布：

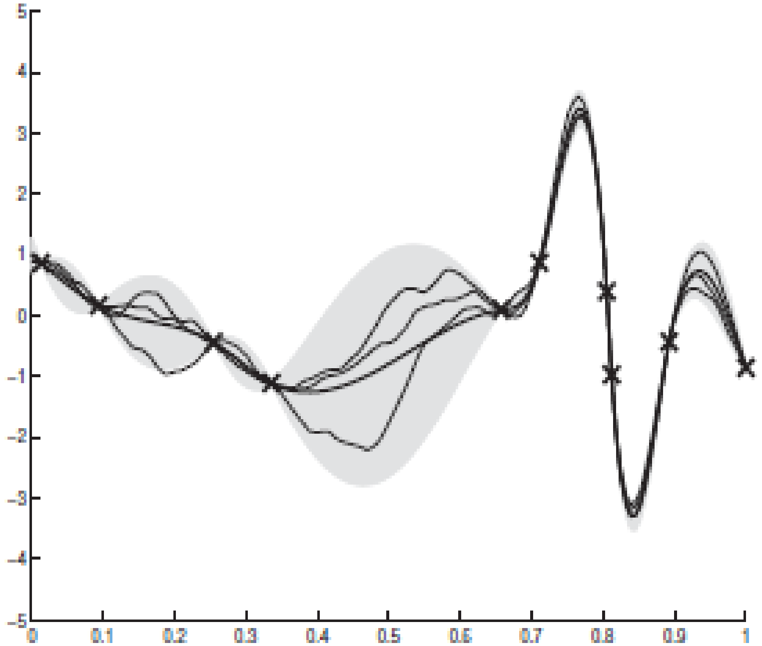
$$p(f_* | f_N) = N(\mu, \Sigma)$$

$$\mu = -\Lambda_{11}^{-1}\Lambda_{12}\boldsymbol{f}_N = -L_1^T\boldsymbol{L}_2\boldsymbol{f}_N$$

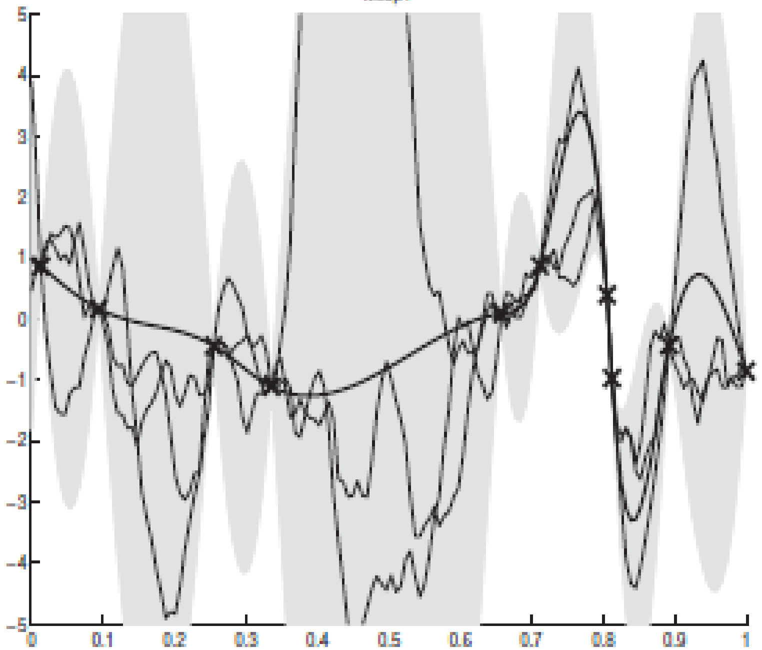
$$\Sigma = \Lambda_{11}^{-1}$$

插值算法举例

- λ不同，得到的拟合曲线平滑性不同 ((a) λ = 3，0. (b) λ = 0.01)



(a)



(b)

线性高斯系统

什么是线性高斯系统

- 我们称: **y = Ax +b为线性高斯系统，如果**
  - 设 **x** $\in$ ***R***^{Dx} *是隐随机变量,*　　***y*** $\in$ ***R***^{Dy} ***是x的具有噪声的观察值.***
  - 其先验与似然:

$$P(x) = N(x \mid \mu_x, \Sigma_x)$$

$$P(y \mid x) = N(y \mid Ax + b, \Sigma_y)$$

  - ***A*** 为 $\mathbf{D_y \times D_x}$.维矩阵
- 线性高斯系统可表示为: ***x*** → ***y,***
- 希望研究如何从***y*** 推断出 ***x***：（***y*** → ***x*** ）

高斯线性系统性质

- *给定一个线性高斯系统* **y = Ax +b**
  - 根据贝叶斯规则，可以得到*后验分布：*

$$P(x|y)=N(x|\mu_{x|y},\Sigma_{x|y})$$

$$\Sigma_{x|y}^{-1}=\Sigma_x^{-1}+A^T\Sigma_y^{-1}A \ ,$$

$$\mu_{x|y}=\Sigma_{x|y}[A^T\Sigma_y^{-1}(y-b)+\Sigma_x^{-1}\mu_x]$$

  - *观测变量先验*

$$P(y)=N(y|\mathbf{A}\mu_x+b,\Sigma_y+\mathbf{A}\Sigma_x\mathbf{A}^T)$$

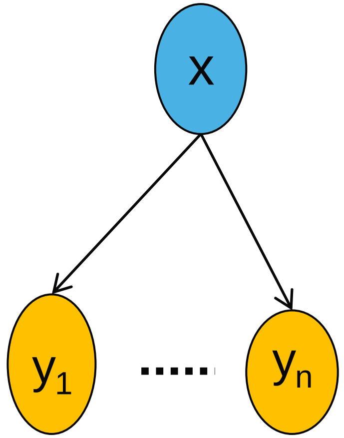
线性高斯系统应用举例

线性高斯系统推断未知标量

基于线性高斯系统推断未知标量

- 从有噪声测量值中推断未知标量
  - 假设：某个隐变量x的N个带噪声测量y_i值
  - 测量噪声具有固定精度$\lambda_y=1/\sigma^2$，
  - 似然：

$$P(y_i \mid x) = N(y_i \mid x, \lambda_y^{-1})$$



- 隐变量服从高斯分布:

$$P(x) = N(x \mid \mu_0, \lambda_0^{-1})$$

- 我们的问题就是要计算隐变量的后验：$p(x|y_1, . . . , y_N, \sigma^2)$.

计算观测值条件下隐变量的后验概率

- 将问题定义成向量形式:
  - **y = (**$y_1, \ldots, y_N$**),**
  - ***A* = 1**^T_N *(an 1 × N row vector of 1's),*
  - $\boldsymbol{\Sigma}^{-1}_y = diag(\lambda y \mathbf{I})$**.**
- 可以得到:

$$P(x \mid y) = N(x \mid \mu_N, \lambda_N^{-1})$$

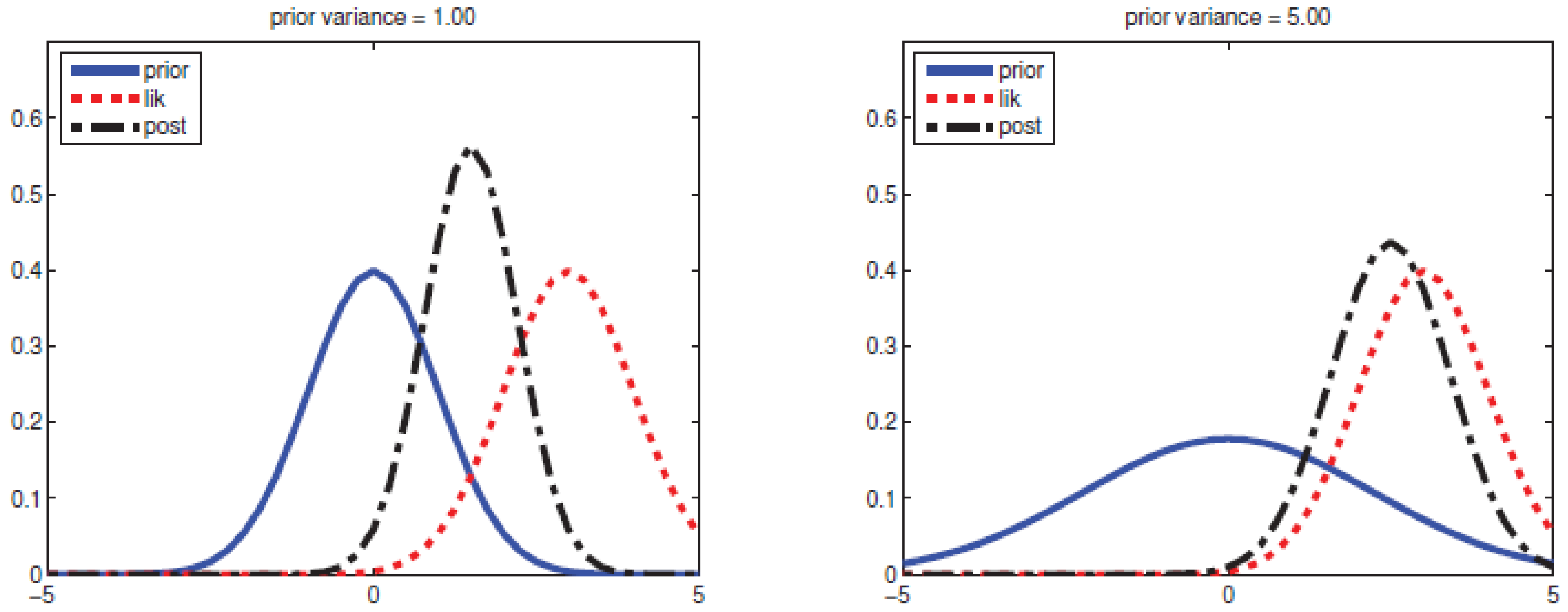
$$\lambda_N = \lambda_0 + N\lambda_y$$

$$\mu_N = \frac{N\lambda_y \bar{y} + \lambda_0 \mu_0}{\lambda_N} = \frac{N\lambda_y}{N\lambda_y + \lambda_0}\bar{y} + \frac{\lambda_0}{N\lambda_y + \lambda_0}\mu_0$$

$$where \;\; \bar{y} = \mathrm{MLE}$$

推理结果

- 给定带噪声的观测值$y = 3$，*推断隐变量x*



线性高斯系统应用举例

线性高斯系统推断未知向量

基于线性高斯系统推断未知向量

- 观测向量 $\boldsymbol{y_i} \sim \boldsymbol{N(x,\Sigma_y)}$　　$1 \le i \le N$

- 隐变量的先验 $\boldsymbol{x} \sim N(\boldsymbol{\mu_0,\Sigma_0})$.

- 设 $A = I$, $b = 0$, 则有：

$$P(\mathbf{x} \mid y_1, \cdots, y_N) = N(\mathbf{x} \mid \mu_N, \Sigma_N)$$

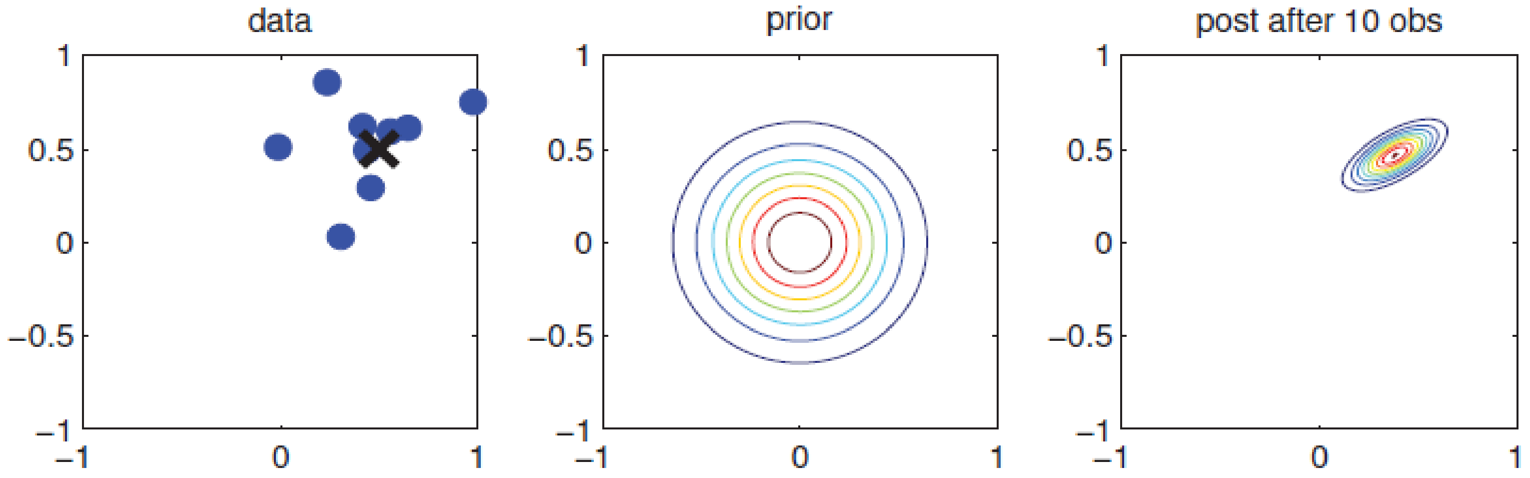
$$\Sigma_N^{-1} = \Sigma_0^{-1} + N\Sigma_y^{-1}$$

$$\mu_N = \Sigma_N(\Sigma_y^{-1}(N\overline{y}) + \Sigma_0^{-1}\mu_0)$$

推断未知向量举例

- *假设:*
  - X为物体在二维空间中的真实但未知位置，比如雷达上的导弹或飞机：$x = [0.5, 0.5]^T$
  - $\mathbf{y}_i$ 为有噪声的观测值，如雷达“闪烁”：$\mathbf{y}_i \sim N(\boldsymbol{x},\boldsymbol{\Sigma}_y)$, $\Sigma y = 0.1[2, 1; 1, 1]$
- 当收到更多信号时，我们可以更好地定位来源。
- 先验分布：$p(x) = N(x|0, 0.1\mathbf{I}_2)$.
- 基于10 个观测数据的后验:

  $P(x|y_1,y_2,\ldots,y_{10})$



基于线性高斯系统的传感器融合

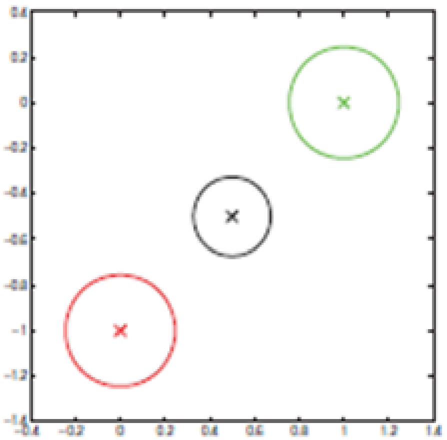
- 我们想把多个测量设备的结果结合起来；
  - $\mathbf{y_i = A_i x + b_i}$
  - 每个传感器的可靠性不同，观测值具有不同协方差
  - 对应的后验概率是数据的适当加权平均值。
  - 融合的任务：*求* $p(x|y_1,\ldots, y_n)$.

基于线性高斯系统的传感器融合举例

- 假设隐变量x服从正态分布，但方差很大，实际不含什么信息:
  - $p(x) = N(\mu_0,\Sigma_0) = N(0, 10^{10}\mathbf{I}_2)$
- 有 2个带噪声的观测量： $y_1 \sim N(x,\Sigma_{y1})$ and $y_2 \sim N(x,\Sigma_{y,2})$
- 融合的任务：求 $p(x|y1, y2)$.
- 讨论 3 种情形
  - 2个传感器可靠性同样
  - 2个传感器可靠性不同
  - 2个传感器各有所长

情形1：2个传感器同样可靠

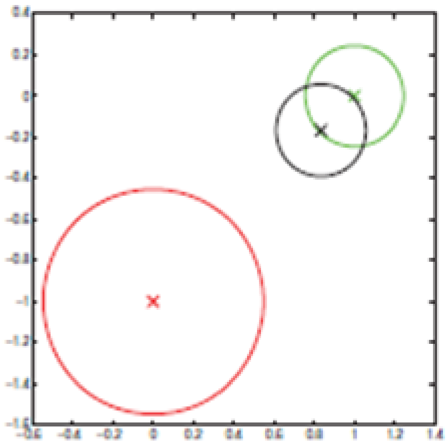
- 两个传感器可靠性相同
  - $\Sigma_{y1} = \Sigma_{y2} = 0.01\mathbf{I}_2$
- 观测到： $y_1 = (0, -1)$ (*红叉*)， $y_2 = (1, 0)$ (*绿叉*)
- *推断*：$E(\mu|y1, y2, \theta)$ (黑叉).
  - 后验均值位于y1 和y2之间的中点



(a)

情形2: 2个传感器可靠性不同

- *两个传感器可靠性不同*
  - $\boldsymbol{\Sigma}_{y1} = 0.05\mathbf{I}_2$ *and* $\boldsymbol{\Sigma}_{y,2} = 0.01\mathbf{I}_2$,
  - 传感器 2 比传感器1 更可靠
  - 后验均值更接近 y2



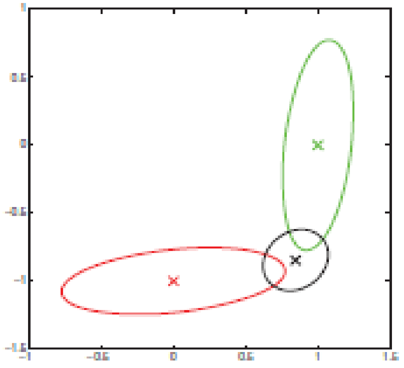
(b)

情形3：两个传感器各有所长

- *传感器1在竖直方向更可靠，传感器2在水平方向更可靠*

$$\Sigma_{y1} = 0.01\begin{pmatrix} 10 & 1 \\ 1 & 1 \end{pmatrix},\quad \Sigma_{y2} = 0.01\begin{pmatrix} 1 & 1 \\ 1 & 10 \end{pmatrix}$$

- 后验均值包含两个传感器的优点
  - **y1的竖直成分**
  - **y2的水平成分**



(c)

线性高斯系统应用举例

基于有噪声数据拟合函数

基于有噪声数据拟合函数

- 有N 个含噪声观测值 y_i; 对应于观测元素 $x_1, ... , x_N$.（这是认为x在不断变化）
- 基于此，构造线性高斯系统:
  - **y = Ax + ε**，其中 ε ~ N(0,Σ_y), $\Sigma_y = \sigma^2 I$, σ^2 为观测噪声
  - **A** 是 **N** × **D** 维投影矩阵，表示选出的观测元素
- 先验采样与无噪声拟合时相同：$\Sigma_x = \left(L^T L\right)^{-1}$
- 根据贝叶斯规则：

$$P(x\,|\,y) = N(x\,|\,\mu_{x|y}, \Sigma_{x|y})$$

$$\Sigma_{x|y}^{-1} = \Sigma_x^{-1} + A^T \Sigma_y^{-1} A\ ,$$

$$\mu_{x|y} = \Sigma_{x|y}[A^T \Sigma_y^{-1}(y-b) + \Sigma_x^{-1}\mu_x]$$

- 即可得出拟合函数

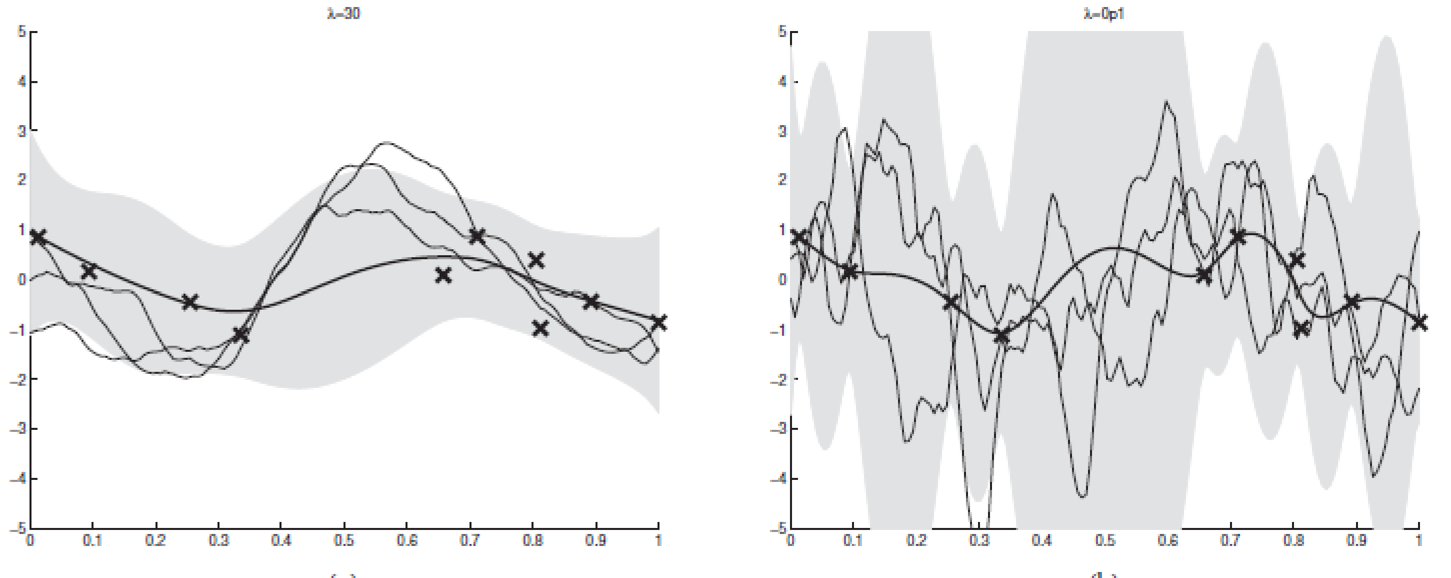
基于有噪声数据拟合函数举例

- 如果观测次数 $N = 2$， *隐变量* x *是*$D = 4$*维向量,* A*为选择观测元素的矩阵，则有：*

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

- $\boldsymbol{\Sigma_x = (L^T L)^{-1}}$ ($\boldsymbol{L}$ is the *是与无噪声数据插值例子中的L相同*

- 噪声方差 $\sigma 2 = 1$
- *斯先验精度参数* λ*.*
  - *(a)* $\lambda = 30$*. (b)* $\lambda = 0.01$*.*



(a) (b)

推断多元高斯分布的参数

问题的提出

- 如何推断高斯分布的参数本身
  - 设数据 ***xi* ~ *N*(*μ*,*Σ*) *for i* = *1* : *N*** .
- 为简化问题, 分三步导出后验分布（通过迭代求解）:
  - 首先，估计 *p*(***μ*|*D*,*Σ*);**
  - 然后，计算*p*(***Σ*|*D*,*μ*);**
  - *最后得到联合分布*：**p(*μ*,*Σ*|D).**

估计均值 μ的后验分布

- 前面已经讨论过基于最大似然估计参数:

$$P(D|\mu) = N(\bar{x}|\mu, \frac{1}{N}\Sigma)$$

- 采用共轭先验*p(μ)* = ***N(μ|m0,V0)*** .
- 可以得到后验分布:

$$P(\mu|D,\Sigma) = N(\mu|m_N, V_N)$$

$$V_N^{-1} = V_0^{-1} + N\Sigma^{-1}$$

$$m_N = V_N(\Sigma^{-1}(N\bar{x}) + V_0^{-1}m_0)$$

- 如果设先验分布几乎等于均匀分布：***V0*** **= ∞I.**
  - 后验分布变成：***p(μ|D,Σ)*** = *N(* $\bar{x}$ ***, (1****/N****)Σ)***，*这意味着:*
    - **后验均值等于最大似然**
    - **后验方差变成了：1/N.**


AUSTRALIA
OCEANIA
Northern Territory
Queensland
NEW ZEALAND
VANUATU
TONGA

Thank You !