回归的线性模型
LINEAR MODELS FOR REGRESSION

张玲玲
计算机学院
zhanglling@xjtu.edu.cn

主要内容

1. 回归简介
2. 线性基函数模型
3. 贝叶斯线性回归

回归简介

- 回归是有监督学习中的一个重要问题。
- 有监督学习：给定N个输入变量的观测值$\{\mathbf{x}_n\}$及对应的目标变量值$\{t_n\}$。输入一个新的变量$\mathbf{x}$，预测其目标变量t的值。
- 回归用于预测输入变量（自变量）和目标变量（因变量）之间的关系，特别是当输入变量的值发生变化时，输出变量的值随之发生的变化。
- 回归的目的：给定一个D维输入变量$\mathbf{x}$，输出一个或多个连续目标变量t的值。

回归简介

- 回归分为学习和预测两个过程。
- 回归的学习：基于训练数据（输入变量值和目标变量值）$\{(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2), \cdots, (\mathbf{x}_N, t_N)\}$，构建一个函数$y(\mathbf{x})$，使其很好地拟合已知数据，且很好地预测未知数据。等价于函数拟合。
- 回归的预测：输入新的变量$\mathbf{x}$，使函数$y(\mathbf{x})$的值是目标变量t的预测值。
- 线性回归模型：使用一组固定的（非）线性函数（基函数）的组合表示回归模型。

主要内容

1. 回归简介
2. 线性基函数模型
3. 贝叶斯线性回归

线性基函数模型

- 最简单的线性回归模型是输入变量的线性组合：

$$y(\mathbf{x},\mathbf{w}) = w_0 + w_1 x_1 + \cdots + w_D x_D$$

  其中，$\mathbf{x} = (x_1, \cdots, x_D)^{\mathrm{T}}$，$D \times 1$维。

- 关键特性：$\mathbf{w} = (w_0, w_1, \cdots, w_D)^{\mathrm{T}}$，$D \times 1$维。

  $y(\mathbf{x},\mathbf{w})$是参数$w_0, w_1, \cdots, w_D$的线性模型。

- $y(\mathbf{x},\mathbf{w})$也是输入变量$\mathbf{x}$的分量x_i的线性模型。

- 但由于这个模型是直接对各分量进行加权求和，因此它的最大局限是描述能力不足。

线性基函数模型

- 一般地

$$\boxed{y(\mathbf{x},\boldsymbol{w}) = \sum_{j=0}^{M-1} w_j\,\phi_j(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x})}$$

  其中，$\boxed{\phi_j(\mathbf{x})\text{称作基函数}}$（basis functions），

  $\mathbf{w} = (w_0, \ldots, w_{M-1})^{\mathrm{T}}$，$\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{M-1})^{\mathrm{T}}$。
- 通常，$\phi_0(\mathbf{x}) = 1$，因此w_0表示偏差（bias）。
- 最简单情况：使用线性基函数，即$\phi_d(\mathbf{x}) = x_d$。

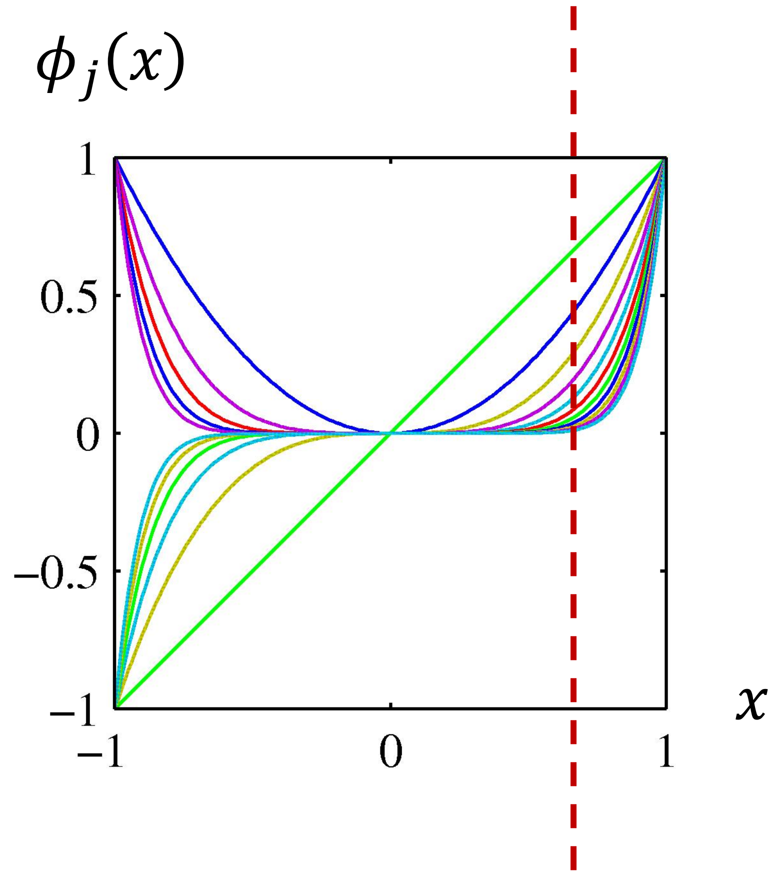
$$y(\mathbf{x},\mathbf{w}) = w_0 + w_1x_1 + \cdots + w_Dx_D$$

线性基函数模型

- 多项式基函数：

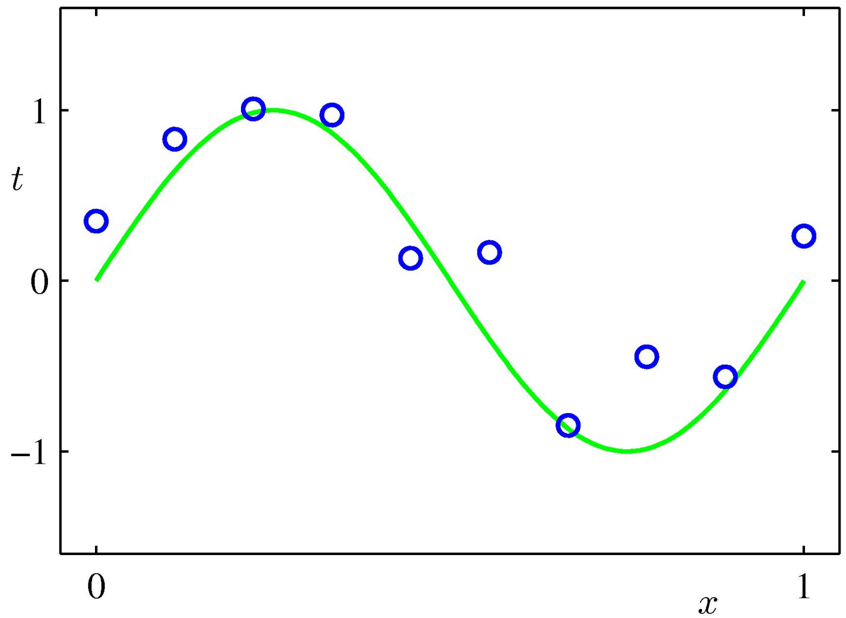
$$\phi_j(x) = x^j$$

- 这些基函数是全局的；x的微小变化也会影响所有基函数。
- 解决办法：将输入空间划分为多个区域，并在每个区域使用不同的多项式来拟合。



线性基函数模型

例子：多项式曲线拟合



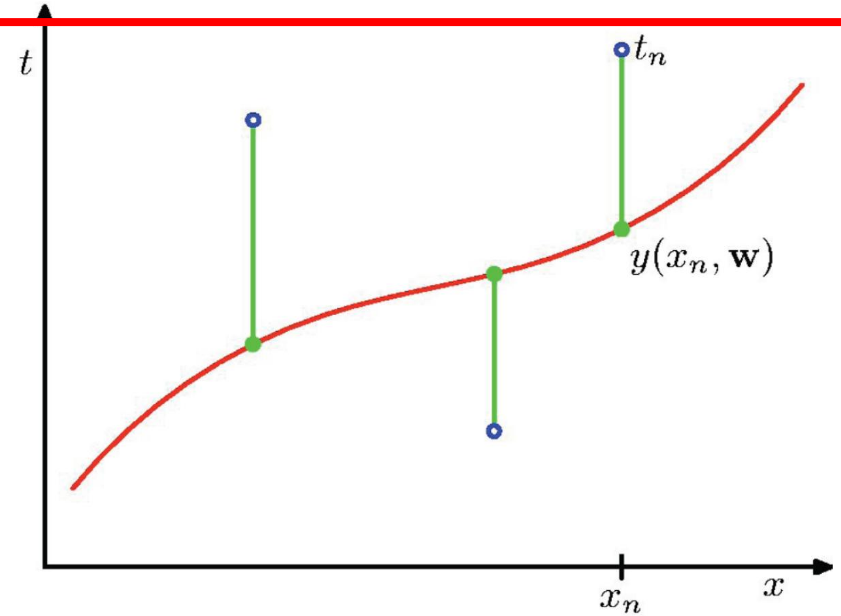
$$\phi_j(x) = x^j$$

$$y(x, \boldsymbol{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

线性基函数模型

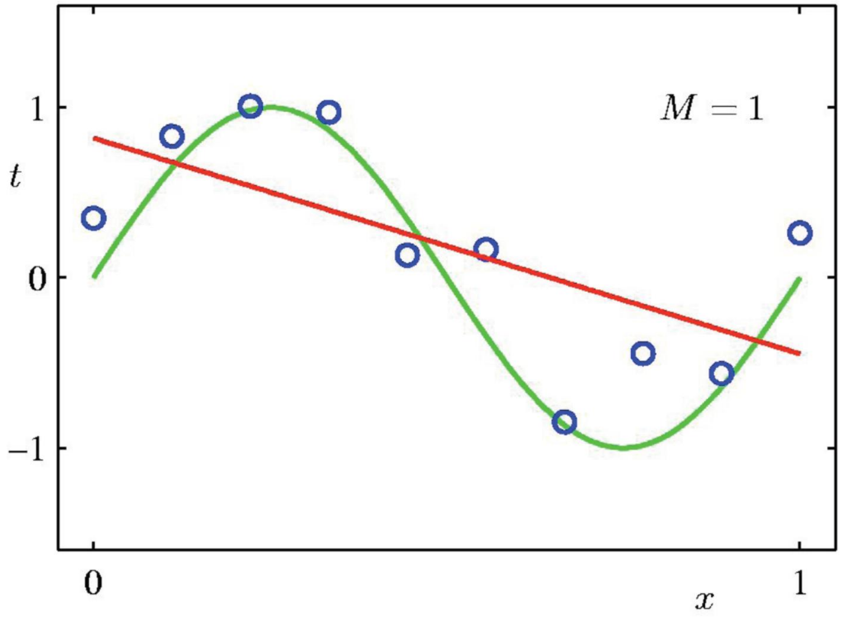
损失函数：平方和误差函数

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \boldsymbol{w}) - t_n\}^2$$



线性基函数模型

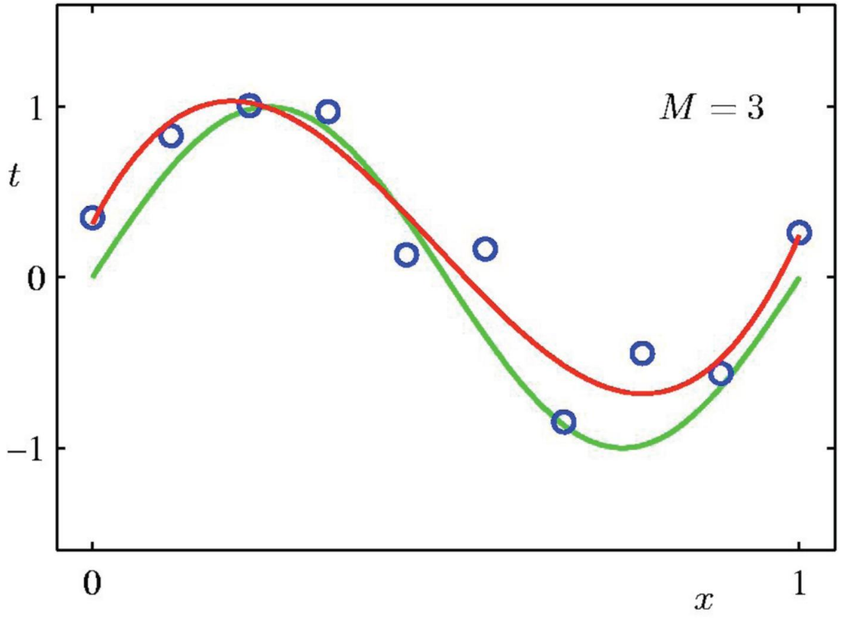
一阶多项式曲线拟合



$$y(x, \boldsymbol{w}) = w_0 + w_1 \boxed{x}$$

线性基函数模型

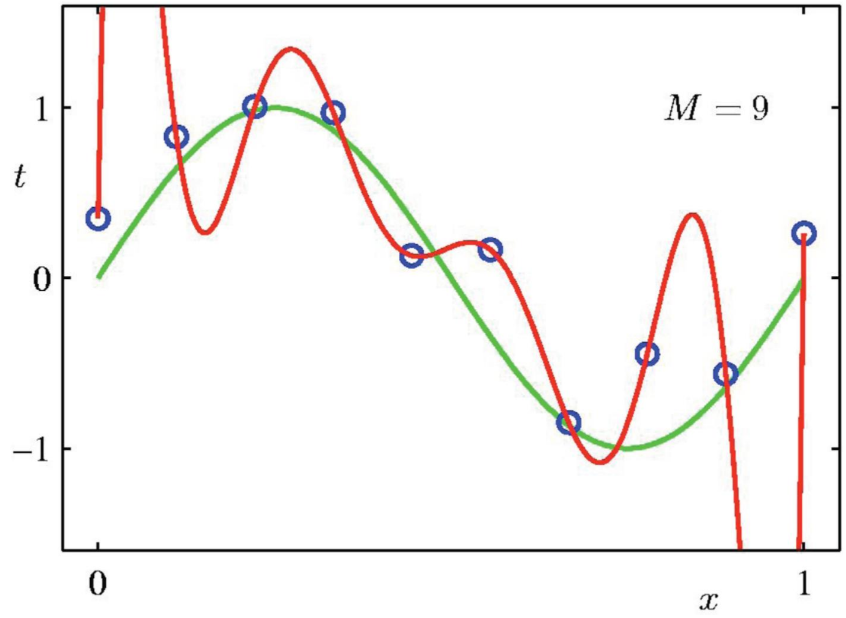
三阶多项式曲线拟合



$$y(x,\boldsymbol{w}) = w_0 + w_1 x + w_2 x^2 + w_3 \boxed{x^3}$$

线性基函数模型

九阶多项式曲线拟合



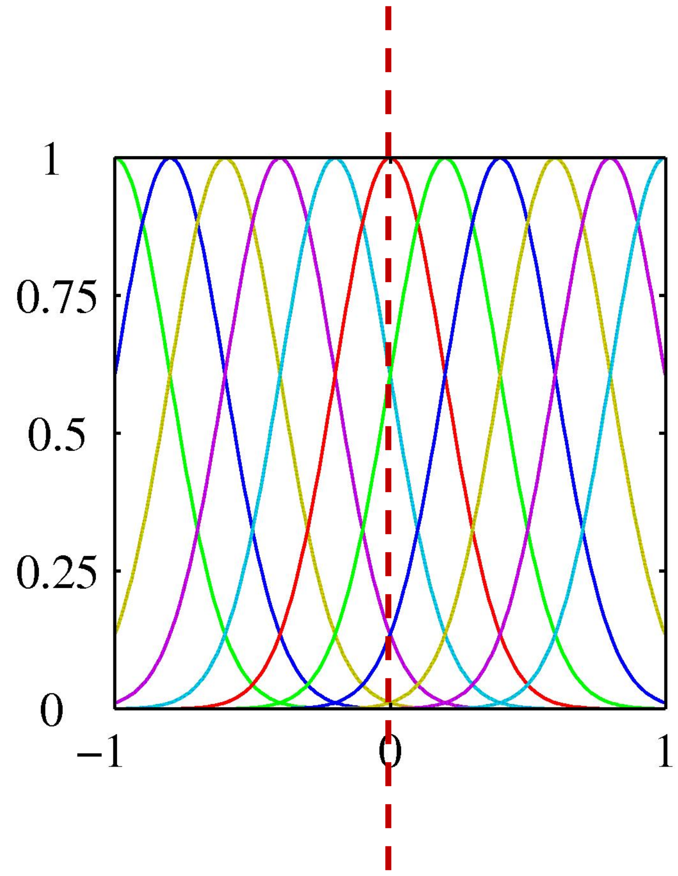
$$y(x, \boldsymbol{w}) = w_0 + w_1 x + \cdots + w_9 \boxed{x^9}$$

线性基函数模型

- 高斯基函数：

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

- μ_j和s分别控制基函数在输入空间中的位置和尺度（宽度）。
- 这些基函数是局部的；x的微小变化仅影响邻近的基函数。



线性基函数模型

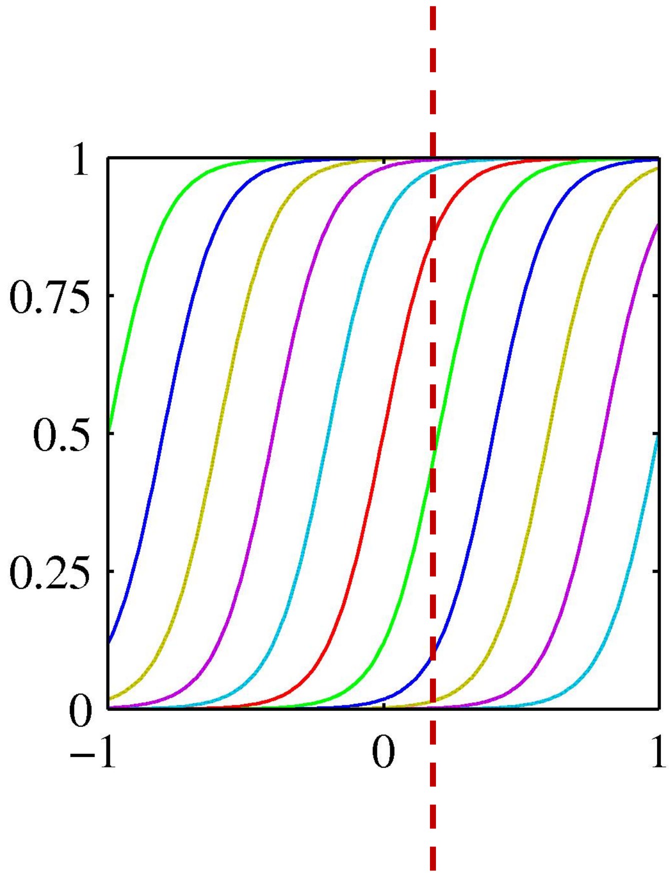
- S型（Sigmoidal）基函数：

  $$\phi_j(x) = \sigma\left(\frac{x-\mu_j}{s}\right)$$

  其中

  $$\sigma(a) = \frac{1}{1+\exp(-a)}$$

- μ_j和s分别控制位置和尺度（斜率）。
- 这些基函数也是局部的；x的微小变化仅影响邻近的基函数。



最大似然和最小二乘

- 假设目标变量t来自带有高斯噪声的确定性函数：
$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$
- 其中，$p(\epsilon|\beta) = \mathcal{N}(\epsilon|0,\ \beta^{-1})$
- 因此，目标变量t的分布等价于，
$$p(t|\mathbf{x}, \mathbf{w}, \beta) = \mathcal{N}(t|y(\mathbf{x}, \mathbf{w}),\ \beta^{-1})$$
- 给定观测输入$\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$，和目标变量$\mathbf{t} = [t_1, \ldots, t_N]^{\mathrm{T}}$，可以得到似然函数

$$p(\mathbf{t}|\mathbf{w}, \beta) = p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n),\ \beta^{-1})$$

最大似然和最小二乘

- 对似然函数使用单变量高斯函数的标准形式：

$$\mathcal{N}(x|\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

并对似然函数取对数，有

$$\ln p(\mathbf{t}|\mathbf{w},\beta)=\sum_{n=1}^{N}\ln\mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n),\beta^{-1})=\frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi)-\beta E_D(\mathbf{w})$$

Proof见下一页

其中，

$$E_D(\mathbf{w})=\frac{1}{2}\sum_{n=1}^{N}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$

是平方和误差。

$$\ln p(\mathbf{t}|\mathbf{w},\beta) = \sum_{n=1}^{N} \ln\mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n),\beta^{-1})$$

$$= \sum_{n=1}^{N} \ln\left(\frac{1}{(2\pi\beta^{-1})^{1/2}}\exp\left\{-\frac{1}{2\beta^{-1}}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2\right\}\right)$$

$$= \sum_{n=1}^{N}\left(\frac{1}{2}\ln\beta-\frac{1}{2}\ln(2\pi)-\frac{\beta}{2}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2\right)$$

$$= \frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi)-\frac{\beta}{2}\sum_{n=1}^{N}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$

$$= \frac{N}{2}\ln\beta-\frac{N}{2}\ln(2\pi)-\beta\boxed{E_D(\mathbf{w})}$$

$\boxed{E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2\text{，平方和误差}}$

最大似然和最小二乘

- 对似然函数$\ln p(\mathbf{t}|\mathbf{w},\beta)$关于$\mathbf{w}$计算梯度并置零：

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = \nabla_{\mathbf{w}} \left(\frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \frac{\beta}{2} \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n)\}^2 \right) = \mathbf{0}$$

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = \nabla_{\mathbf{w}} \left(-\frac{\beta}{2} \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n)\}^2 \right) = \mathbf{0}$$

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = -\beta \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n)\} \nabla_{\mathbf{w}} \left(-\mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n) \right) = \mathbf{0}$$

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = \beta \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}} \phi(\mathbf{x}_n)\} \phi(\mathbf{x}_n)^{\mathrm{T}} = \mathbf{0}$$

最大似然和最小二乘

- 对似然函数$\ln p(\mathbf{t}|\mathbf{w},\beta)$关于$\mathbf{w}$计算梯度并置零：

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = \beta \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}\,\phi(\mathbf{x}_n)^{\mathrm{T}} = \mathbf{0}$$

- 求解$\mathbf{w}$，可得

$$\mathbf{w}_{\mathrm{ML}} = \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}$$

Moore-Penrose 伪逆，
$\mathbf{\Phi}^{\dagger} \equiv \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathrm{T}}$

proof见下一页

其中，

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_{M-1}(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_{M-1}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \phi_1(\mathbf{x}_N) & \cdots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}.$$

$$\Phi_{nj} = \phi_j(\mathbf{x}_n)$$

$$\nabla_{\mathbf{w}} \ln p(\mathbf{t}|\mathbf{w},\beta) = \beta \sum_{n=1}^{N} \{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}\,\phi(\mathbf{x}_n)^{\mathrm{T}} = 0$$

$$\sum_{n=1}^{N} t_n \phi(\mathbf{x}_n)^{\mathrm{T}} = \sum_{n=1}^{N} \{\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}\phi(\mathbf{x}_n)^{\mathrm{T}}$$

$$\sum_{n=1}^{N} \phi(\mathbf{x}_n) t_n = \sum_{n=1}^{N} \phi(\mathbf{x}_n)\{\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^{\mathrm{T}}$$

$$\sum_{n=1}^{N} \phi(\mathbf{x}_n) t_n = \sum_{n=1}^{N} \phi(\mathbf{x}_n)\phi(\mathbf{x}_n)^{\mathrm{T}}\mathbf{w}$$

$$[\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \cdots, \phi(\mathbf{x}_n)] \begin{bmatrix} t_1 \\ t_2 \\ \cdots \\ t_n \end{bmatrix} = [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \cdots, \phi(\mathbf{x}_n)] \begin{bmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \cdots \\ \phi(\mathbf{x}_n) \end{bmatrix} \mathbf{w}$$

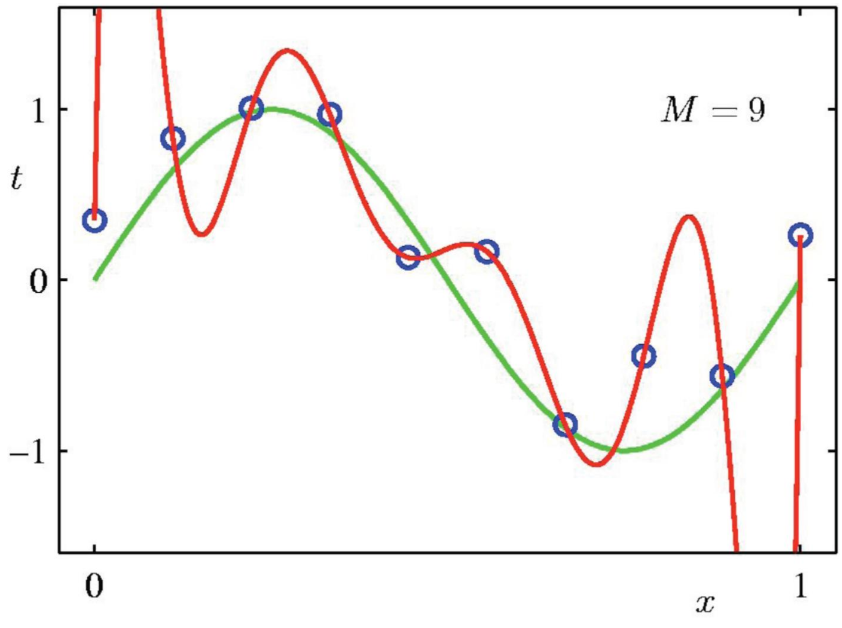
$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{t} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w}$$

$$(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\mathbf{w} \Rightarrow \mathbf{w}_{\mathrm{ML}} = (\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{T}}\mathbf{t}$$

正则化最小二乘

- 误差函数

$$E_D(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$



正则化最小二乘

- 考虑误差函数：

$$\underbrace{E_D(\mathbf{w})}_{\text{数据项}} + \overbrace{\lambda}^{\text{正则化系数}} \underbrace{E_W(\mathbf{w})}_{\text{正则化项}}$$

- 选择二次正则器和平方和误差函数，可得到

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

正则化最小二乘

- 考虑误差函数：

正则化系数

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

数据项

正则化项

- 选择二次正则器和平方和误差函数，可得到

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

- 关于$\mathbf{w}$求梯度并置零，有

$$\mathbf{w} = (\lambda \mathrm{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}$$

$$\boxed{\mathbf{w}_{\mathrm{ML}} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}}$$

主要内容

1. 回归简介
2. 线性基函数模型
3. 贝叶斯线性回归

贝叶斯线性回归

- 在$\mathbf{w}$上定义一个共轭高斯先验（conjugate prior）
$$\boxed{p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0).}$$
- 将先验与似然函数相结合，并使用边际和条件高斯分布的结果，得出$\mathbf{w}$上的后验分布
$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$
其中
$$\begin{aligned}
\mathbf{m}_N &= \mathbf{S}_N\left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}\right) \\
\mathbf{S}_N^{-1} &= \mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}.
\end{aligned}$$
$$\mathbf{w}_{\mathrm{MAP}} = \mathbf{m}_N$$

贝叶斯线性回归

- 通常选择的先验是零均值各向同性高斯分布：

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

- 在$\mathbf{w}$上的后验分布的均值和方差变成

$$\begin{aligned} \mathbf{m}_N &= \beta \mathbf{S}_N \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_N^{-1} &= \alpha \mathbf{I} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}. \end{aligned}$$

$$\boxed{\begin{aligned} \mathbf{m}_N &= \mathbf{S}_N \left(\mathbf{S}_0^{-1} \mathbf{m}_0 + \beta \boldsymbol{\Phi}^{\mathrm{T}} \mathbf{t} \right) \\ \mathbf{S}_N^{-1} &= \mathbf{S}_0^{-1} + \beta \boldsymbol{\Phi}^{\mathrm{T}} \boldsymbol{\Phi}. \end{aligned}}$$

贝叶斯线性回归

- 对后验分布取对数，它就变成了似然的对数和先验的对数（**w**的函数）之和：

$$\ln p(\mathbf{w}|\mathbf{t}) = -\frac{\beta}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2 - \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \mathrm{const}$$

似然对数：

$$\ln p(\mathbf{t}|\mathbf{w},\beta) = \frac{N}{2}\ln\beta - \frac{N}{2}\ln(2\pi) - \frac{\beta}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n)\}^2$$

先验对数：

$$\ln p(\mathbf{w}) = -\frac{1}{2}\ln(2\pi) + \frac{1}{2}\ln\alpha - \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

贝叶斯线性回归

- 对后验分布取对数，它就变成了似然的对数和先验的对数（$\mathbf{w}$的函数）之和：

$$\ln p(\mathbf{w}|\mathbf{t}) = -\frac{\beta}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 - \frac{\alpha}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \mathrm{const}$$

- 对后验分布的对数关于$\mathbf{w}$求最大化，等同于对平方和误差函数加上正则化项求最小化，有

$$\frac{1}{2}\sum_{n=1}^{N}\{t_n - \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}_n)\}^2 + \frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$

$$\lambda = \alpha/\beta$$

贝叶斯线性回归

- 以直线拟合为例，说明线性基函数模型中的贝叶斯学习，以及后验分布的顺序更新。
- 考虑
  - 一个单输入变量x
  - 一个单目标变量t
  - 一个线性模型$y(x, \mathbf{w}) = w_0 + w_1 x$
  - 由于只有两个可调参数，我们可以在参数空间中直接画出先验分布和后验分布。

贝叶斯线性回归

- 生成合成数据
  - 首先从均匀分布$\mathrm{U}(x|-1,+1)$中采样得到x_n的值
  - 然后计算函数$f(x,\mathbf{a})=-0.3+0.5x$的值
  - 再给其加上标准偏差为0.2的高斯噪声
  - 最终得到目标变量值t_n

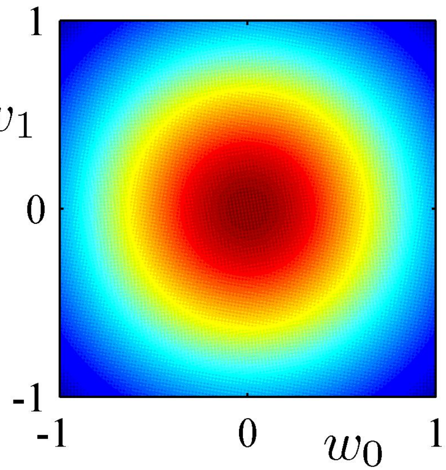
- 将(x_n,t_n)作为观测数据

贝叶斯线性回归

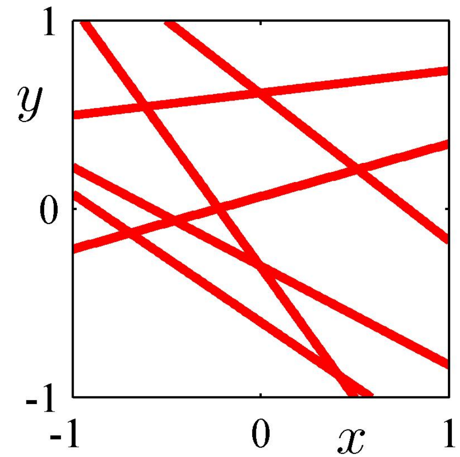
- 观测到0个数据点

真实函数$f(x, \mathbf{a}) = -0.3 + 0.5x$

先验



数据空间



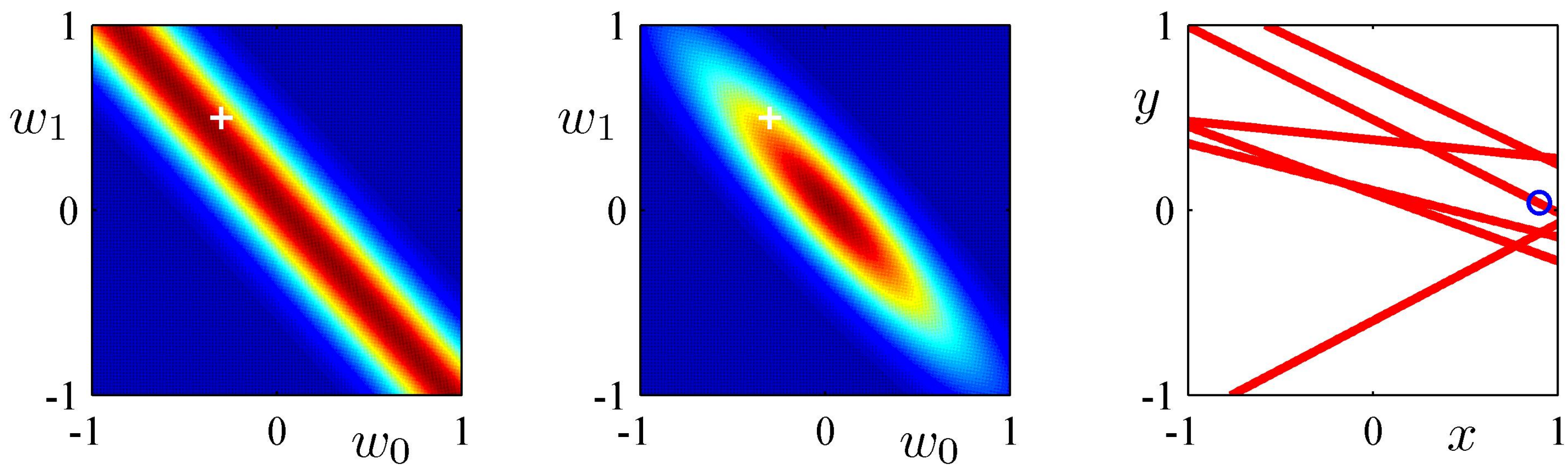
$p(\mathbf{w})$

$$y(x, \mathbf{w}) = w_0 + w_1 x$$

贝叶斯线性回归

- 观测到1个数据点　　真实函数$f(x, \mathbf{a}) = -0.3 + 0.5x$

似然　　后验　　数据空间

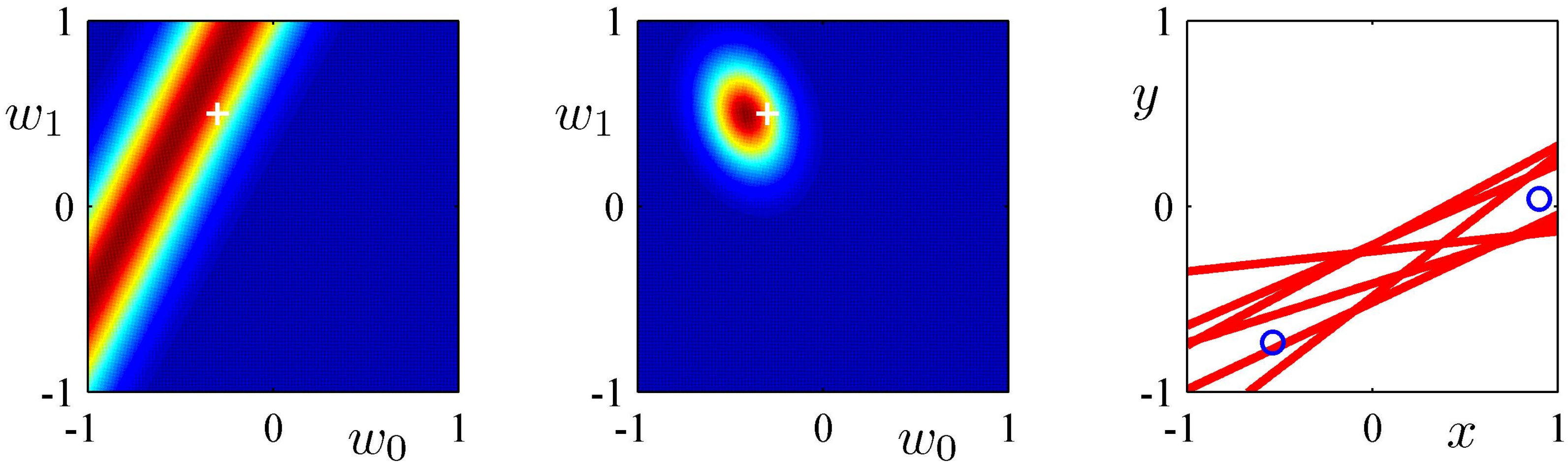


$p(t|x, \mathbf{w})$　　$p(\mathbf{w}|t)$　　$y(x, \mathbf{w}) = w_0 + w_1 x$

贝叶斯线性回归

- 观测到2个数据点 真实函数$f(x, \mathbf{a}) = -0.3 + 0.5x$

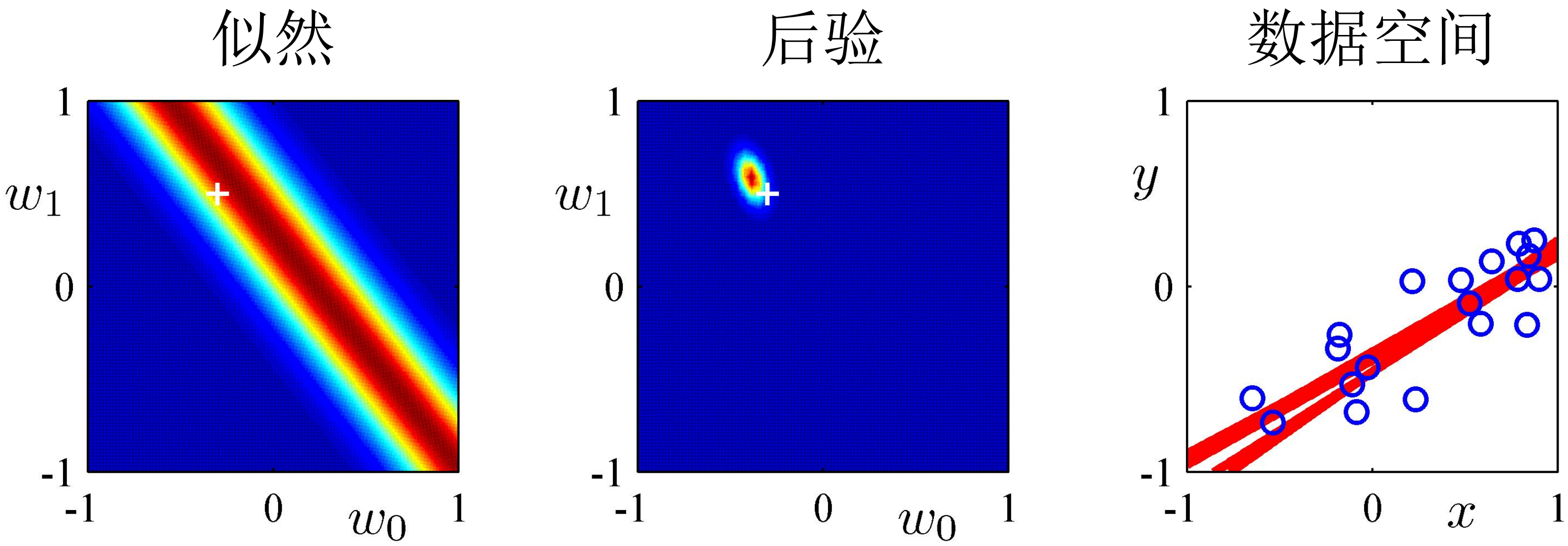
似然　　后验　　数据空间



$p(t|x, \mathbf{w})$　　$p(\mathbf{w}|t)$　　$y(x, \mathbf{w}) = w_0 + w_1 x$

贝叶斯线性回归

- 观测到20个数据点　　真实函数$f(x,\mathbf{a}) = -0.3 + 0.5x$



$p(t|x,\mathbf{w})$　　$p(\mathbf{w}|t)$　　$y(x,\mathbf{w}) = w_0 + w_1 x$

预测分布

- 对新的输入$\mathbf{x}$预测目标变量t的值：

$\mathbf{x},\mathbf{w}$

$$p(t|\mathbf{t},\alpha,\beta) \quad = \quad \int p(t|\mathbf{w},\beta)p(\mathbf{w}|\mathbf{t},\alpha,\beta)\,\mathrm{d}\mathbf{w}$$

其中，目标变量t的条件分布是：

$$p(t|\mathbf{w},\beta) = p(t|\mathbf{x},\mathbf{w},\beta) = \mathcal{N}(t|y(\mathbf{x},\mathbf{w}),\beta^{-1})$$

参数$\mathbf{w}$的后验分布是：

$$p(\mathbf{w}|\mathbf{t},\alpha,\beta) = p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N,\mathbf{S}_N)$$

预测分布

- 预测分布对**w**进行积分，有：

$$
\begin{aligned}
p(t|\mathbf{t},\alpha,\beta) &= \int p(t|\mathbf{w},\beta)p(\mathbf{w}|\mathbf{t},\alpha,\beta)\,\mathrm{d}\mathbf{w} \\
&= \mathcal{N}(t|\mathbf{m}_N^{\mathrm{T}}\boldsymbol{\phi}(\mathbf{x}),\sigma_N^2(\mathbf{x}))
\end{aligned}
$$

其中

$$
\sigma_N^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\phi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_N\boldsymbol{\phi}(\mathbf{x}).
$$

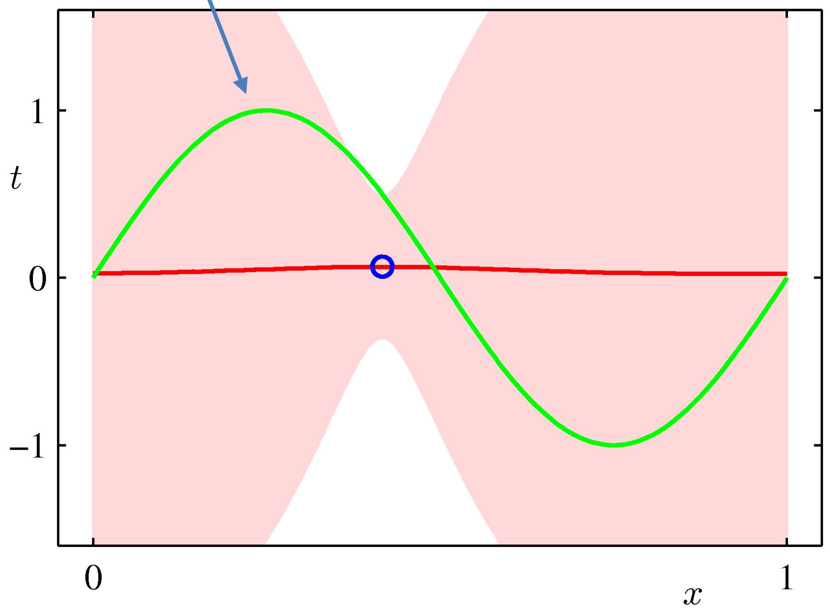
- 对于预测分布来讲，有

$$
\sigma_{N+1}^2(\mathbf{x}) \leqslant \sigma_N^2(\mathbf{x})
$$

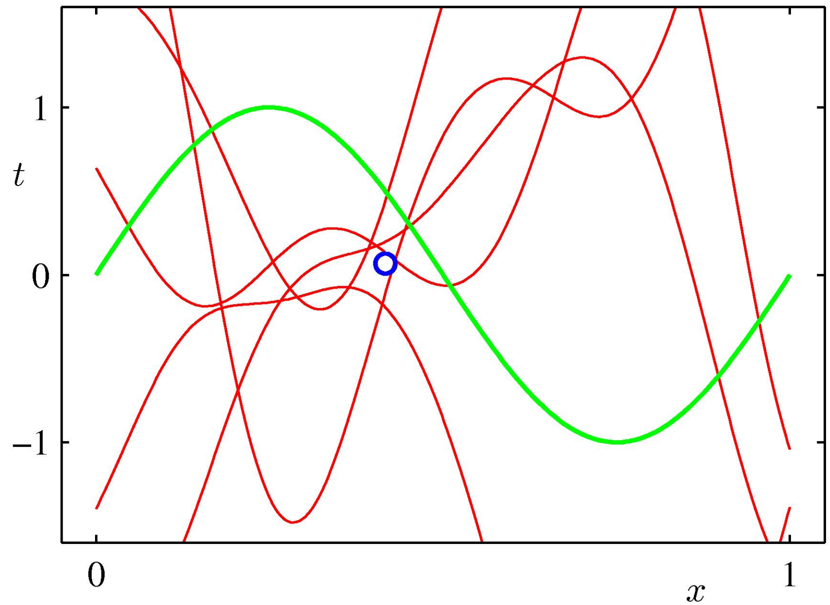
预测分布

示例：正弦数据，9个高斯基函数，1个数据点

$\sin(2\pi x)$



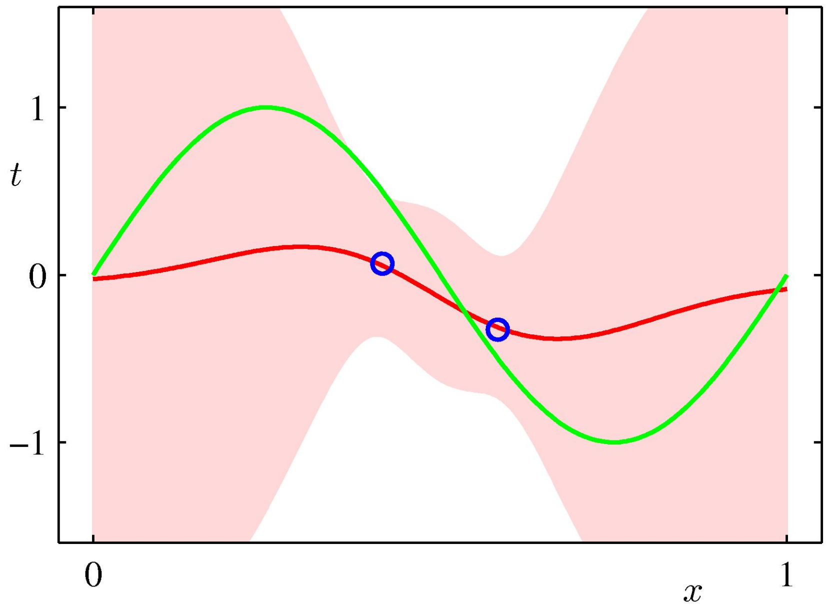
$p(t|\mathbf{t}, \alpha, \beta)$



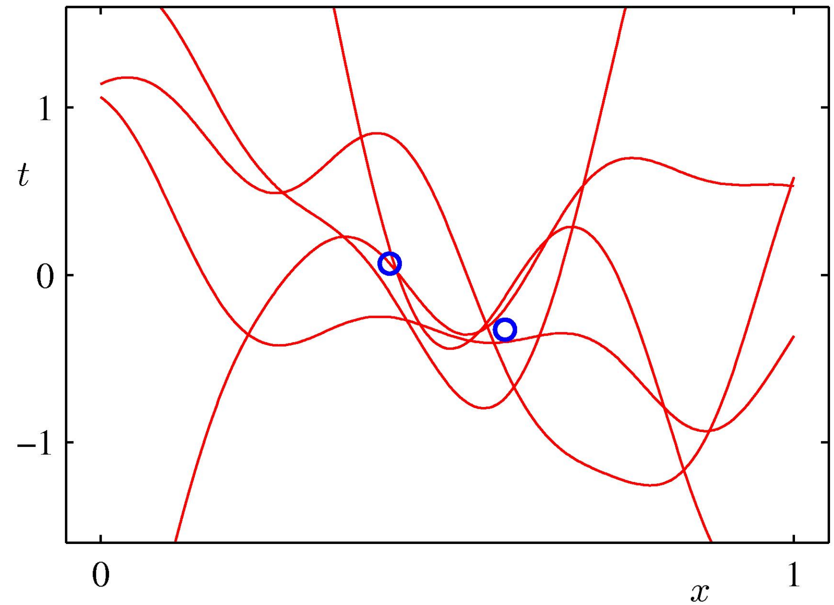
$y(x, \mathbf{w})$

预测分布

示例：正弦数据，9个高斯基函数，2个数据点



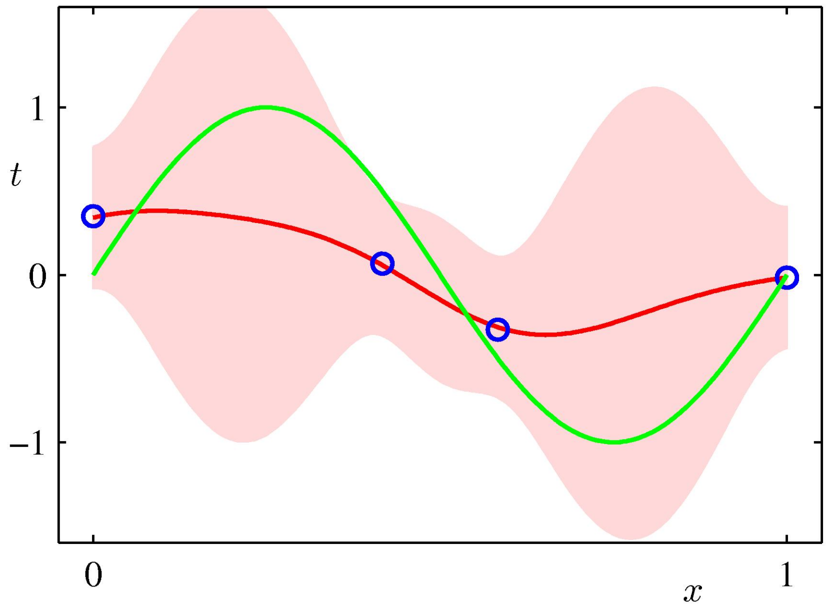
$p(t|\mathbf{t}, \alpha, \beta)$



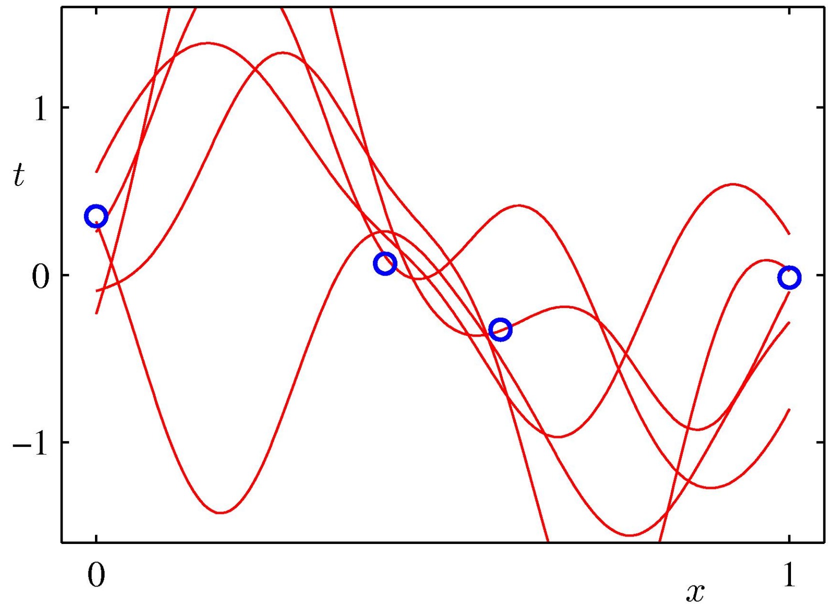
$y(x, \mathbf{w})$

预测分布

示例：正弦数据，9个高斯基函数，4个数据点



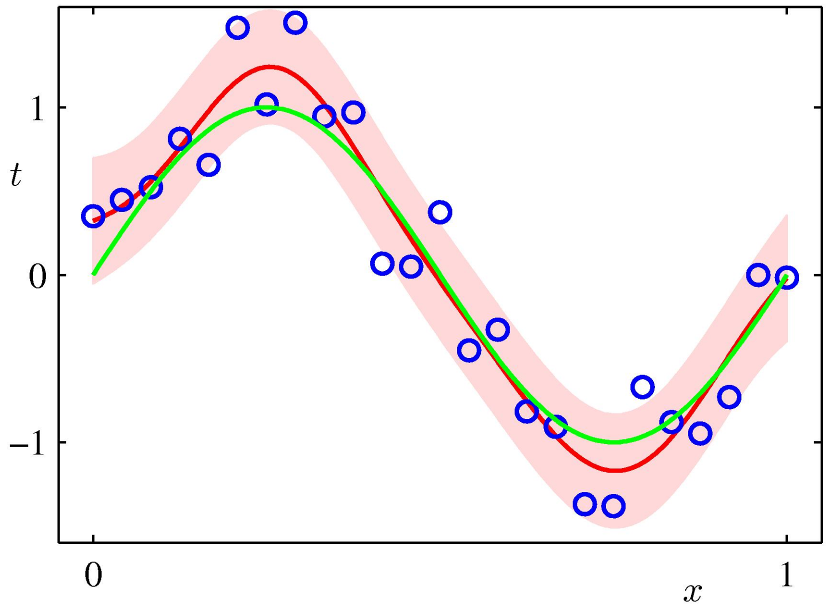
$p(t|\mathbf{t}, \alpha, \beta)$



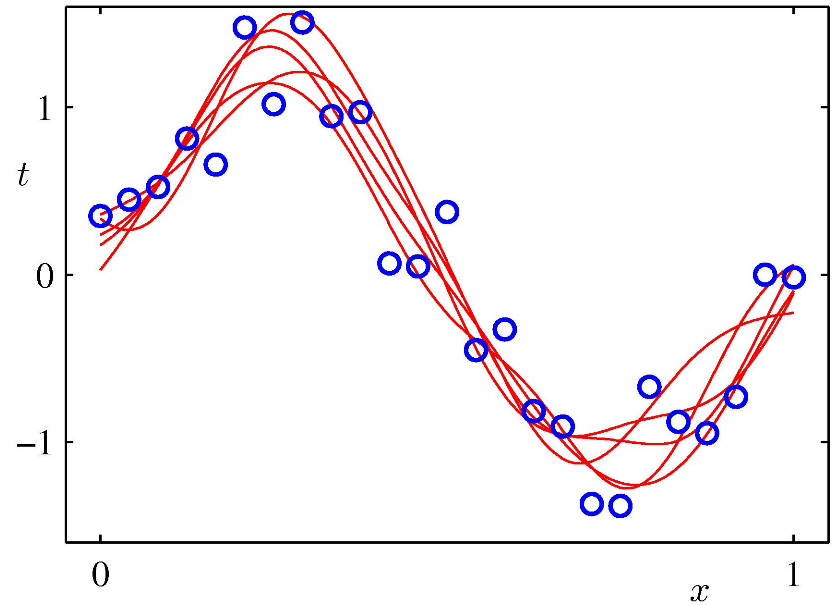
$y(x, \mathbf{w})$

预测分布

示例：正弦数据，9个高斯基函数，25个数据点



$p(t|\mathbf{t}, \alpha, \beta)$



$y(x, \mathbf{w})$

本章小结

1. 回归简介
   - 回归的目的，回归的学习和预测
2. 线性基函数模型
   - 一般形式，几种基函数形式
   - 最大似然和最小二乘
   - 正则化最小二乘
3. 贝叶斯线性回归
   - $\mathbf{w}$的先验分布、后验分布 （最大化）
   - 预测分布