计算建模

实验五 基于RANSAC和最小二乘的直线和曲线拟合

刘绍辉，范晓鹏
计算机科学与技术学院 哈尔滨工业大学
fxp,shliu@hit.edu.cn

2021年秋季

小波去噪模型

◆采用小波变换来去噪的基本思想

➢都是假设噪声与信号相比较，要远小于信号，否则就是噪声是主要成分

2层小波变换后的近似系数

2层小波变换后水平中频小波系数

2层小波变换后垂直中频小波系数

2层小波变换后对角中频小波系数

硬阈值：

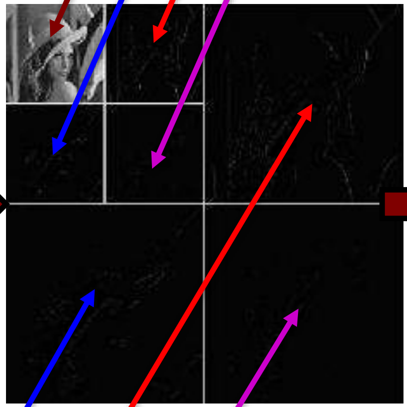
$$y=\begin{cases} y, 如果|y| \ge T \\ 0, 如果|y| < T \end{cases}$$

软阈值：

$$y=\begin{cases} y-T, 如果y \ge T \\ 0, 如果 -T \le y < T \\ y+T, 如果y < -T \end{cases}$$

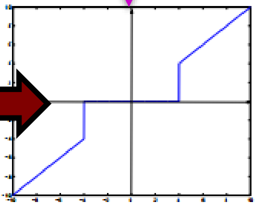
原始图像

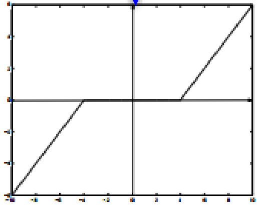
添加噪声后的图像



2层小波变换后的近似图像和小波系数图像

对高频和中频小波系数做阈值化处理，阈值T，例如$T = 0.5$





对阈值化处理之后的小波系数和逼近系数做逆小波变换

去噪声后的图像

2层小波变换水平高频小波系数

2层小波变换垂直高频小波系数

2层小波变换对角高频小波系数

基本内容

- **给定一些点，如何来进行直线拟合？（ransac方法和最小二乘方法）**
  - 如果是根据直线方程+噪声，生成的一些点，如何来拟合直线？如果有很多外点(outlier)如何处理？
  - 如果是从图像中检测出来的边缘点，这些边缘点正好在直线上，如何来进行直线拟合？
  - 如果这些离散点形成了平行直线，如何来拟合出平行直线的方向？
- **具体内容（分别使用ransac方法和最小二乘方法来进行拟合）**
  - 1.直线拟合
    - ✓ 根据直线方程ax+by+c=0,产生(x_i, y_i)随机点，然后增加随机噪声成为:$(\boldsymbol{x_i} + \boldsymbol{n_i}, \boldsymbol{y_i} + \boldsymbol{m_i})$,根据这些点，拟合直线$\boldsymbol{ax} + \boldsymbol{by} + \boldsymbol{c} = \boldsymbol{0}$中的参数
    - ✓ 如果有一系列平行直线$\boldsymbol{ax} + \boldsymbol{by} + \boldsymbol{c_1} = \boldsymbol{0}, \boldsymbol{ax} + \boldsymbol{by} + \boldsymbol{c_2} = \boldsymbol{0}, \boldsymbol{ax} + \boldsymbol{by} + \boldsymbol{c_3} = \boldsymbol{0}$,然后对直线上的点添加类似的噪声，拟合这些平行直线
  - 2.曲线拟合
    - ✓ 自己设计曲线方程，例如圆方程，椭圆方程，然后添加适当的噪声（例如高斯噪声），然后分别采用ransac和最小二乘方法进行拟合
    - ✓ 如果添加一些外点，拟合效果如何？是否有方法改进！

RANSAC

◆基本思想

- 随机选择两点(确定一条直线所需要的最小点集)；由这两个点确定一条线l；
- 根据阈值t，确定与直线l的几何距离小于t的数据点集$S(l)$，并称它为直线l的一致集；
- 重复若干次随机选择，得到直线$l_1, l_2, \ldots, l_n$和相应的一致集$S(l_1), S(l_2), \ldots, S(l_n)$；
- 使用几何距离，求最大一致集的最佳拟合直线，作为数据点的最佳匹配直线

RANSAC

- 推广到一般的模型，估计模型参数p
  - 确定求解模型M，即确定模型参数p，所需要的最小数据点的个数n。由n个数据点组成的子集称为模型M的一个样本；
  - 从数据点集D中随机地抽取一个样本J，由该样本计算模型的一个实例$M_p(J)$，确定与$M_p(J)$之间几何距离<阈值t的数据点所构成的集合，并记为$S(M_p(J))$，称为实例$M_p(J)$的一致集；
  - 如果在一致集$S(M_p(J))$中数据点的个数# $S(M_p(J))$>阈值T, 则用$S(M_p(J))$重新估计模型M，并输出结果；如果# $S(M_p(J))$ < 阈值T，返回到步骤2；
  - 经过K次随机抽样，选择最大的一致集$S(M_p(J))$，用$S(M_p(J))$重新估计模型M，并输出结果

RANSAC

◆实现中需要注意的事项

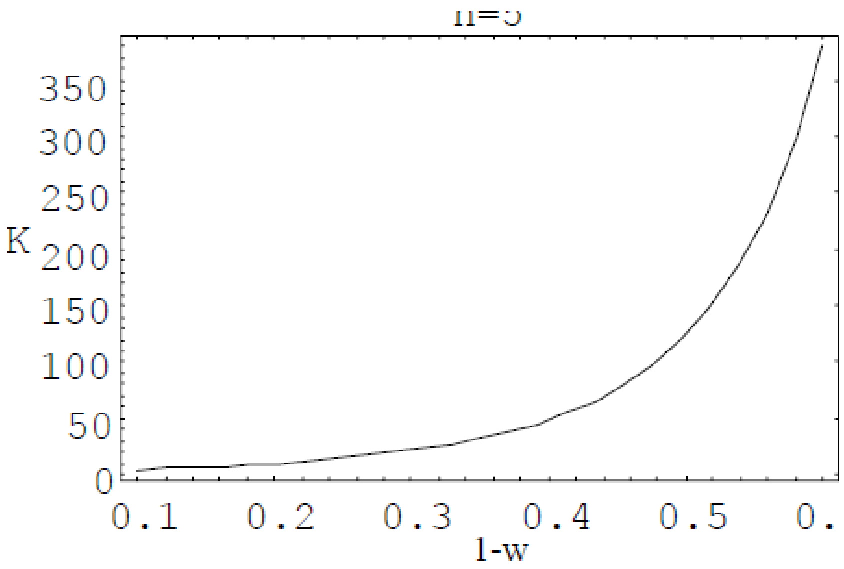
➢抽样次数的考虑

表：样本所含数据点个数 n 与抽样次数 K 的一些对应值($w=0.45$ ， $z=0.02$)

n	2	3	4	5	7	9	13	16	20
K	18	41	94	210	1045	5168	126076	1.38×10^6	3.37×10^7

➢内点比例为w, K 次取中所有样本均为坏样本的概率，从而可得

$$K=\frac{\log z}{\log(1-w^n)}$$



图：抽样次数 K 与外点的比例($1-w$)之间的变化关系

RANSAC

- **距离阈值*t***
  - 经验选取
  - 如果测量误差服从0均值，sigma方差的高斯分布，则可以计算*t*
    - ✓ *因为这时候点到模型几何距离的平方是高斯变量的平方和，服从自由度为n的卡方分布*

$$p_{\chi_n^2}(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}e^{-x/2}x^{n/2-1},0<x<\infty\\0,\text{其它}\end{cases}$$

RANSAC

◆终止阈值

- 不好设置
- 一般规则：如果根据内点比例的估计值，内点数目与一致集大小相当的时候就停止
- 自适应规则

自适应算法(终止 **RANSAC** 抽样)：

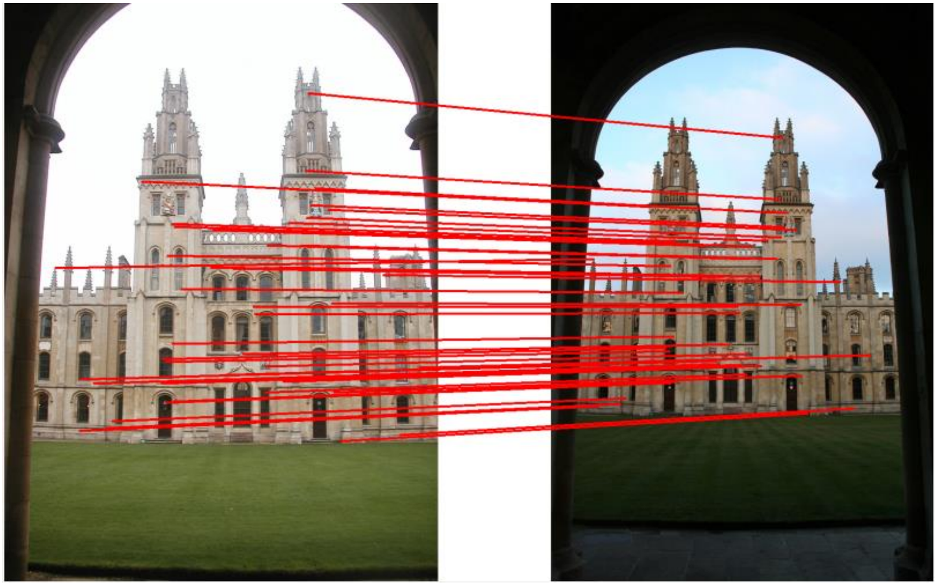
1) 对内点比例作最保守估计 $w = w_0$ (如 $w_0 = 0.1$，这意味着在数据点集中可能有 90%的外点，这确实是一个保守估计)，应用公式$K = \frac{\log z}{\log(1-w^n)}$得到抽样次数 K 的初始值 K_0；

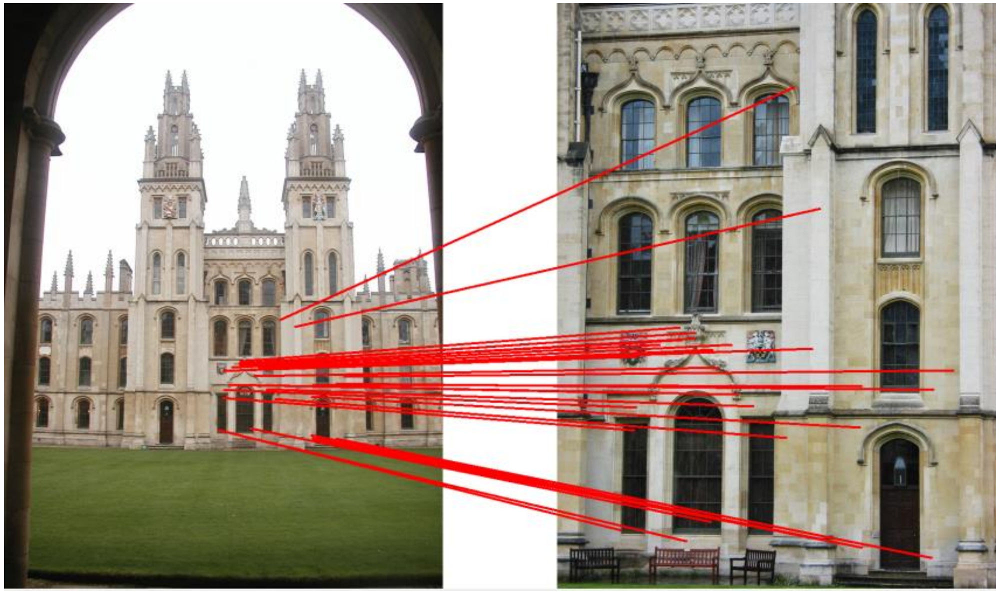
2) 抽样并更新 w_0, K_0：令当前抽样的一致集所含数据点占整个数据点的比例为 w，若 $w > w_0$ 则更新 $w_0 := w$，并且应用公式(17.1.2)更新抽样数 K_0；否则，保持原来的 w_0, K_0；

3) 如果抽样次数已达到或超过 K_0，则终止抽样；否则，返回步骤 2)。

RANSAC

◆实践证明非常有效！

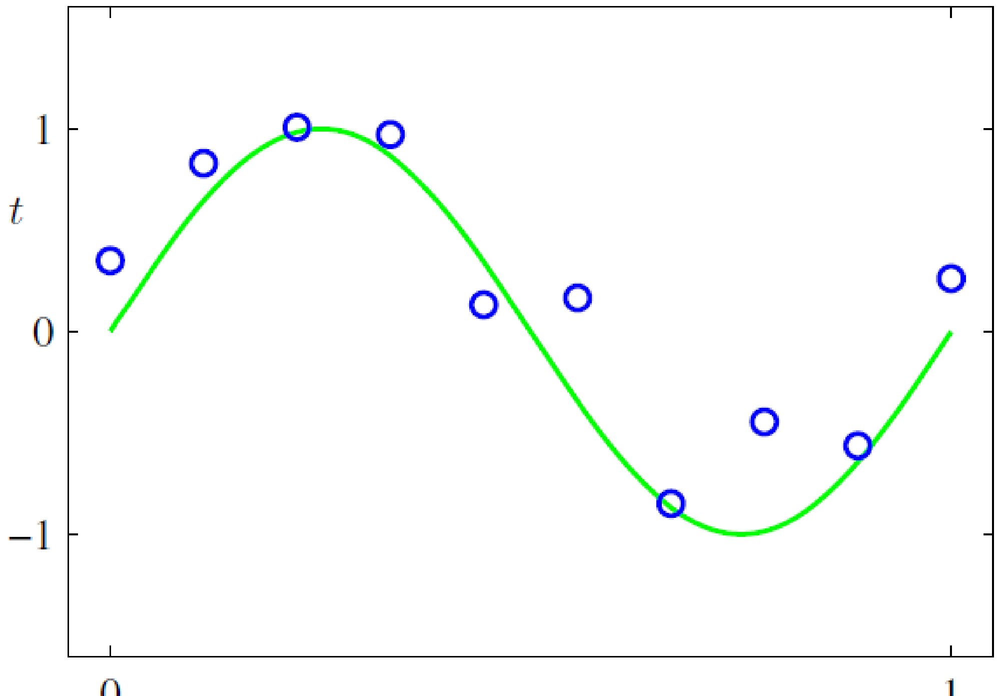




[1] R. Szeliski. Image alignment and stitching: A tutorial. Foundations and Trends in Computer Graphics and Vision, 2(1):1－104, 2006.

最小二乘数据拟合仿真

- **10个训练数据点：圆圈代表 输入变量x的观察值，t代表对应的目标值，绿色线表示$sin(2\pi x)$，用来生成数据**
- **问题：如何预测任意输入变量x所对应的输出值t?注意,绿色线未知**

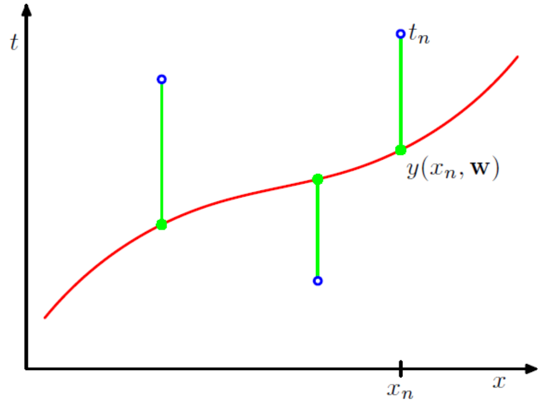


最小二乘数据拟合仿真

◆简单采用多项式来进行逼近

$$y(x,w)=w_0+w_1x+\cdots+w_Mx^M=\sum_{j=0}^{M}w_jx^j$$

◆未知量与目标，函数之间为线性关系，称为线性模型，如何来确定系数w_i呢?



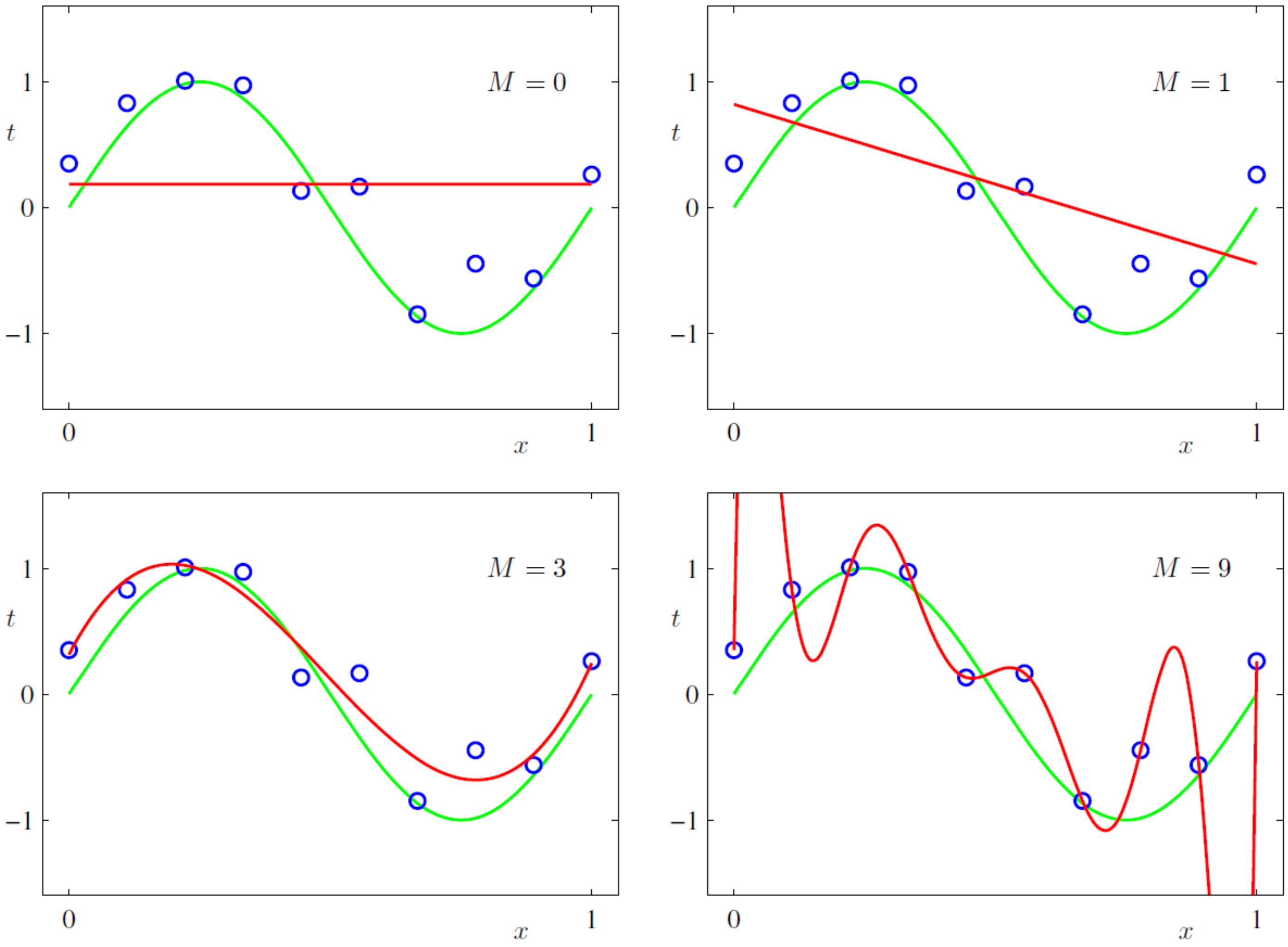
◆通过训练数据，如何做?

◆形式化为一个最优化问题

$$\min E(w), E(w)=\frac{1}{2}\sum^{N}\{y(x_n,w)-t_n\}^2$$

最小二乘数据拟合仿真

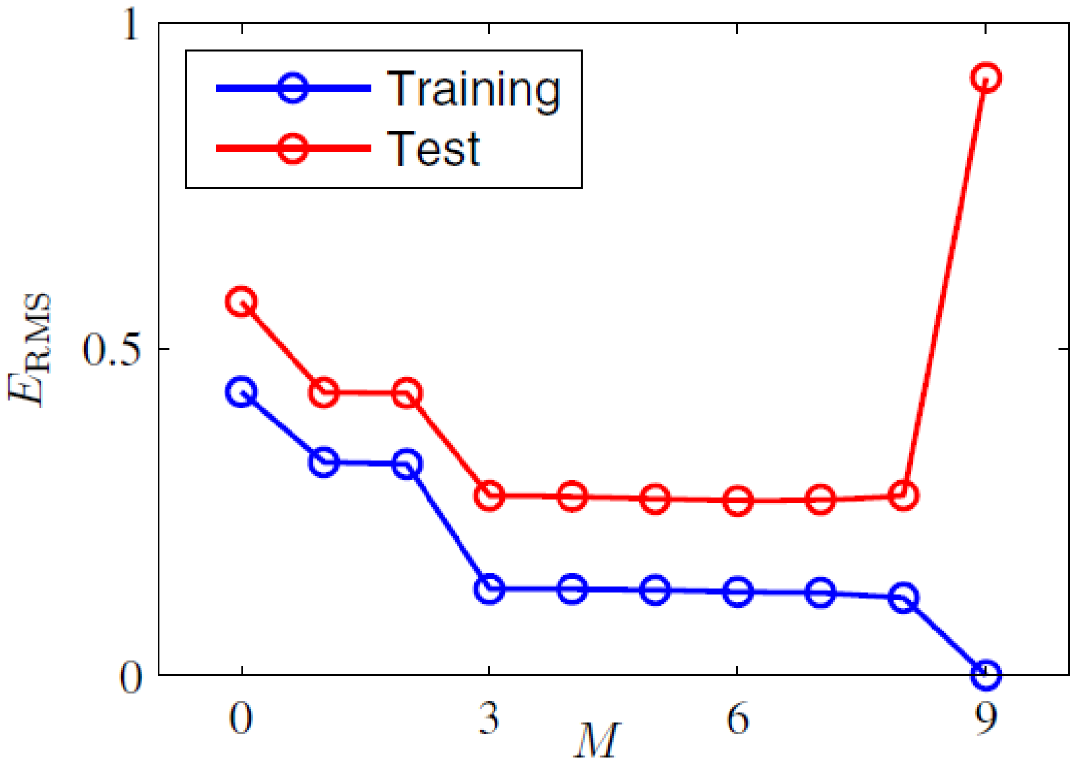
◆如何求解？不同的*M*有什么影响？



最小二乘数据拟合仿真

◆采用$Root-Mean-Squre(RMS)$误差评价：

$$E_{RMS}=\sqrt{2E(w^*)/N}$$



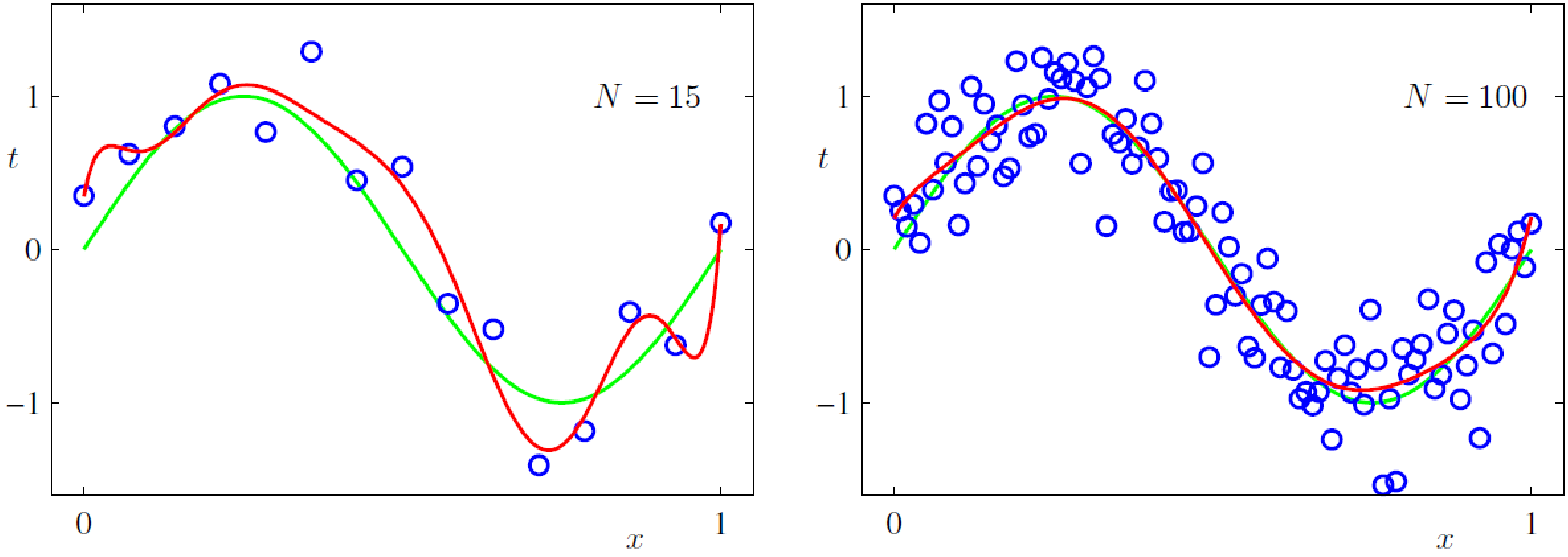
最小二乘数据拟合仿真

◆权系数的情况

	$M=0$	$M=1$	$M=6$	$M=9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

最小二乘数据拟合仿真

◆提升训练数据量($M = 9$)



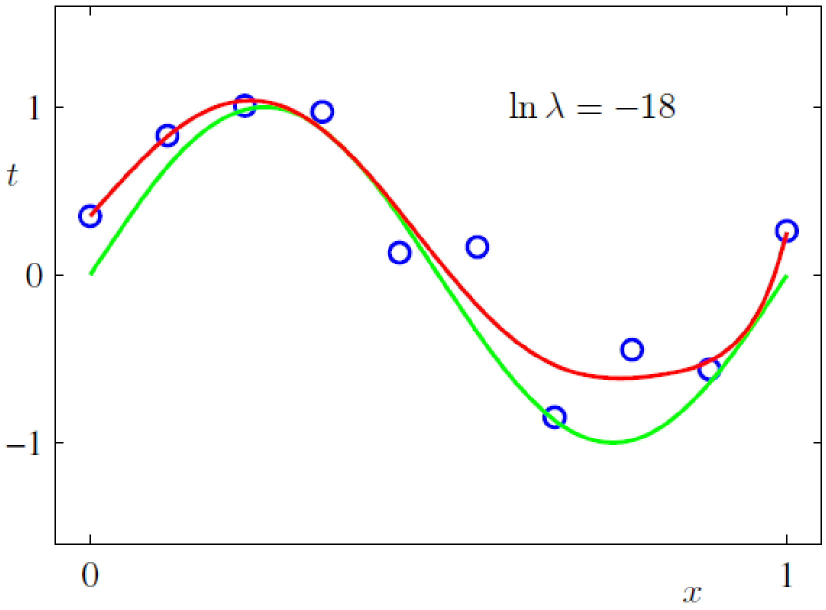
◆一般数据大小为参数的5-10倍左右，那是否需要限制模型参数的数量呢？Deep Learning(深度学习)

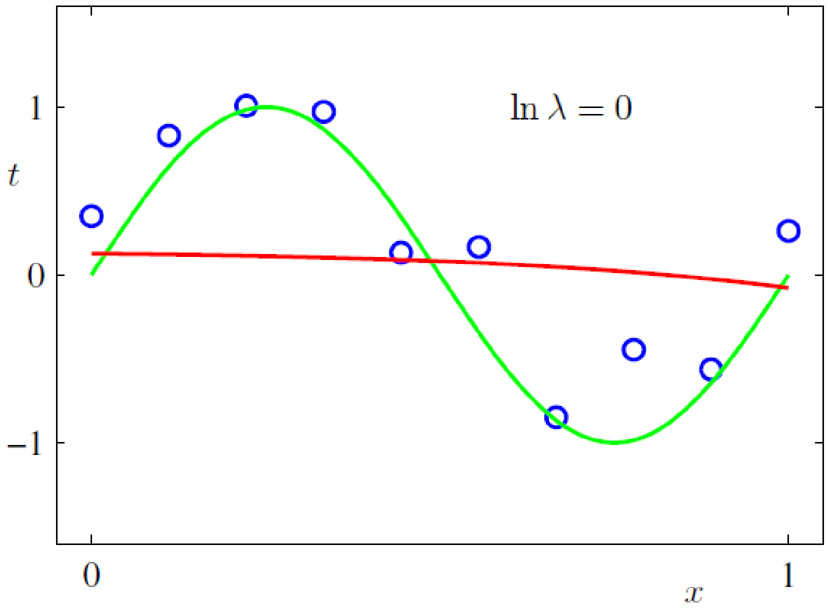
最小二乘数据拟合仿真

- ◆如何改进？正则化方法(Regularization)
- ◆对权系数进行限定，对近似模型进行约束

$$\widetilde{E}(w)=\frac{1}{2}\sum_{n=1}^{N}\{y(x_n,w)-t_n\}^2+\frac{\lambda}{2}||w||_2^2$$

◆





最小二乘数据拟合仿真

◆正则因子对权系数的影响

	$\ln\lambda = -\infty$	$\ln\lambda = -18$	$\ln\lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01