中国科学技术大学
University of Science and Technology of China

数学建模
Mathematical Modeling

陈仁杰
中国科学技术大学

数据拟合模型

问题1：　农业生产实验模型

- 在研究农业生产的试验中，为分析某地区土豆产量与化肥的关系，得到了每公顷地的氮肥的施肥量与土豆产量的对应关系如下

氮肥量(kg)	0	34	67	101	135	202	259	336	404	471
土豆产量(kg)	15.18	21.36	25.72	32.29	34.03	39.45	43.15	43.46	40.83	30.75

问题2： 血药浓度模型

- 通过实验测得一次性快速静脉注射300mg药物后随时间变化的血药浓度数据

t(h)	0.25	0.5	1	1.5	2	3	4	6	8
y(ug/ml)	19.21	18.15	15.36	14.10	12.89	9.32	7.45	5.24	3.01

问题3：牙膏的销售量

问题

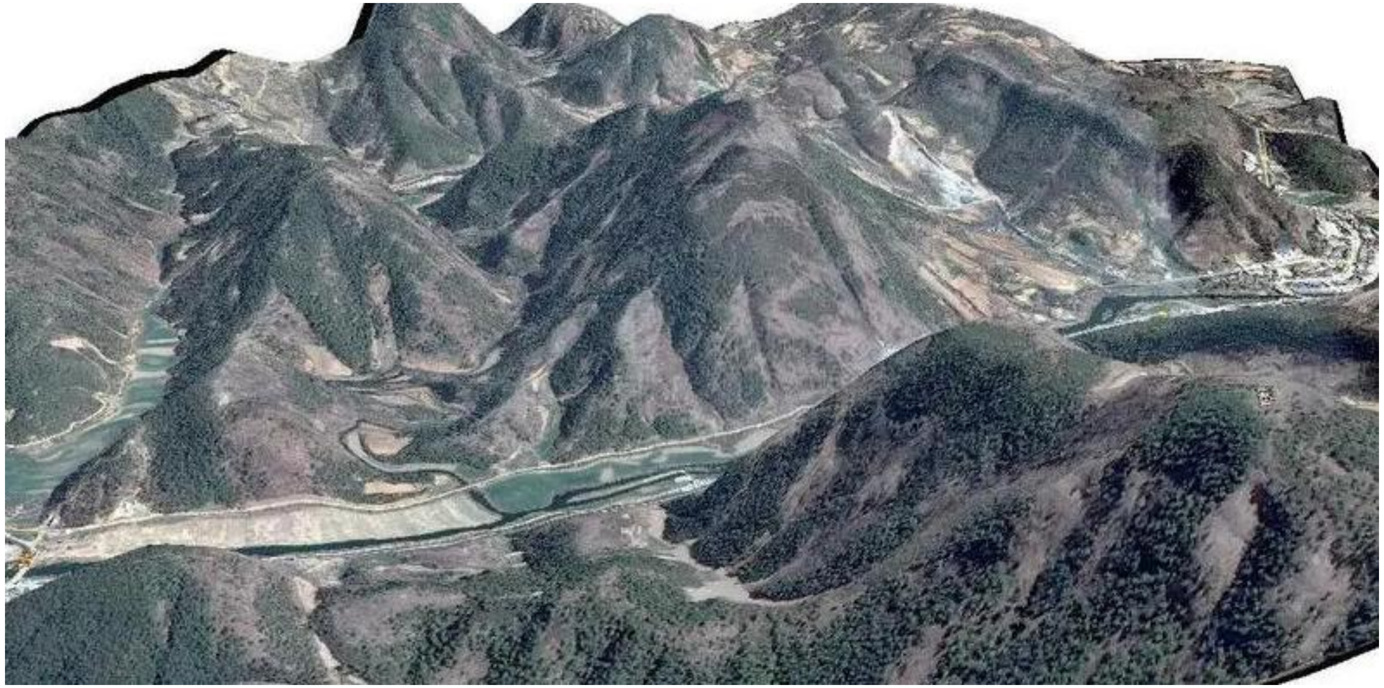
建立牙膏销售量与价格、广告投入之间的模型

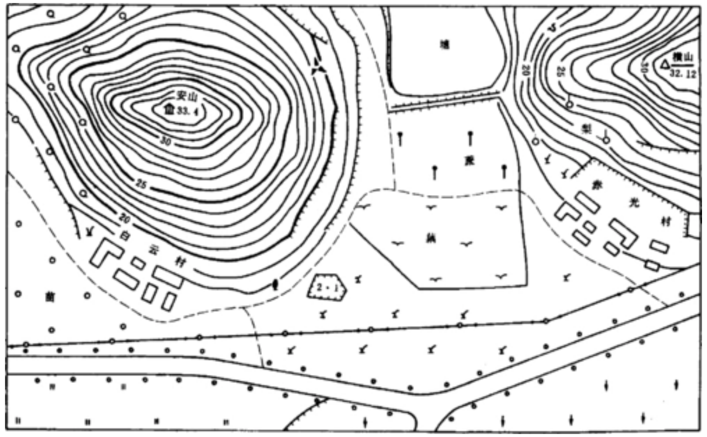
预测在不同价格和广告费用下的牙膏销售量

收集了30个销售周期本公司牙膏销售量、价格、广告费用，及同期其它厂家同类牙膏的平均售价

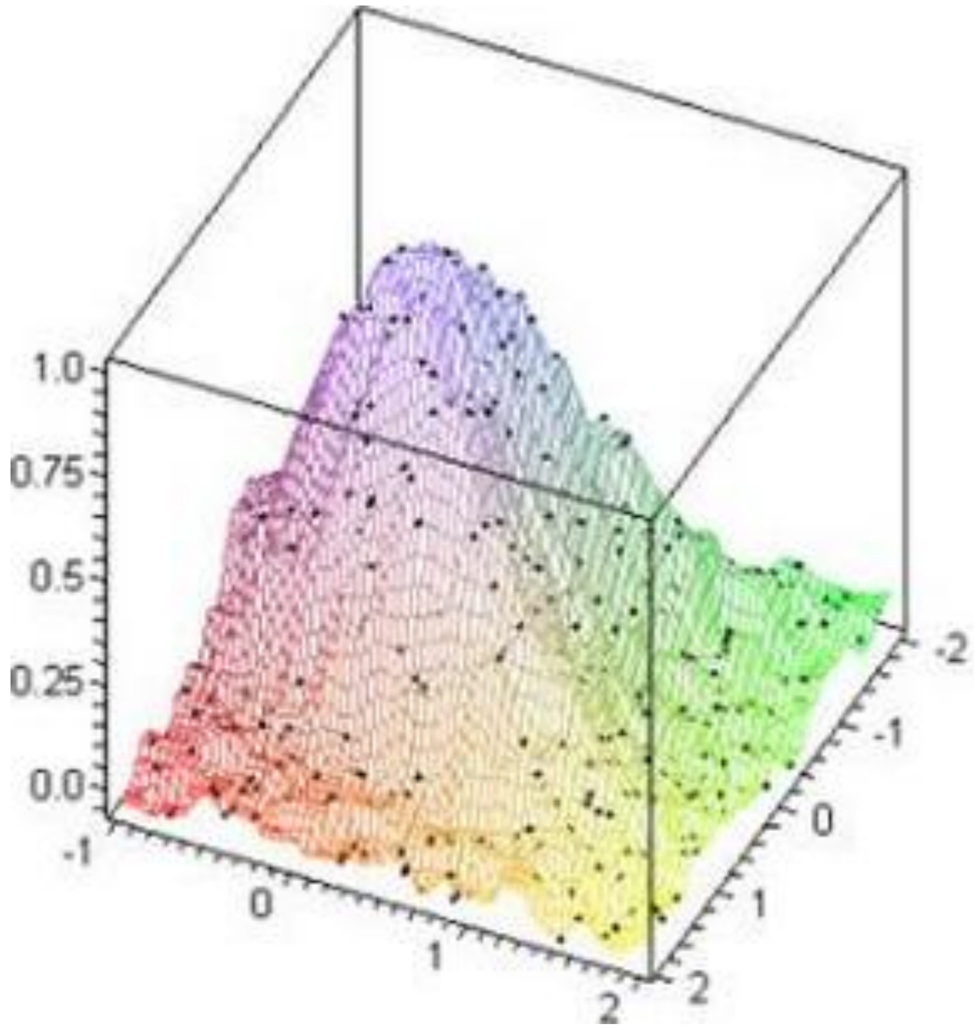
销售周期	本公司价格(元)	其它厂家价格(元)	广告费用(百万元)	价格差(元)	销售量(百万支)
1	3.85	3.80	5.50	-0.05	7.38
2	3.75	4.00	6.75	0.25	8.51
…	…	…	…	…	…
29	3.80	3.85	5.80	0.05	7.93
30	3.70	4.25	6.80	0.55	9.26

问题4：地形高度测量



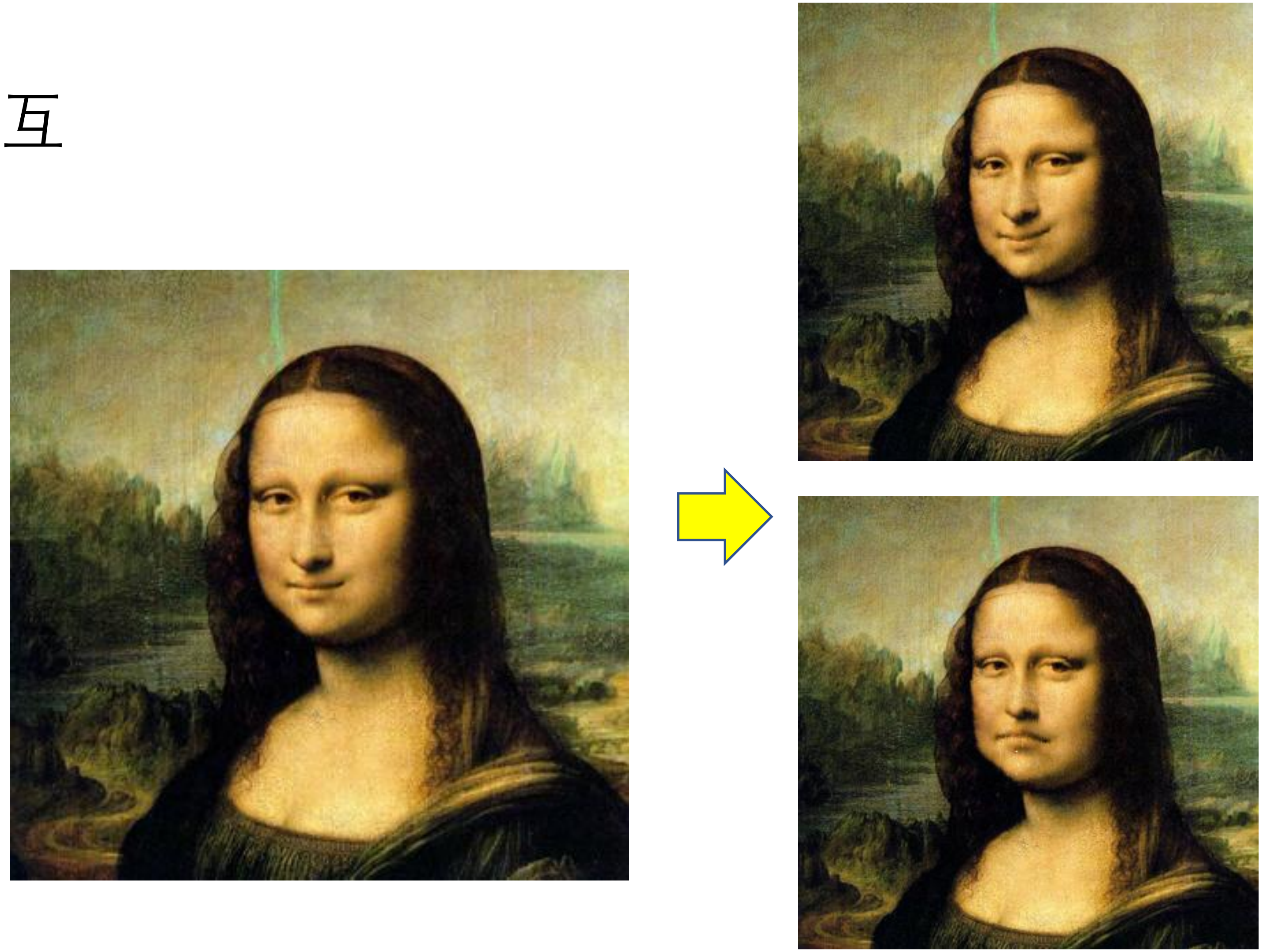


数学模型： 散乱点插值

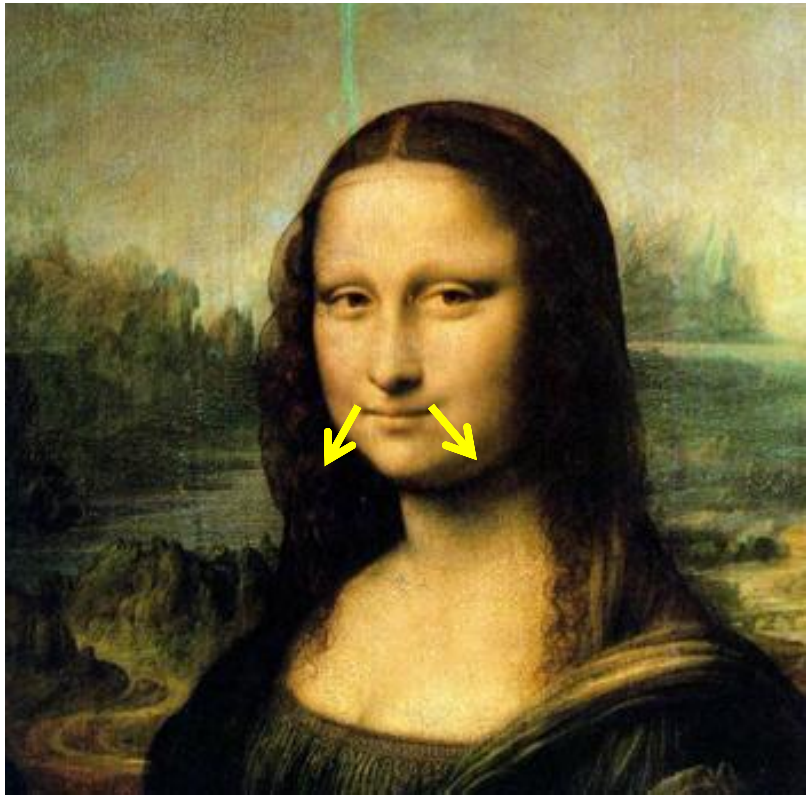


问题5： 图像变形

- 用户交互



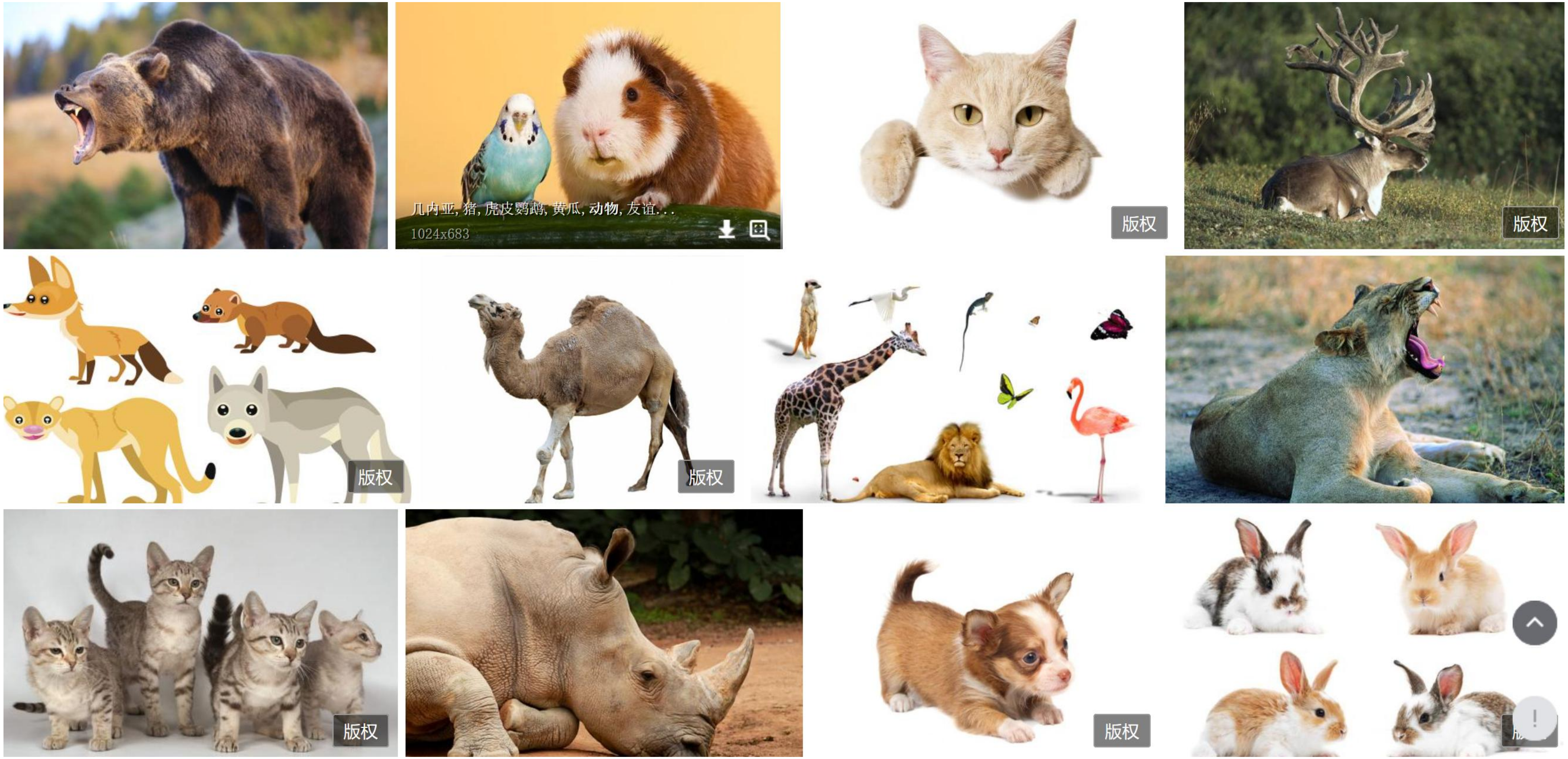
图像变形： 用户交互



数学模型?

0 50 100 150 200 250 300 350 400 450 500
0 50 100 150 200 250 300 350 400 450 500
0 50 100 150 200 250 300 350 400 450 500
50 100 150 200 250 300 350 400 450 500

问题6：图像识别：判断图片中是否有“猫”？



判断图片中是否有“猫”？

- 获得观察数据：标注的图片

判断图片中是否有“猫”？

- 求一个函数f，使得

$f($  $)=$ “1”

$f($  $)=$ “0”

$f($  $)=$ “1”

$f($  $)=$ “0”

如果该函数在观察数据上都成立，则可相信它近似反映（逼近）了规律

$f($  $)=$ “1”

$f($  $)=$ “0”

$f($  $)=$ “1”

$f($  $)=$ “0”

将输入表达为同一维度的数据

1000x1000的矩阵

$$\begin{pmatrix} p_{1,1} \\ p_{1,2} \\ \\ p_{i,j} \\ \\ p_{n,n} \end{pmatrix} \xrightarrow{f} 0/1$$

维数太高！

Find a function:

1000x1000的矩阵

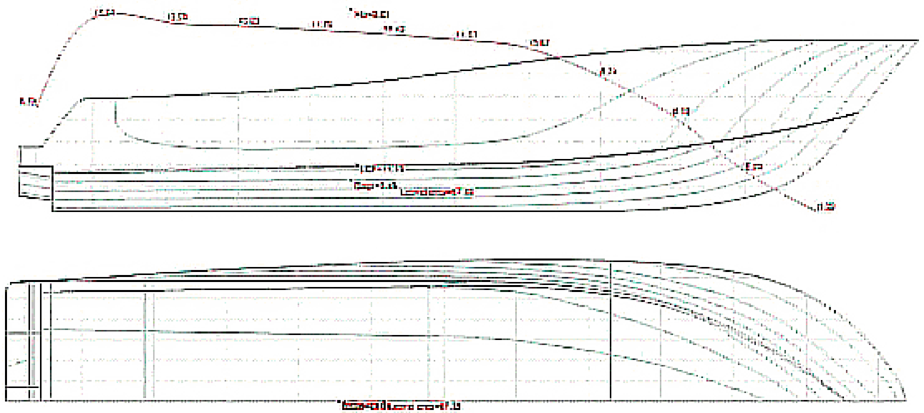
$$\begin{pmatrix} p_{1,1} \\ p_{1,2} \\ \\ p_{i,j} \\ \\ p_{n,n} \end{pmatrix}$$

特征向量：降维、抽象

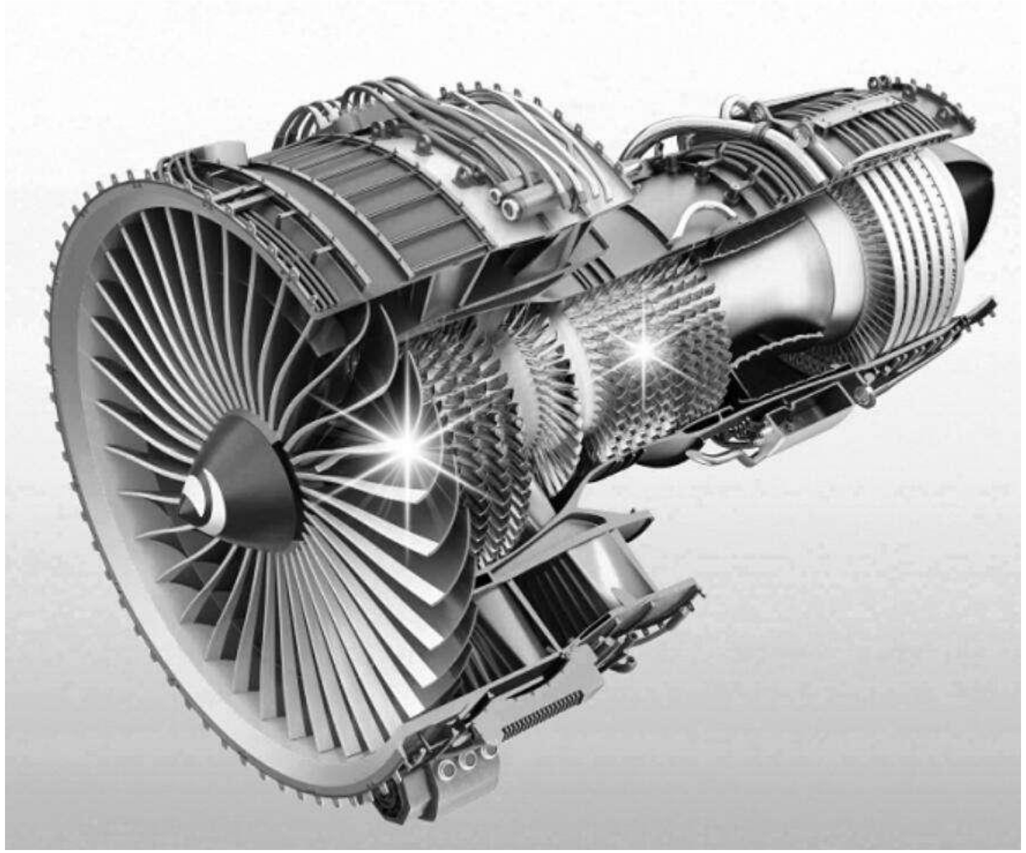
$$\begin{pmatrix} r \\ g \\ b \\ n \\ A \\ D \end{pmatrix} \xrightarrow{f} 0/1$$

问题7：求船型曲线

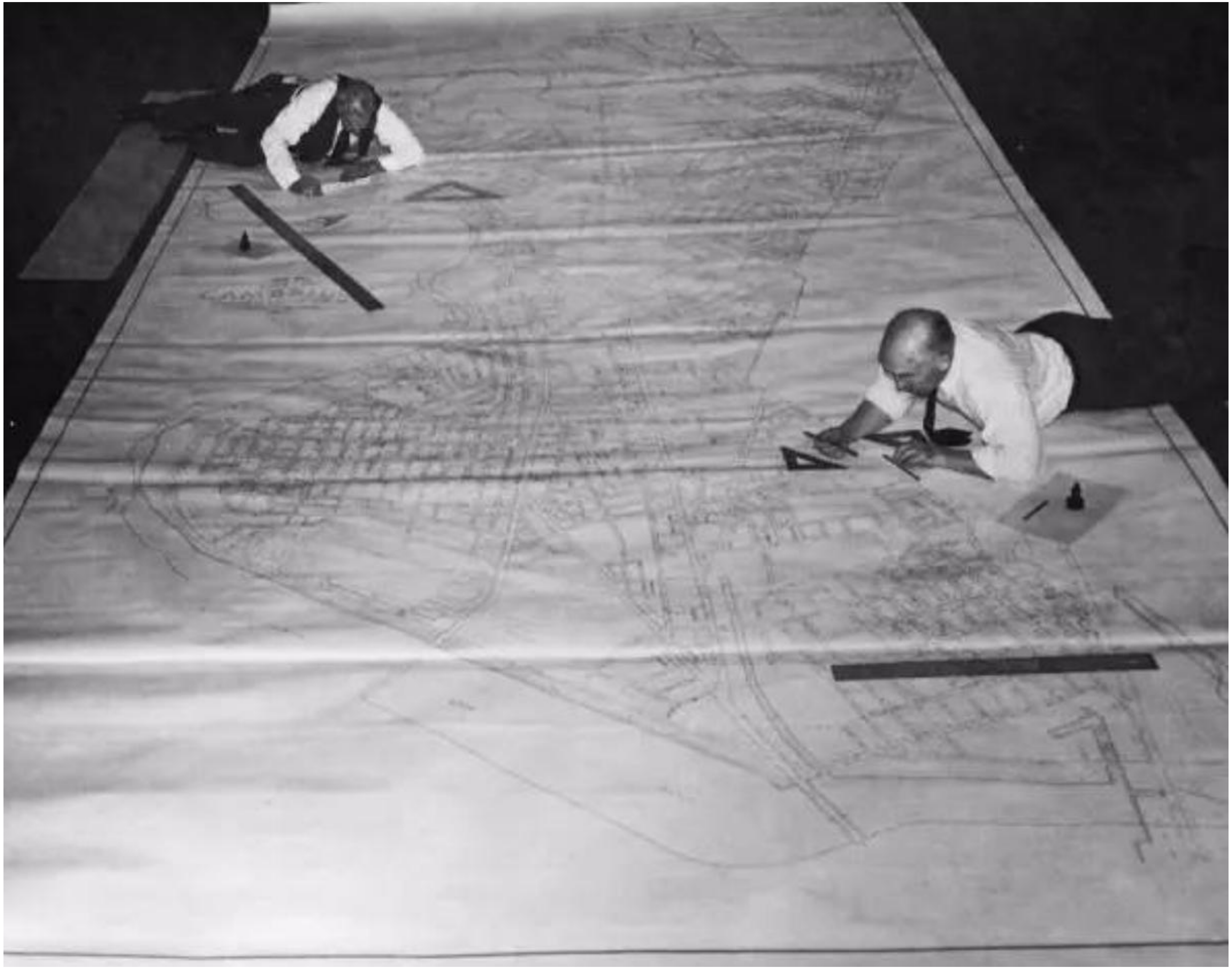
- 已知：某条船的侧面投影曲线图
- 求：该投影轮廓线的表达函数？
- 方法：
  - 从投影曲线上描（采样）一系列点
  - 找一个函数拟合这些采样点



工业产品

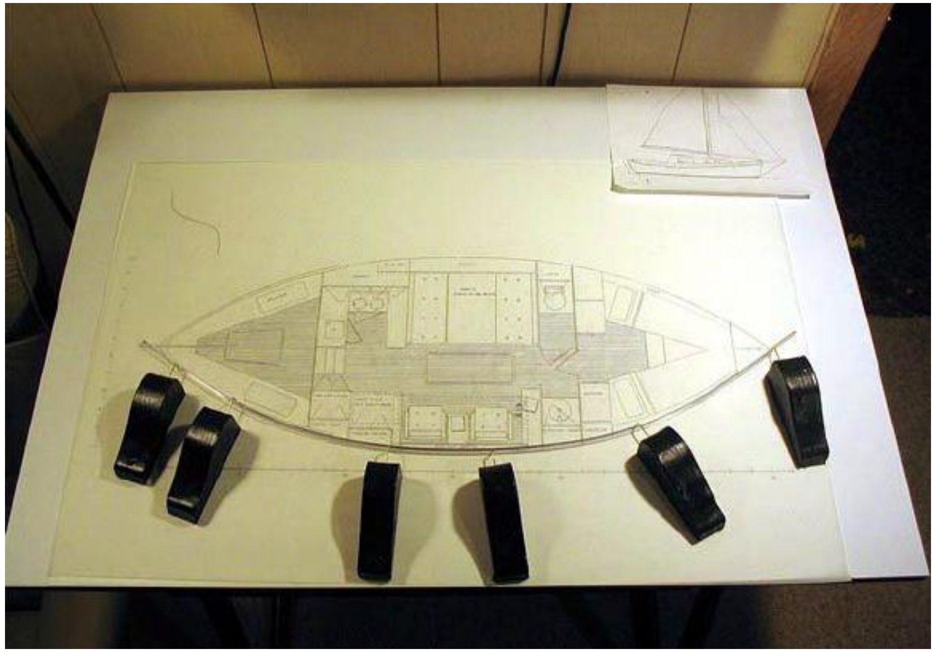


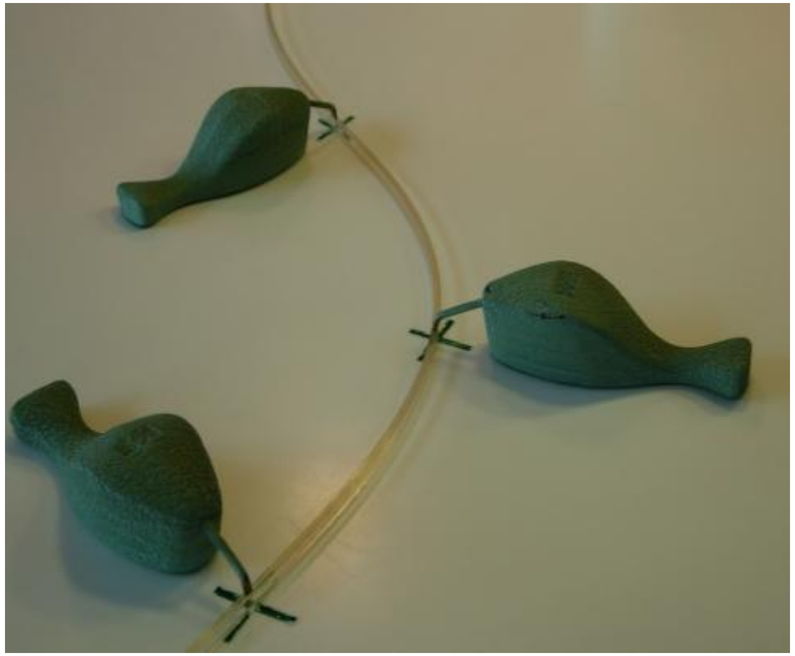
没有电脑之前，
设计师是如何画图设计的？



样条与压铁

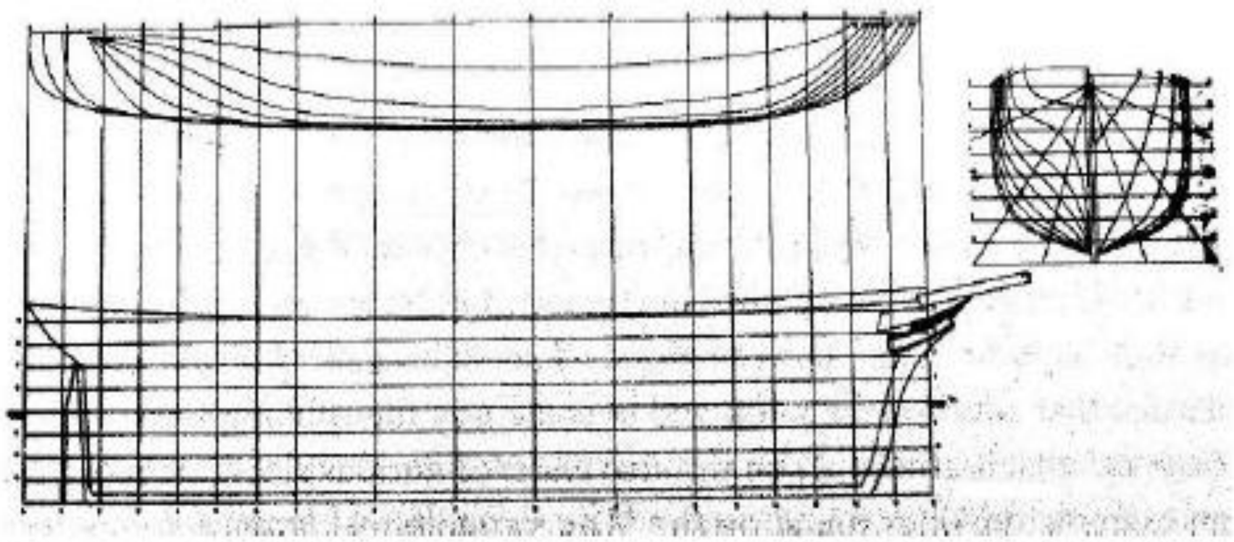
- 样条：具有一定弹性的软木条
- 压铁：较重的铁块，用来固定样条所经过的点



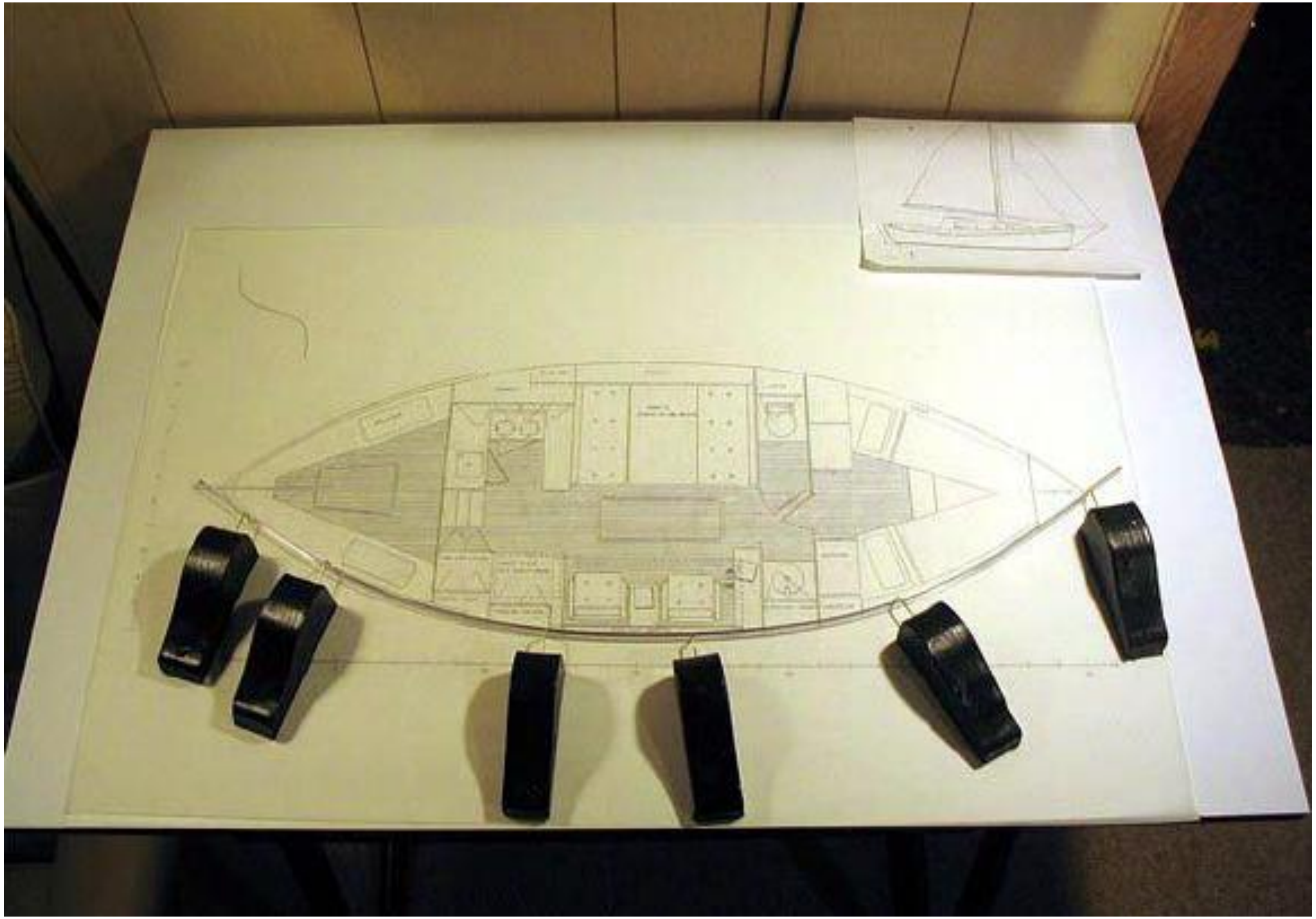


自由曲线

- 在产品初始设计阶段，描述其外形的曲线或曲面常常只有大致形状或只知道它通过一些空间点列，称为**型值点。**



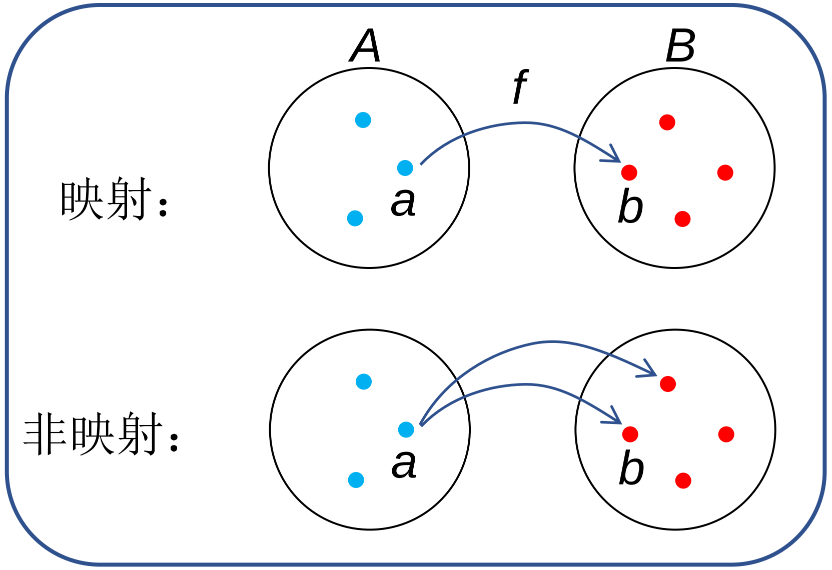
样条所形成的曲线的数学表达？

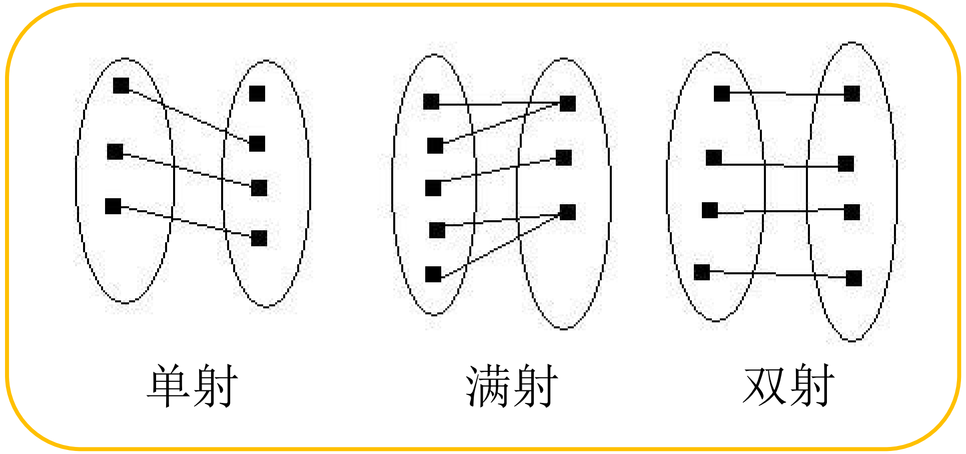


映射与函数

映射 (mapping)

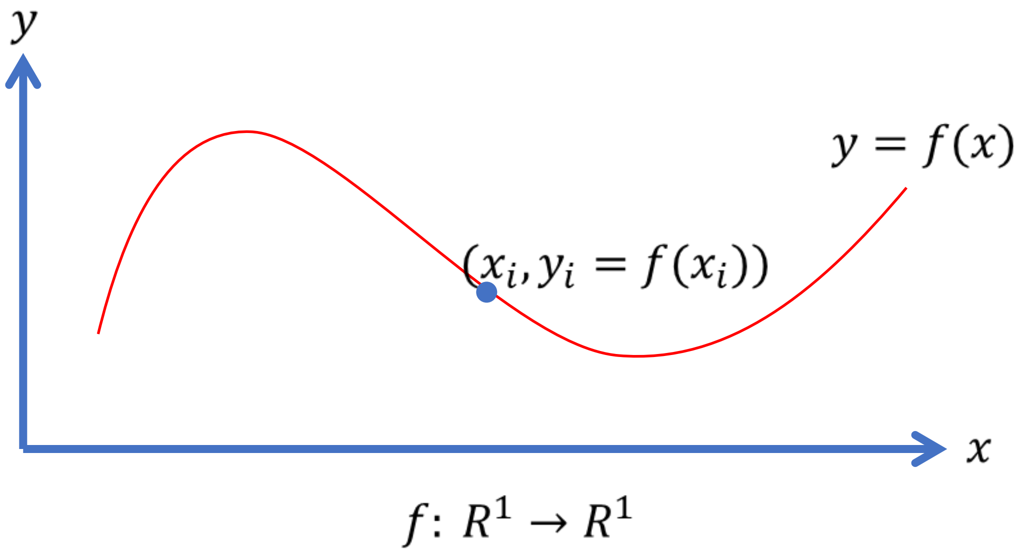
- 两个非空集合A和B的映射$f: A \to B$：对A中的任何一个元素a，有唯一的一个B中的元素b与之对应，记为$f(a) = b$
  - b称为a的象，a称为b的原象
  - A称为定义域，B称为值域





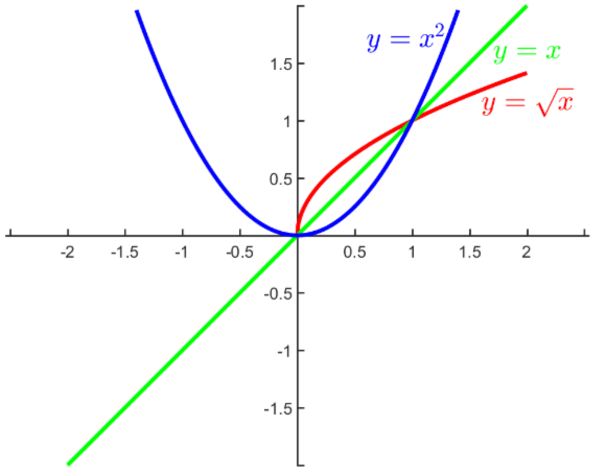
函数 (Function)

- 非空实数集之间的映射称为（一元）函数$y = f(x)$，或变换
- 函数的图像（函数的可视化）：所有有序对$(x, f(x))$组成的集合

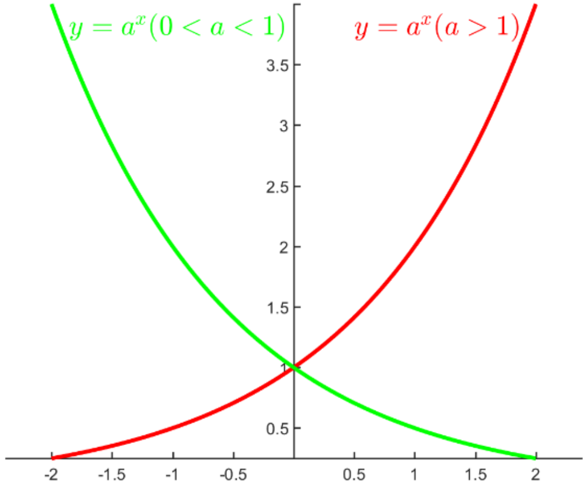


一元函数

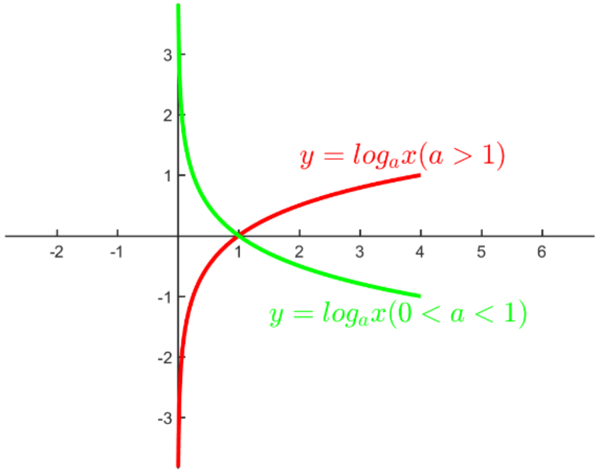
- 有哪些函数?
  - 幂函数
  - 三角函数
  - 对数函数
  - 指数函数
  - 三角函数
  - 反三角函数
  - …



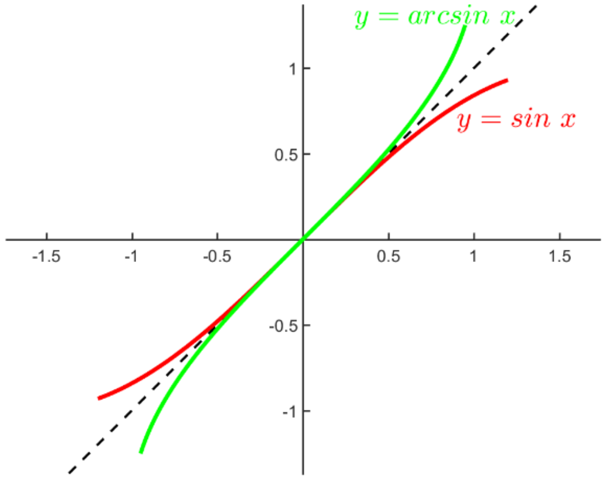
幂函数



指数函数



对数函数



(反)三角函数

函数的集合（函数空间）

- 用若干简单函数（“基函数”）线性组合张成一个函数空间
  - $L = span\{f_1, f_2, \dots, f_n\} = \{\sum_{i=1}^{n} a_i f_i(x) \,|a_i \in R\}$
  - 每个函数就表达（对应）为n个实数，即系数向量$(a_1, a_2, \dots, a_n)$

幂基 $\{x^k, k = 0,1, \dots, n\}$

$$f(x) = \sum_{k=0}^{n} w_k x^k$$

多项式函数空间

三角函数基

$$f(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

三角函数空间

空间的完备性：这个函数空间是否可以表示（逼近）任意函数？

赋范空间

- 内积诱导范数、距离
  - $\langle f, g \rangle = \int_a^b f(x)g(x)dx$
- 度量空间：可度量函数之间的距离
  - Lp范数
- 赋范空间+完备性=巴拿赫空间
- 内积空间（无限维）+完备性=希尔伯特空间

万能逼近定理：Weierstrass逼近定理

- 定理1：闭区间上连续函数可用多项式级数一致逼近
- 定理2：闭区间上周期为2π的连续函数可用三角函数级数一致逼近

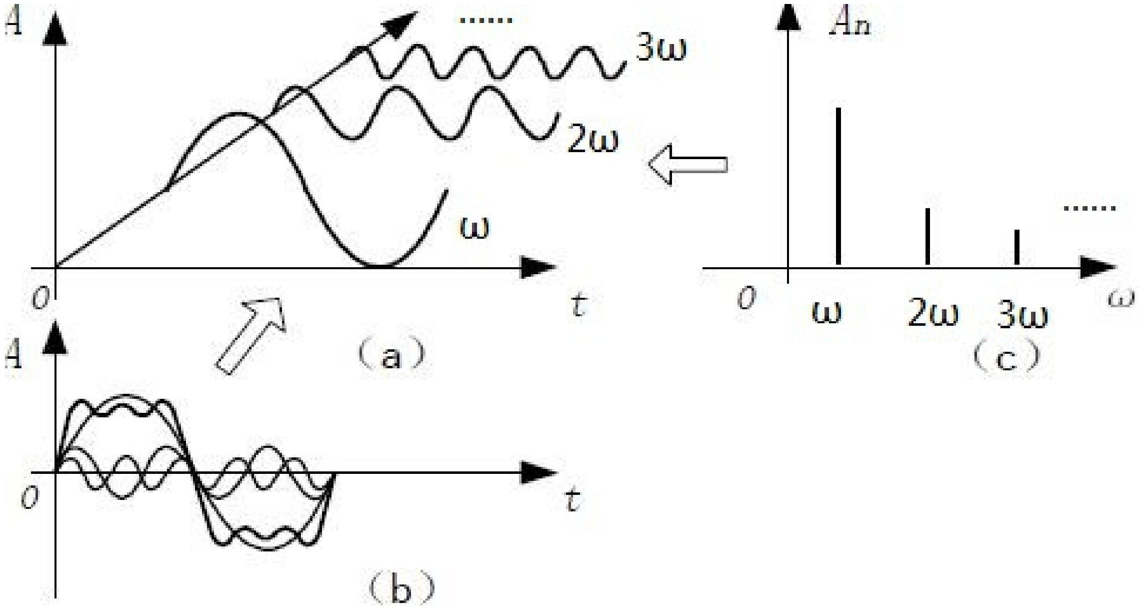
对 $[a, b]$上的任意连续函数g，及任意给定的$\varepsilon > 0$，必存在n次代数多项式$f(x) = \sum_{k=0}^{n} w_k x^k$，使得

$$\min_{x \in [a,b]} |f(x) - g(x)| < \varepsilon.$$

傅里叶级数

$$f(t) = A_0 + \sum_{n=1}^{\infty} [a_n cos(n\omega t) + b_n sin(n\omega t)]$$

$$\boxed{f(t) = A_0 + \sum_{n=1}^{\infty} A_n sin(n\omega t + \psi_n)}$$



更复杂的函数：函数复合

$$\underbrace{x \xrightarrow{f_0} x^1 \xrightarrow{f_1} x^2 \longrightarrow \cdots \longrightarrow x^k \xrightarrow{f_k}}_{f} y$$

$$f = f_k \circ f_{k-1} \circ \cdots \circ f_0$$

$$\frac{1}{1+\left(\frac{2x}{1+x^2}\right)^2}. \qquad f(x)=\frac{e^x}{x^2+1}$$

问题：如何求满足要求的函数？

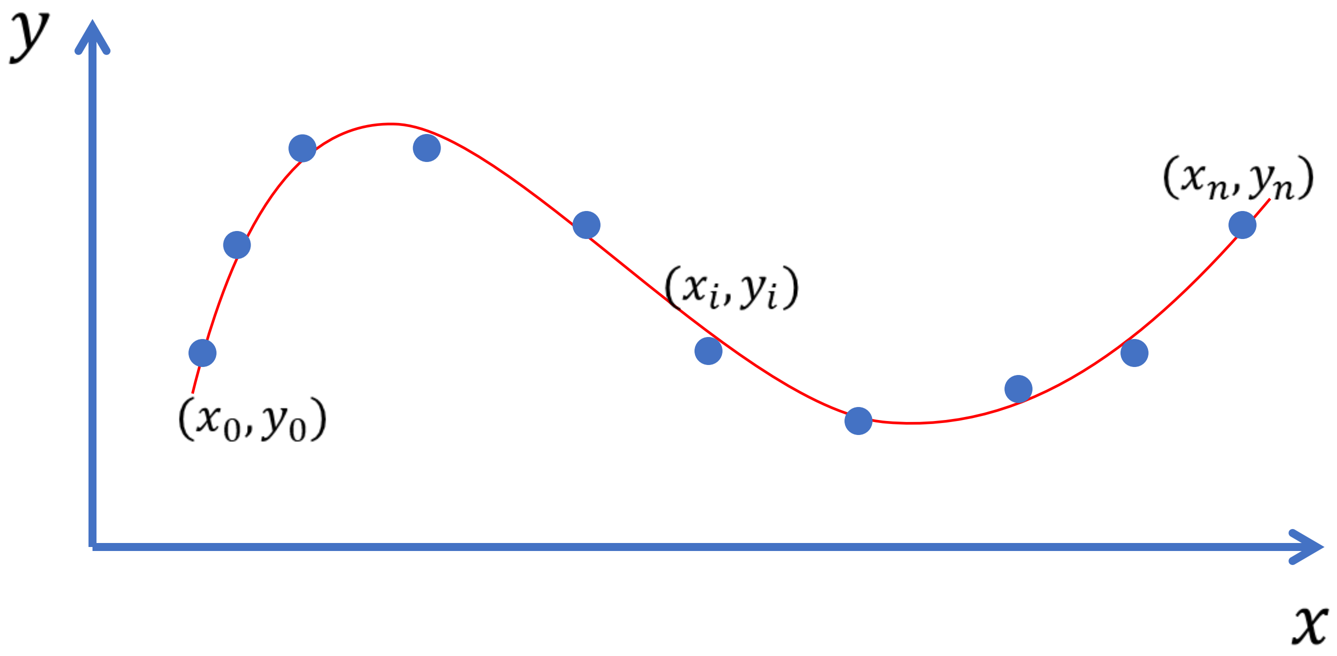
- 大部分的实际应用问题
  - 可建模为：找一个映射/变换/函数
  - 输入不一样、变量不一样、维数不一样
- 如何找函数的三步曲：
  - 到哪找？
    - 确定某个函数集合/空间
  - 找哪个？
    - 度量哪个函数是好的/“最好”的
  - 怎么找？
    - 求解或优化：不同的优化方法与技巧，既要快、又要好…
- 【注】这里先暂时限定为单变量的函数形式

数据拟合

拟合(Fitting)问题

- 输入：一些观察的数据点
- 输出：反映这些数据规律的函数

$$y = f(x)$$



$$x \in R^1, y \in R^1$$

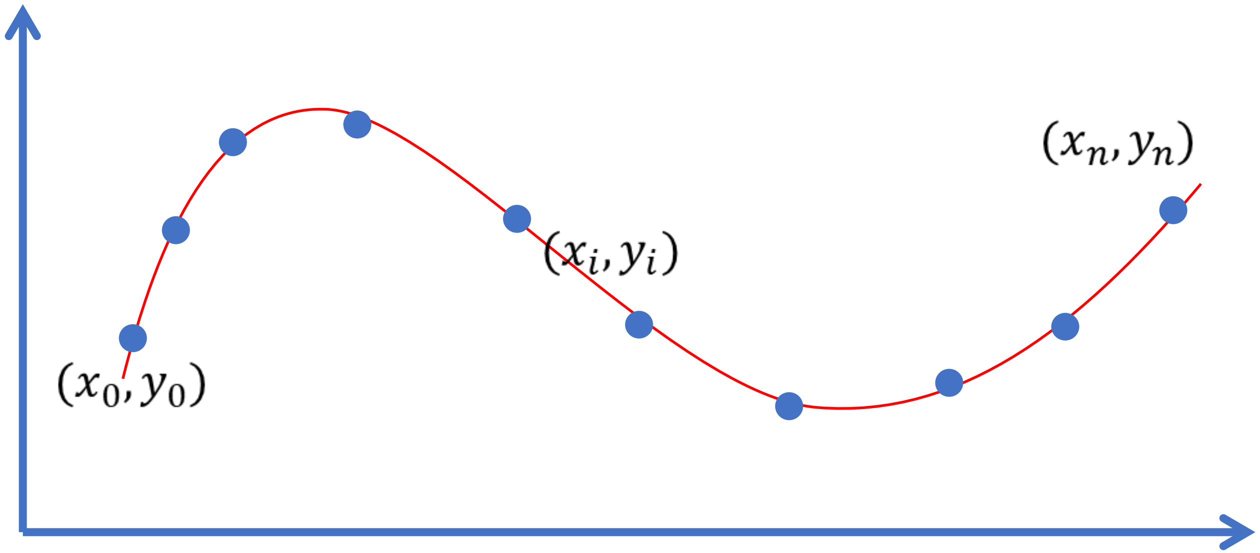
1. 到哪找？

- 选择一个函数空间
  - 线性函数空间$A = span\{B_0(x), \dots, B_n(x)\}$
    - 多项式函数$span\{1, x, x^2, \dots, x^n\}$
    - RBF函数
    - 三角函数
- 函数表达为
  - $f(x) = \sum_{k=0}^{n} a_k B_k(x)$
  - 求$n + 1$个系数$(a_0, \dots, a_n)$

待定系数

2. 找哪个？

- 目标1：函数经过每个数据点（**插值**）
  - $y_i = f(x_i), i = 0,1,\ldots,n$



3. 怎么找?

- 目标1：每个数据点都要插值（零误差）
  - $y_i = f(x_i), i = 0,1,\dots,n$
- 联立，求解线性方程组：
  - $\sum_{k=0}^{n} a_k B_k(x_i) = y_i, i = 0,1,\dots,n$
    - 求解$(n+1)\times(n+1)$线性方程组
    - n次Langrange插值多项式
  - 病态问题：系数矩阵条件数高时，求解不稳定

Lagrange插值函数

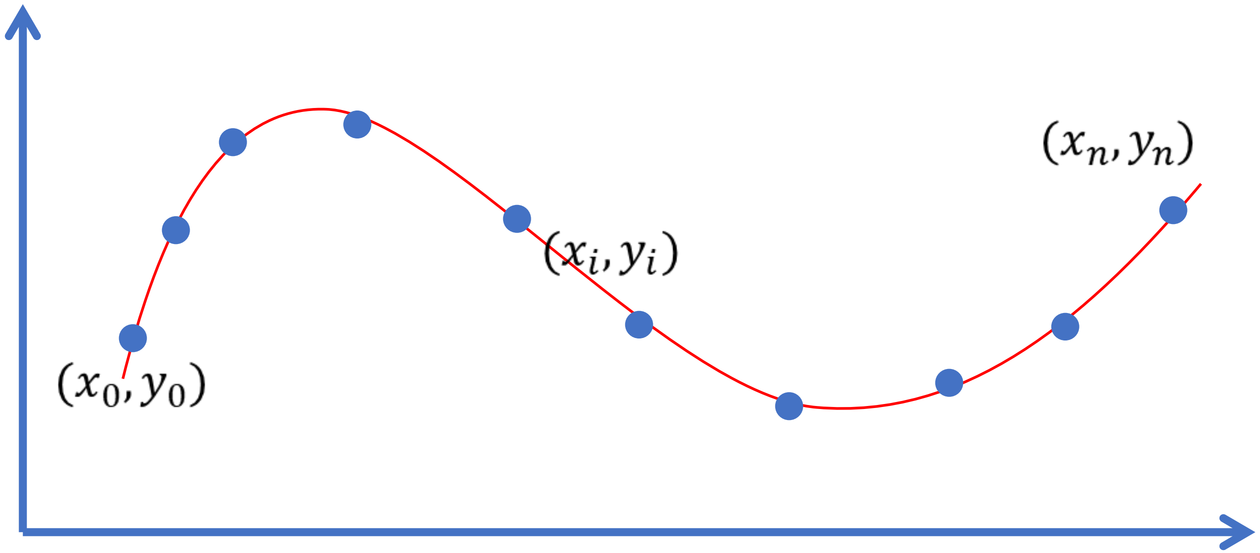
- 插值$n+1$个点、次数不超过n的多项式是存在而且是唯一的
  - （ $n+1$个变量，$n+1$个方程）

$$p_k(x)=\prod_{i\in B_k}\frac{x-x_i}{x_k-x_i}$$

插值函数的自由度 = 未知量个数 – 已知量个数

2-2. 找哪个？

- 目标2：函数尽量靠近数据点（**逼近**）
  - $\min \sum_{i=0}^{n}(y_i - f(x_i))^2$



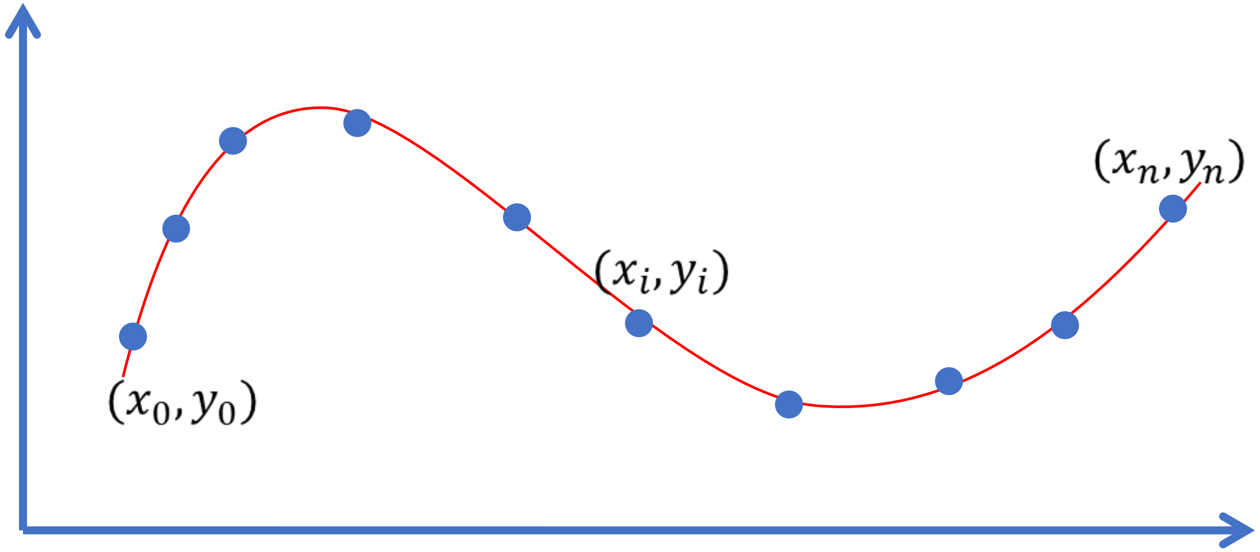
3-2. 怎么找?

- 目标2：函数尽量靠近数据点（逼近）
  - $min\sum_{i=0}^{n}(y_i - f(x_i))^2$
- 对各系数求导，得法方程（线性方程组）
  - $AX = b$

最小二乘法

- 问题：
  - 点多，系数少？
  - 点少，系数多？

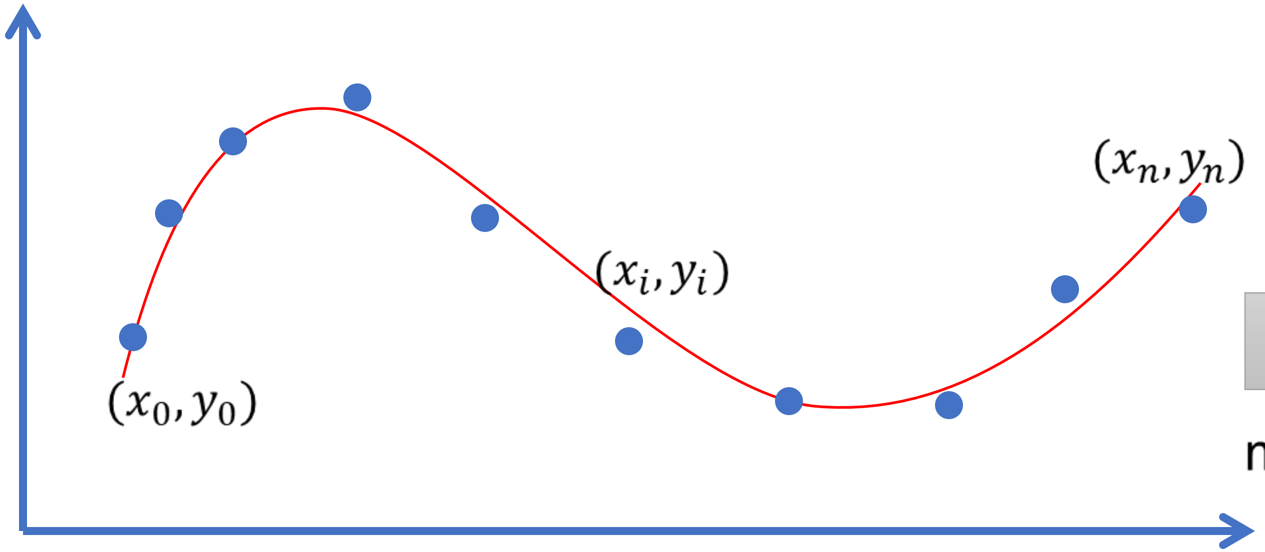
Recap：拟合--插值或逼近



零误差

插值

$y_i = f(x_i)$



最小平方误差

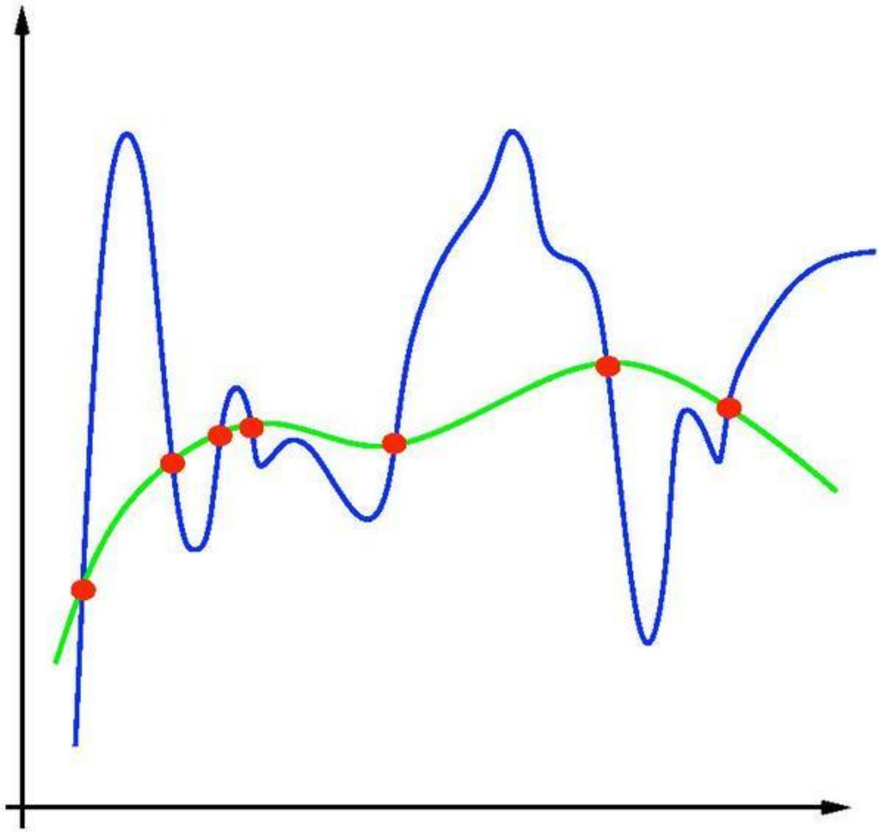
（最小二乘拟合）

逼近

$\min \sum_{i=0}^{n} (y_i - f(x_i))^2$

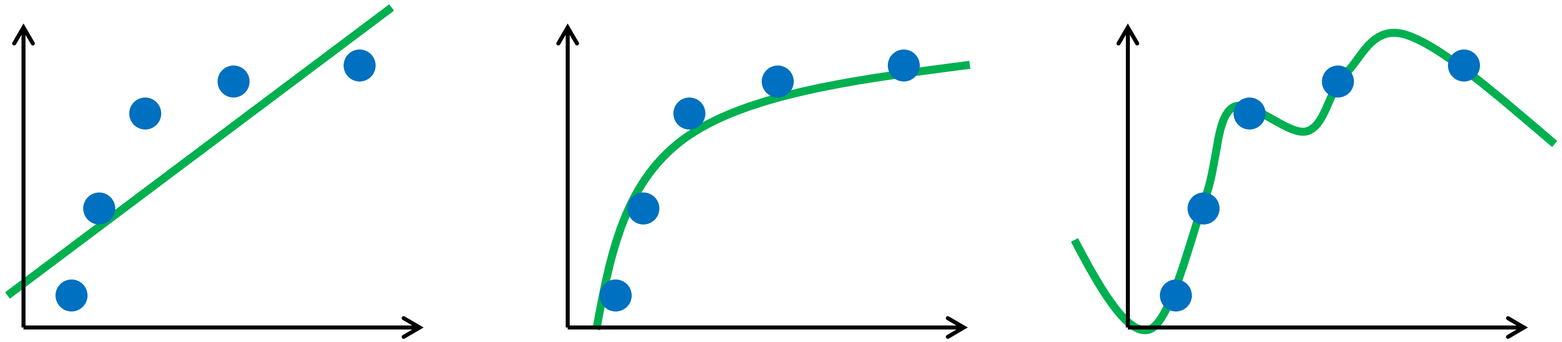
Overfitting（过拟合）

- 误差为0，但是拟合的函数并无使用价值！



欠拟合或过拟合

- 如何选择合适的基函数?



$w_0 + w_1 x$ | $w_0 + w_1 x + w_2 x^2$ | $w_0 + w_1 x + w_2 x^2 + w_3 x^3 +$

High bias (underfitting) | "Just right" | High variance (overfitting)

- 需要根据不同的应用与需求，不断尝试（不断“调参”）

避免过拟合的常用方法

$$y = \sum_{i=0}^{n} w_i B_i(x)$$

- 数据去噪
  - 剔除训练样本中噪声
- 数据增广
  - 增加样本数，或者增加样本的代表性和多样性
- 模型简化
  - 预测模型过于复杂，拟合了训练样本中的噪声
  - 选用更简单的模型，或者对模型进行裁剪
- 正则约束
  - 适当的正则项，比如方差正则项、稀疏正则项

岭回归正则项

- 选择一个函数空间
  - 基函数的线性表达

$$W = (w_0, w_1, \ldots, w_n)$$

$$y = f(x) = \sum_{i=0}^{n} w_i B_i(x)$$

- 最小二乘拟合

$$\min_{W} \|Y - XW\|^2$$

- Ridge regression （岭回归）

$$\min_{W} \|Y - XW\|^2 + \mu \boxed{\|W\|_2^2}$$

稀疏学习：稀疏正则化

- 冗余基函数（过完备）
- 通过优化来选择合适的基函数
  - 系数向量的L_0模（非0元素个数）尽量小
  - 挑选（“学习”）出合适的基函数

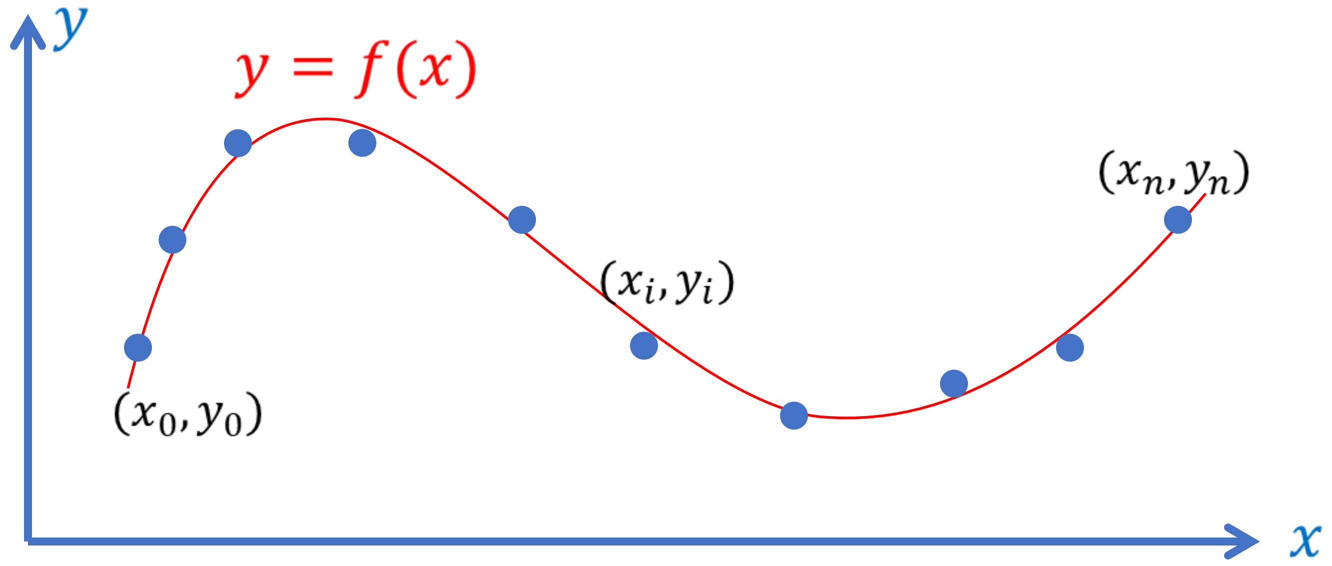
$$\min_{\alpha} \|Y - XW\|^2 + \mu \|W\|_0$$

$$\min_{\alpha} \|Y - XW\|^2, \quad \text{s.t. } \|W\|_0 \leq \beta$$

数据拟合模型求解

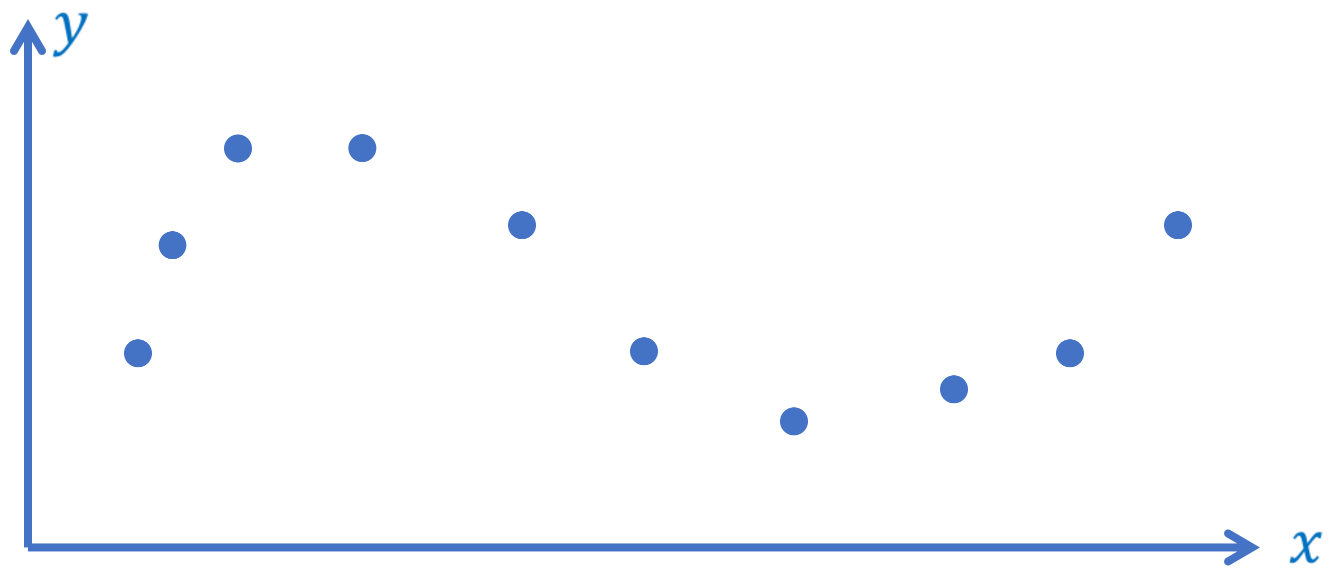
函数拟合问题

- 输入：一些观察（采样）的数据点$\{x_i, y_i\}_{i=0}^{n}$
- 输出：拟合数据点的函数$y = f(x)$，并用于预测



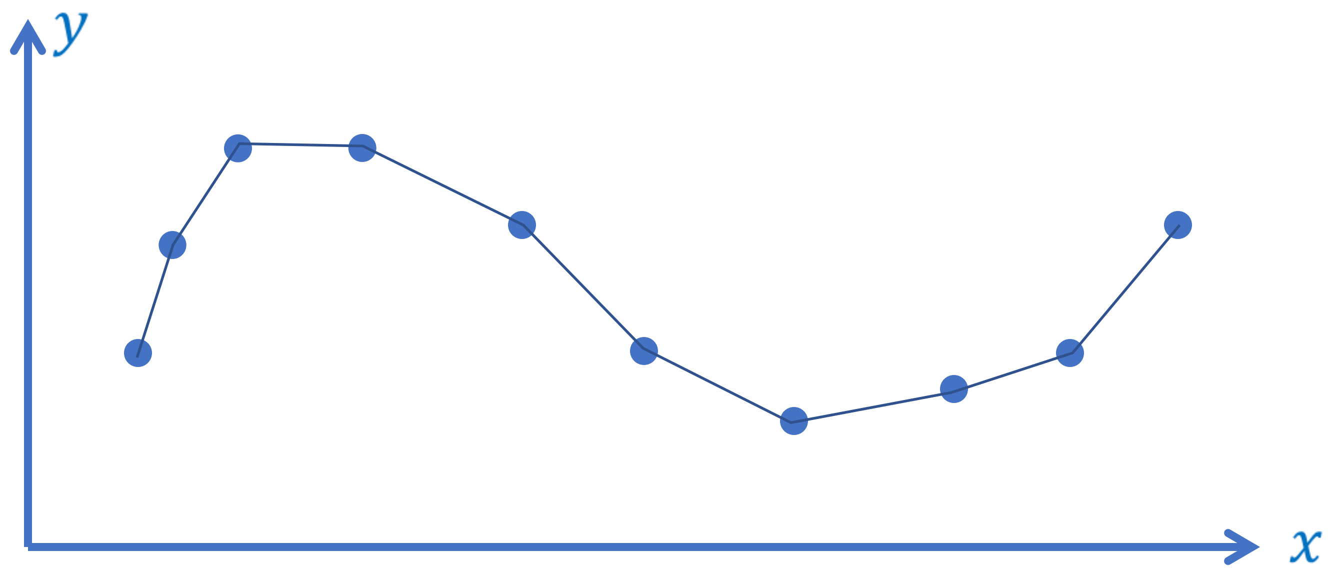
这种拟合函数有多少个？

拟合函数的“好坏”



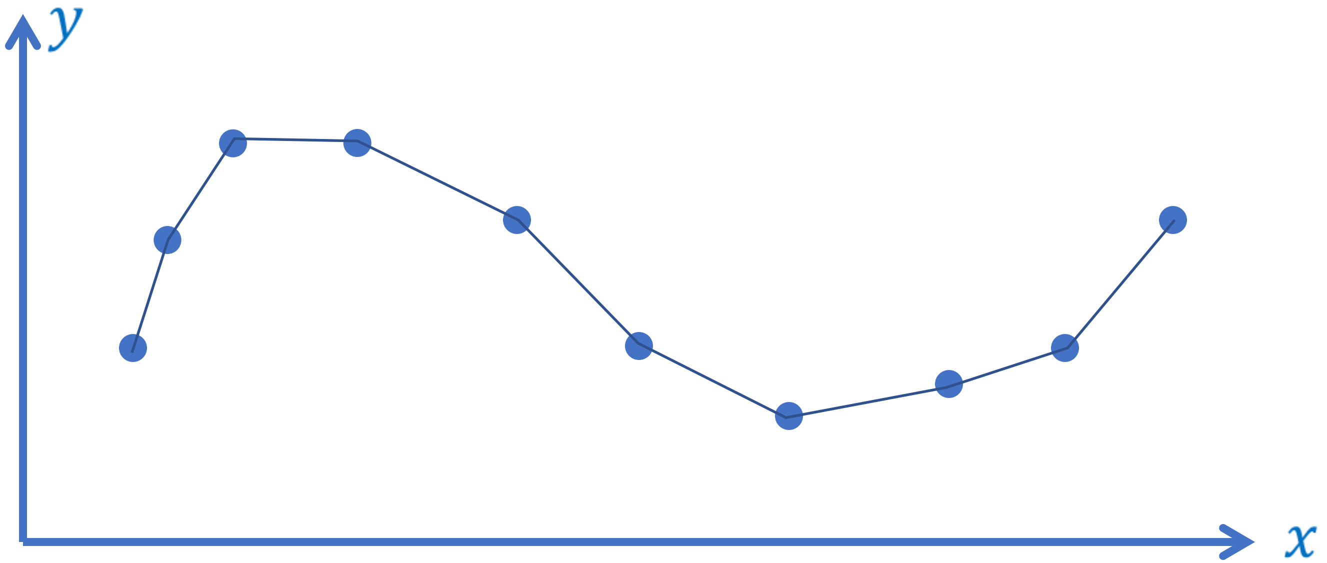
拟合函数的“好坏”

- 分段线性插值函数$y = f_1(x)$
  - 数据误差为0



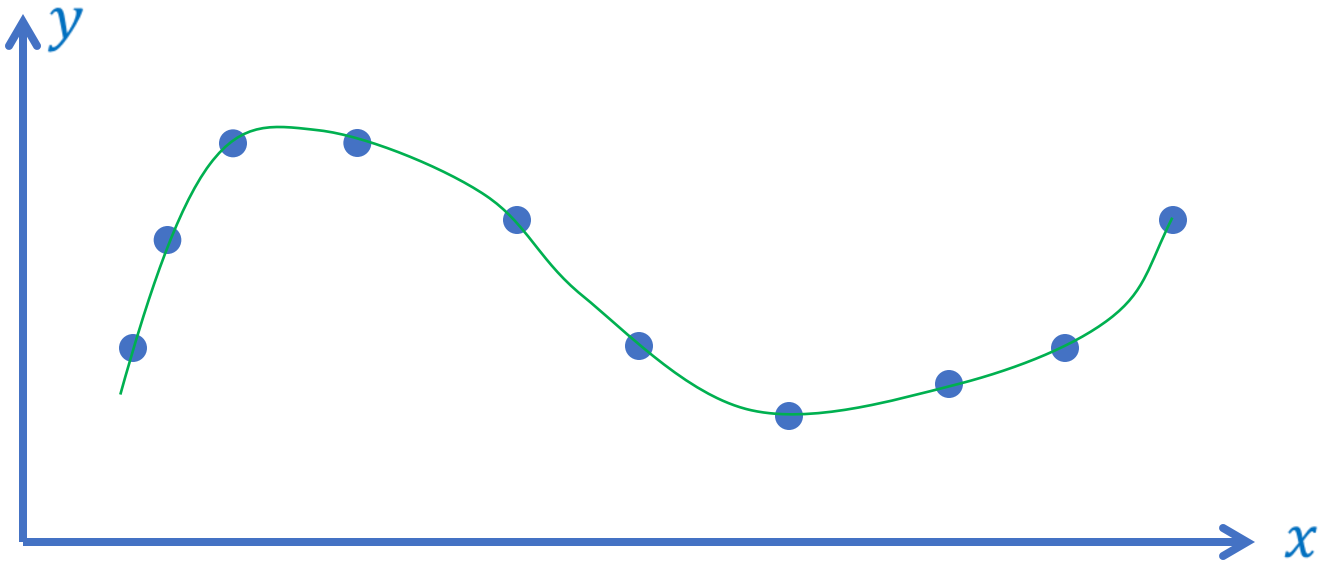
拟合函数的“好坏”

- 分段线性插值函数$y = f_1(x)$
  - 数据误差为0
  - 函数性质不够好：只有C^0连续，不光滑（数值计算）



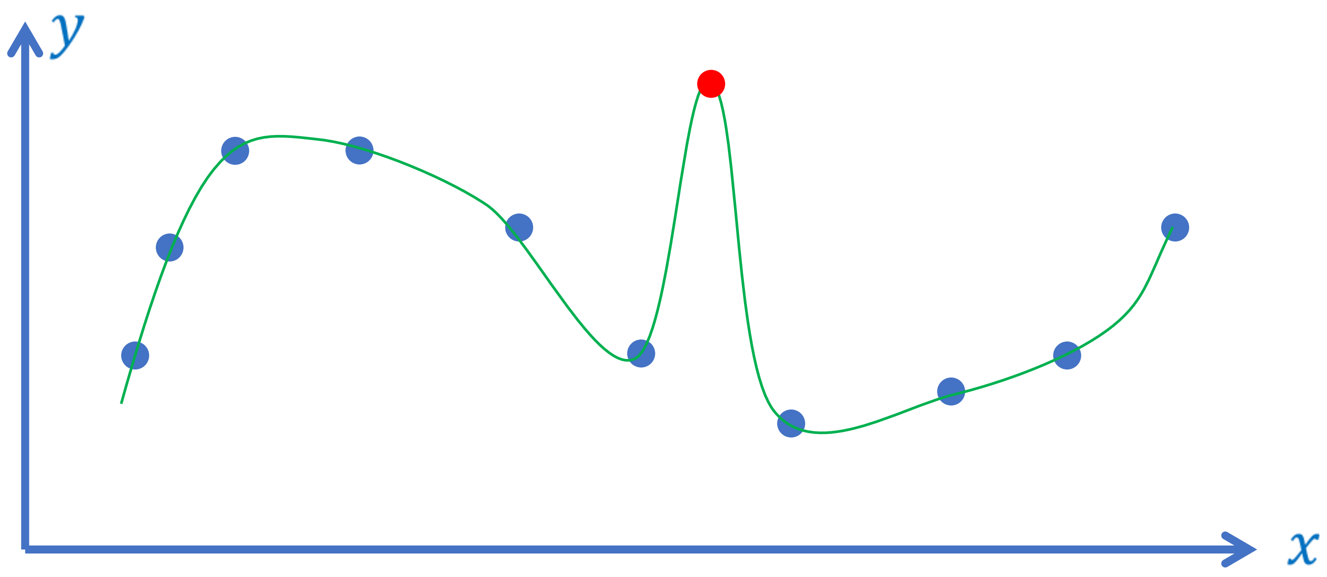
拟合函数的“好坏”

- 光滑插值函数$y = f_2(x)$
  - 数据误差为0



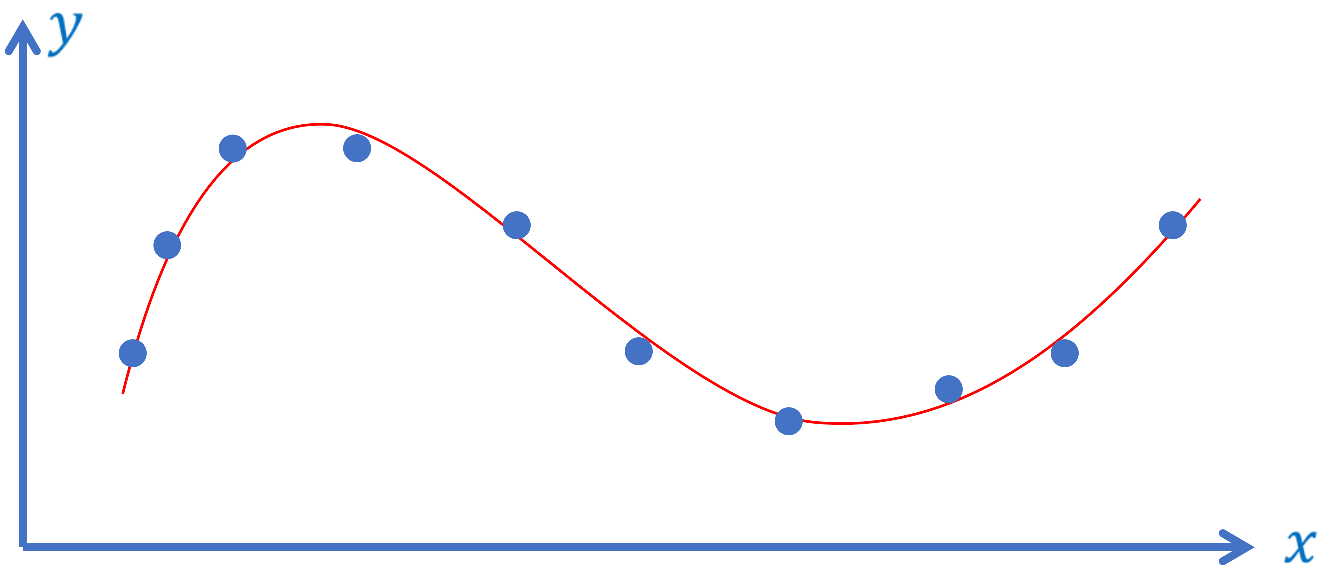
拟合函数的“好坏”

- 光滑插值函数$y = f_2(x)$
  - 数据误差为0
  - 可能被“差数据”（噪声、outliers）带歪，导致函数性质不好、预测不可靠



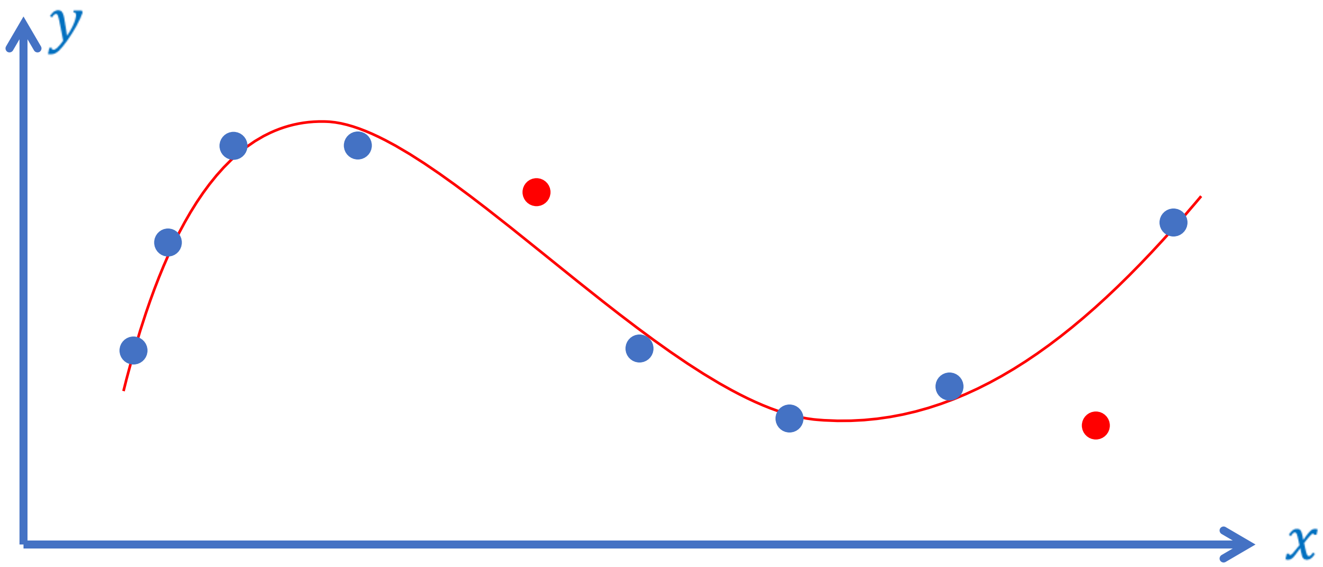
拟合函数的“好坏”

- 逼近拟合函数$y = f_3(x)$
  - 数据误差不为0，但足够小



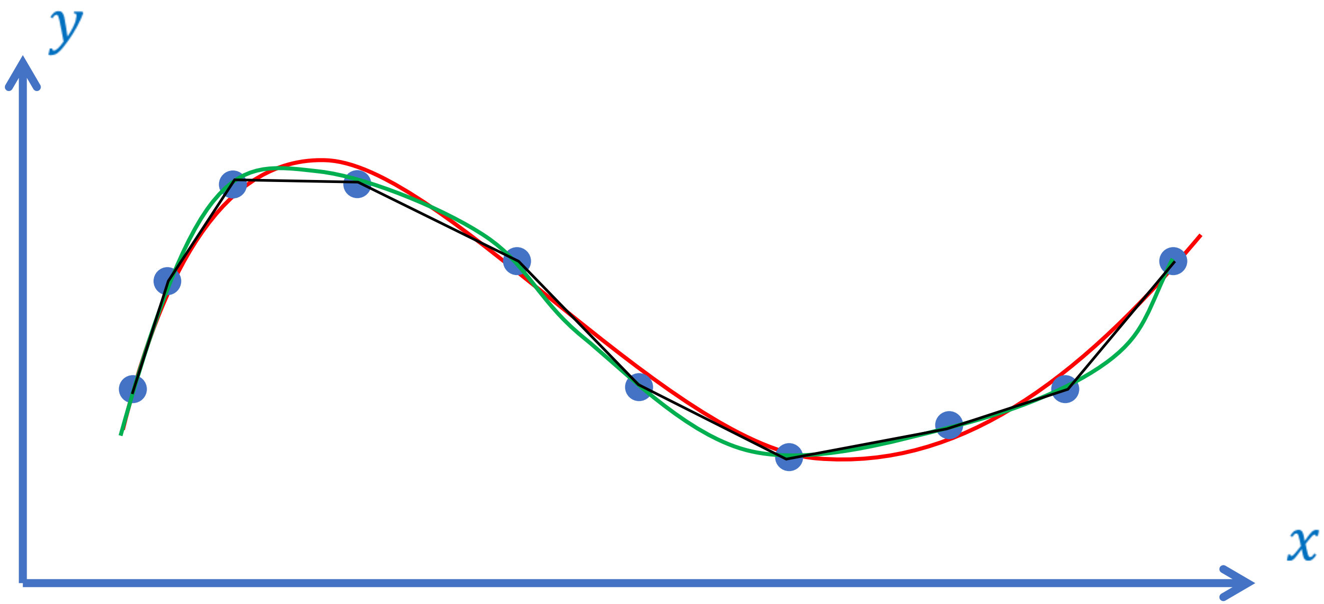
拟合函数的“好坏”

- 逼近拟合函数$y = f_3(x)$
  - 数据误差不为0，但足够小



拟合函数的“好坏”

- 分段线性函数$f = f_1(x)$
- 光滑插值函数$f = f_2(x)$
- 逼近拟合函数$f = f_3(x)$



求拟合函数：应用驱动

- 大部分的实际应用问题
  - 可建模为：找一个映射/变换/函数
  - 输入不一样、变量不一样、维数不一样
- 三步曲方法论：
  - 到哪找？
    - 确定某个函数集合/空间
  - 找哪个？
    - 度量哪个函数是好的/“最好”的
  - 怎么找？
    - 求解或优化：不同的优化方法与技巧，既要快、又要好…

Recap: 数据拟合的方法论

- 到哪找？
  - 确定函数的表达形式（函数集、空间）$L = span\{b_0(x), \dots, b_n(x)\}$
  - 待定基函数的组合系数（求解变量）$f_\lambda(x) = \sum_{k=0}^{n} \lambda_i b_i(x)$
- 找哪个？
  - 优化模型（最小化问题）
    - 能量项 = 误差项 + 正则项
  - 统计模型、规划模型…
- 怎么找？
  - 求解误差函数的驻点（导数为0之处）
  - 转化为系数的方程组
    - 如果是欠定的（有无穷多解），则修正模型
      - 改进/增加各种正则项：Lasso、岭回归、稀疏正则项…
      - 返回修改模型

f由待定系数$\lambda = \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_n \end{pmatrix}$确定

1. 多项式插值

多项式插值定理

定理：若 x_i 两两不同，则对任意给定的 y_i，存在唯一的次数至多是 n 次的多项式 p_n，使得 $p_n(x_i) = y_i \ i = 0, \cdots, n$。

证明：在幂基 $\{1, x, \cdots, x^n\}$ 下待定多项式 p 的形式为：

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

由插值条件 $p(x_i) = y_i, i = 0, \cdots, n$，得到如下方程组：

$$\begin{pmatrix} 1 & x_0 & x_0^2 & \dots & x_o^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

如果基函数选取不一样，方程组的系数矩阵不同

系数矩阵为 **Vandermonde** 矩阵，其行列式非零，因此方程组有唯一解。

技巧1：构造插值问题的通用解

- **构造插值问题的通用解**
  - 给定$n+1$个点$\{(x_0, y_0), \dots, (x_n, y_n)\}$，寻找一组次数为$n$的多项式基函数$l_i$使得

$$l_i(x_j) = \begin{cases} 1, 若 i = j \\ 0, 若 i \neq j \end{cases}$$

  - 插值问题的解为：

$$P(x) = y_0 l_0(x) + y_1 l_1(x) + \cdots + y_n l_n(x) = \sum_{i=0}^{n} y_i l_i(x)$$

一般形式

- 怎么计算多项式$l_i(x)$?
  - n阶多项式，且有以下n个根

$$x_0, x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n$$

  - 故可表示为

$$l_i(x) = C_i(x - x_0)(x - x_1)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n) = C_i \prod_{j \neq i} (x - x_j)$$

  - 由$l_i(x_i) = 1$可得

$$1 = C_i \prod_{j \neq i} (x_i - x_j) \Rightarrow C_i = \frac{1}{\prod_{j \neq i} (x_i - x_j)}$$

技巧1：构造插值问题的通用解

- 最终多项式基函数为

$$l_i(x) = \frac{\prod_{j \neq i}(x - x_j)}{\prod_{j \neq i}(x_i - x_j)}$$

- 多项式$l_i(x)$被称为拉格朗日多项式

技巧2：更方便的求解表达

- Newton插值：具有相同“导数”（差商）的多项式构造（*n*阶Taylor展开）

定义：

一阶差商：

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

k 阶差商：

设 $\{x_0, x_1, \cdots, x_k\}$ 互不相同，$f(x)$ 关于 $\{x_0, x_1, \cdots, x_k\}$ 的 k 阶差商为：

$$f[x_0, x_1, \cdots, x_k] = \frac{f[x_1, \cdots, x_k] - f[x_0, x_1, \cdots, x_{k-1}]}{x_k - x_0}$$

所以 Newton **插值多项式**表示为：

$$N_n(x) = f(x_0) + f[x_0, x_1](x - x_0) + \cdots + f[x_0, x_1, \cdots, x_n](x - x_0)\cdots(x - x_{n-1})$$

多项式插值存在的问题

- 系统矩阵稠密
- 依赖于基函数选取，矩阵可能病态，导致难于求解（求逆）

病态矩阵示例

- 考虑二元方程组
  - 解为$(1,1)$

$$x_1 + 0.5x_2 = 1.5$$
$$0.667x_1 + 0.333x_2 = 1$$

- 对第二个方程右边项扰动0.001
  - 解为$(0,3)$

$$x_1 + 0.5x_2 = 1.5$$
$$0.667x_1 + 0.333x_2 = 0.999$$

- 对矩阵系数进行扰动
  - 解为$(2,-1)$

$$x_1 + 0.5x_2 = 1.5$$
$$0.667x_1 + 0.334x_2 = 1$$

病态问题

- 输入数据的细微变化导致输出(解)的剧烈变化
- 将线性方程看成直线（超平面）
  - 当系统病态时，直线变为近似平行
  - 求解（即直线求交）变得困难、不精确

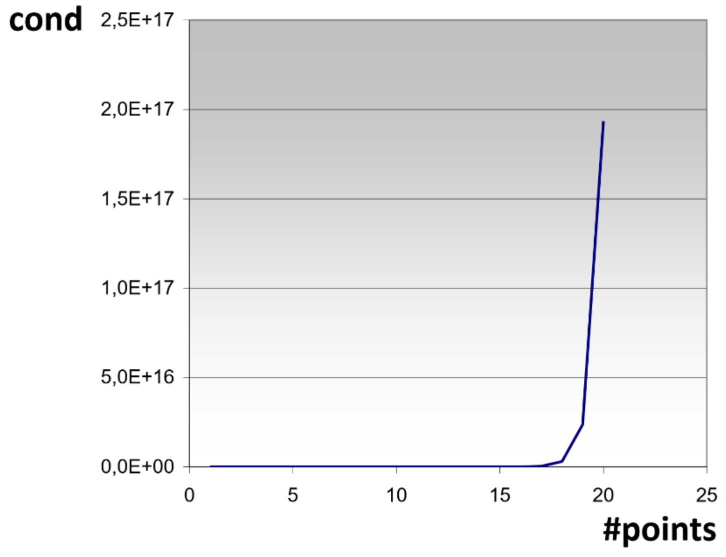
矩阵条件数

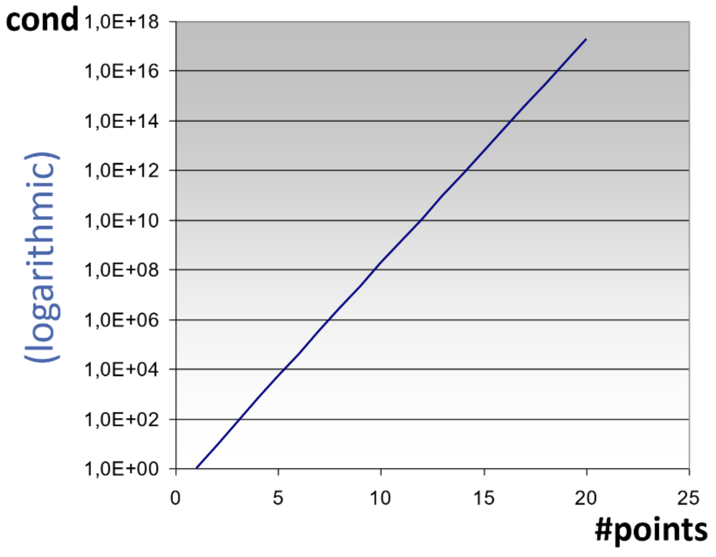
$$\kappa_2(A) = \frac{\max\limits_{x \neq 0} \frac{\|Ax\|}{\|x\|}}{\min\limits_{x \neq 0} \frac{\|Ax\|}{\|x\|}}$$

- 等于最大特征值和最小特征值之间比例
- 条件数大意味着基元之间有太多相关性

矩阵条件数

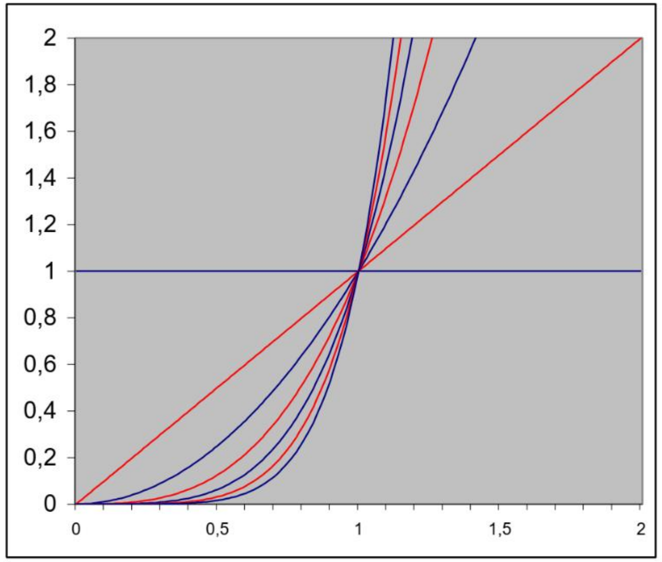
- 多项式插值问题是病态的
  - 对于等距分布的数据点x_i，范德蒙矩阵的条件数随着数据点数n呈指数级增长（多项式的最高次数为$n-1$）





为什么？

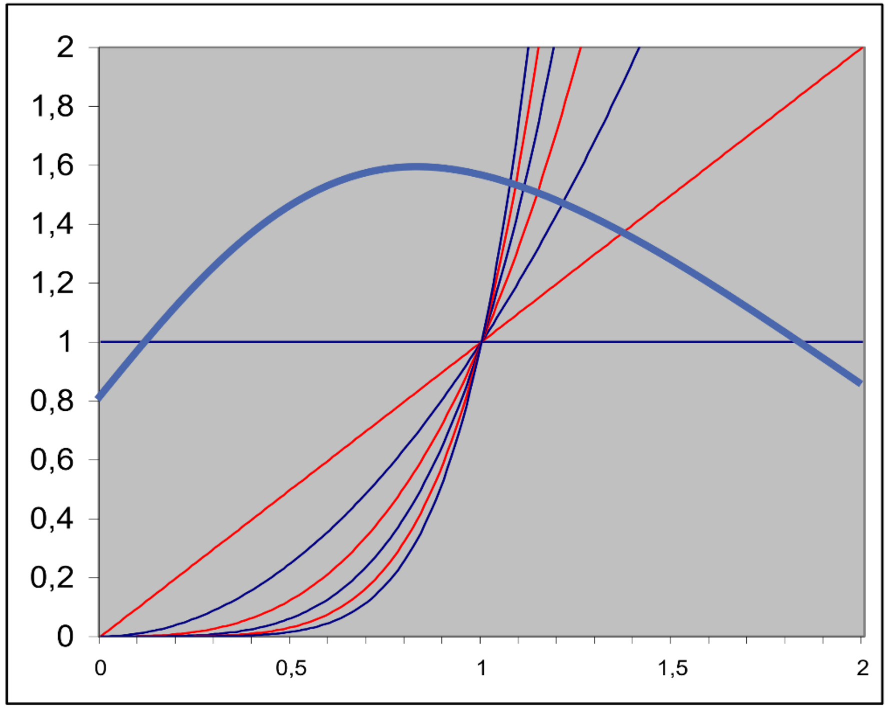
- 幂（单项式）函数基
  - 幂函数之间差别随着次数增加而减小
  - 不同幂函数之间唯一差别为增长速度(x^i比x^{i-1}增长快)



幂(单项式) 函数

函数互相抵消

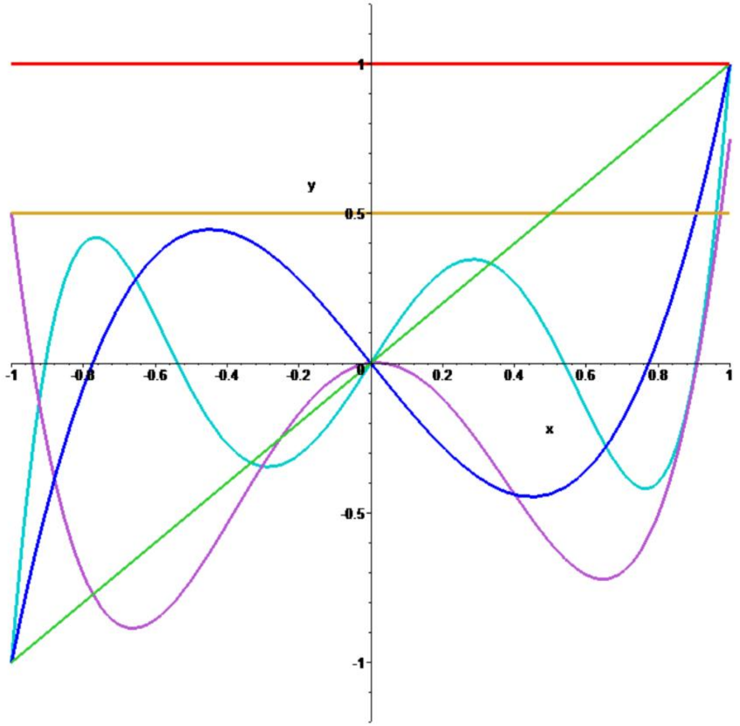
- 单项式：
  - 从左往右
  - 首先常函数1主宰
  - 接着x增长最快
  - 接着x^2增长最快
  - 接着x^3增长最快
  - …
- 趋势
  - 好的基函数一般需要系数交替
  - 互相抵消问题



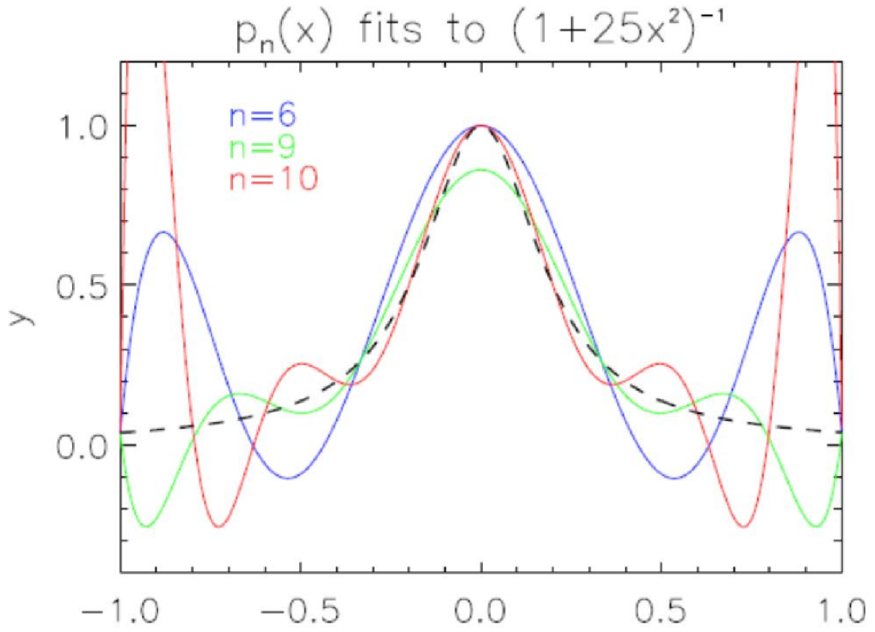
解决方法

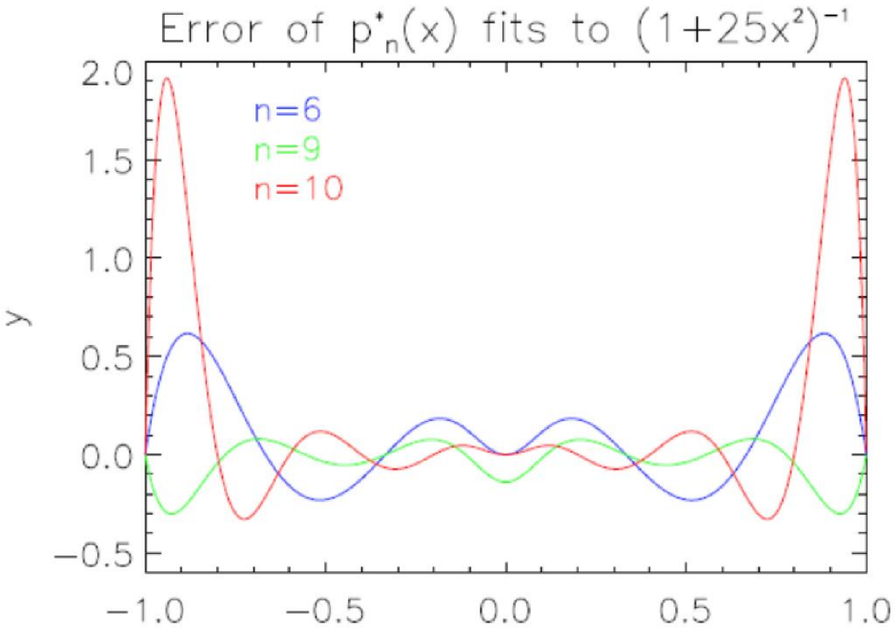
- 使用正交多项式基

- 如何获得？
  - Gram-Schmidt正交化



多项式插值结果好吗？





振荡(龙格Runge)现象和对插值点数的高度敏感性

观察$n = 9$(10个数据点)和$n = 10$(11个数据点)的差别

结论

- 多项式插值不稳定
  - 控制点的微小变化可导致完全不同的结果
- 振荡(Runge)现象
  - 多项式随着插值点数(可以是细微)增加而摆动
- 需要更好的基函数来做插值
  - Bernstein基函数?
  - 分片多项式?

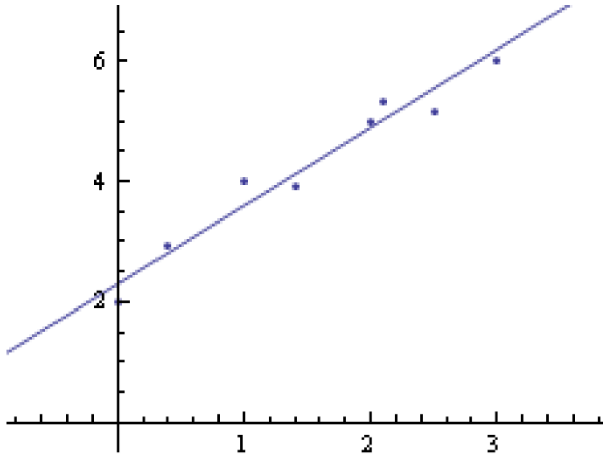
2. 多项式逼近

为什么逼近?

- 数据点含噪声、outliers等
- 更紧凑的表达
- 计算简单、更稳定

最小二乘逼近

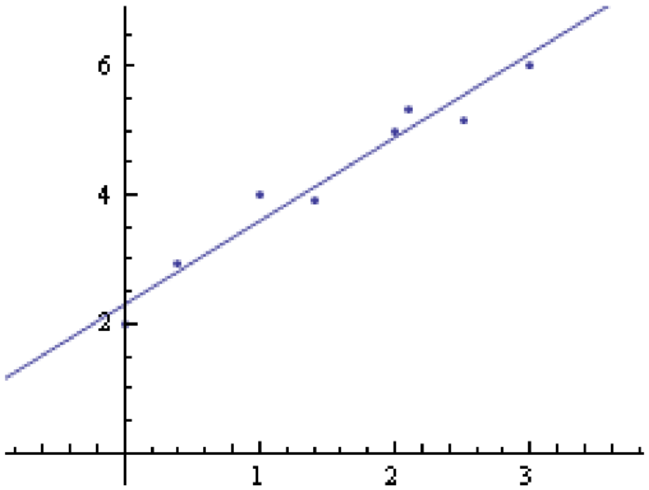
- 逼近问题
  - 给定一组线性无关的连续函数集合$B = \{b_1, \ldots b_n\}$和一组结点$\{(x_1, y_1), \ldots, (x_m, y_m)\}$，其中$m > n$
  - 在B张成空间中哪个函数$f \in span(B)$对结点逼近最好？
  - 示例：给定一组点，找到最佳逼近的线性函数
  - 怎么定义“最佳逼近”?



最佳逼近的定义

- **最小二乘逼近**

$$\underset{f\in span(B)}{\operatorname{argmin}} \sum_{j=1}^{m} \left(f(x_j) - y_j\right)^2$$



$$\sum_{j=1}^{m} \left(f(x_j) - y_j\right)^2 = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} \lambda_i b_i(x_j) - y_j\right)^2$$

$$= (M\lambda - y)^T (M\lambda - y)$$

$$= \lambda^T M^T M\lambda - y^T M\lambda - \lambda^T M^T y + y^T y$$

$$= \lambda^T M^T M\lambda - 2y^T M\lambda + y^T y$$

$$M = \begin{pmatrix} b_1(x_1) & \dots & b_n(x_1) \\ \dots & \dots & \dots \\ b_1(x_m) & \dots & b_n(x_m) \end{pmatrix}$$

求解

- 关于λ的二次多项式

$$\lambda^T M^T M \lambda - 2y^T M \lambda + y^T y$$

- 法方程
  - 最小解满足

$$M^T M \lambda = M^T \mathrm{y}$$

- 提示
  - 最小化二次目标函数 $x^T A x + b^T x + c$
  - 充分必要条件: $2Ax = -b$

3. 函数空间及基函数

为什么用多项式？

- 易于计算，表现良好，光滑，…
- 稠密性与完备性：表达能力足够！
  - **魏尔斯特拉斯Weierstrass定理**：令f为闭区间$[a,b]$上任意连续函数，则对任意给定ε，存在n和多项式P_n使得

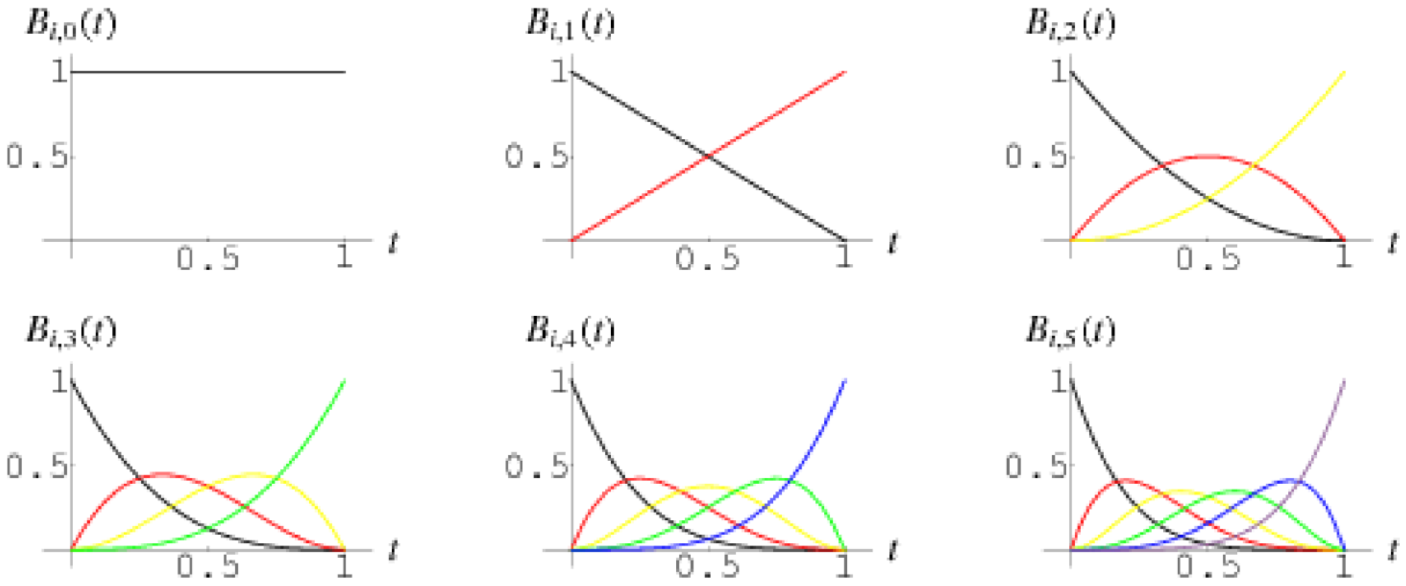
$$|f(x)-P_n(x)|<\varepsilon, \forall x\in[a,b]$$

  - Weierstrass只证明了存在性，未给出多项式

用Bernstein多项式做逼近

- 伯恩斯坦Bernstein给出了构造性证明（强大!）
  - 对$[\mathbf{0},\mathbf{1}]$区间上任意连续函数$f(x)$和任意正整数n，以下不等式对所有$x \in [0,1]$成立
  $$|f(x) - B_n(f,x)| < \frac{9}{4} m_{f,n}$$
  - $m_{f,n} = \underset{y_1,y_2 \in [0,1] \text{且} |y_1 - y_2| < \frac{1}{\sqrt{n}}}{\text{lower upper bound}} |f(y_1) - f(y_2)|$
  - $B_n(f,x) = \sum_{j=0}^{n} f(x_j) b_{n,j}(x)$，其中$x_j$为$[0,1]$上等距采样点
  - $b_{n,j} = \binom{n}{j} x^j (1-x)^{n-j}$为**Bernstein多项式**

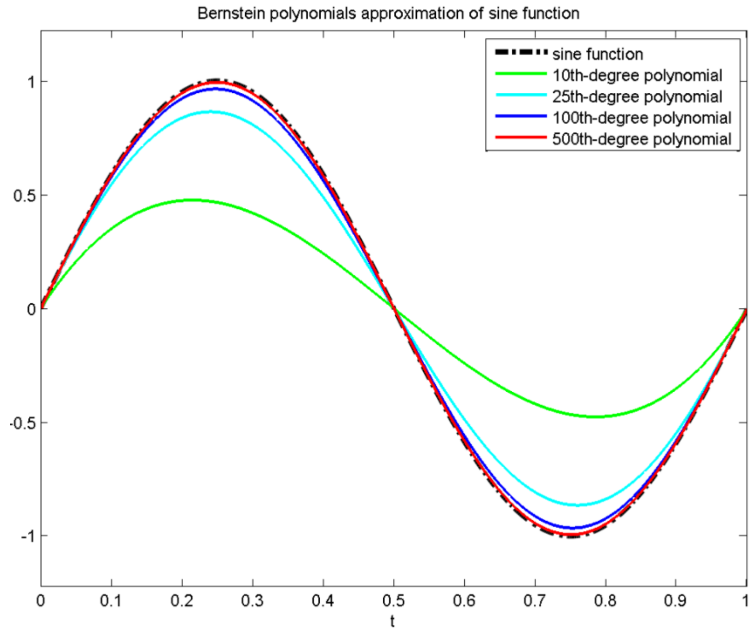
Bernstein多项式



- $b_{0,0}(x) = 1$
- $b_{0,1}(x) = 1 - x, \quad b_{1,1} = x$
- $b_{0,2}(x) = (1 - x)^2, \; b_{1,2} = 2x(1 - x), \quad b_{2,2} = x^2$
- $b_{0,3}(x) = (1 - x)^3, \; b_{1,3} = 3x(1 - x)^2, \; b_{2,3} = 3x^2(1 - x), \quad b_{3,3} = x^3$
- $b_{0,4}(x) = (1 - x)^4, \; b_{1,4} = 4x(1 - x)^3, \; b_{2,4} = 6x^2(1 - x)^2, \; b_{3,4} = 4x^3(1 - x), \quad b_{4,4} = x^4$

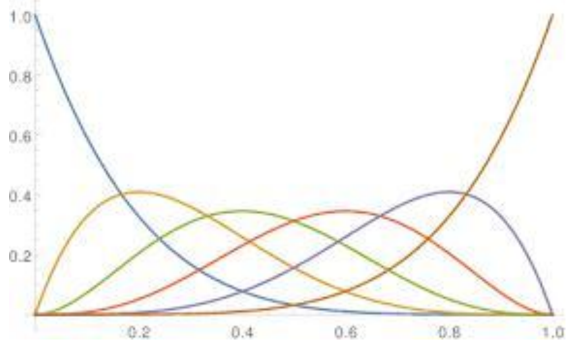
用Bernstein多项式做逼近

- Bernstein基函数的良好性质：非常好的几何意义！
  - 正性、权性（和为1）→凸包性
  - 变差缩减性
  - 递归线性求解方法
  - 细分性
  - …
- Bernstein多项式逼近示例
  - 逼近结果优秀
  - 需要高阶

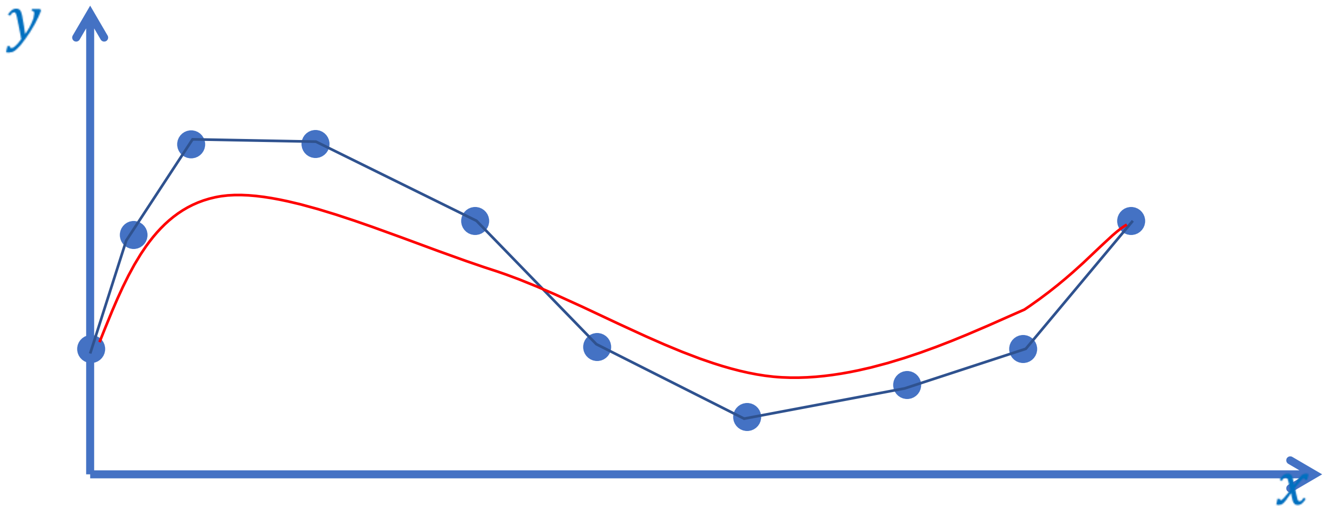


丰富的理论：CAGD课程

关于Bernstein函数...



- $B_n(f,x) = \sum_{j=0}^{n} f\left(\frac{i}{n}\right) b_{n,j}(x)$
- 两种观点：
  – 几何观点、代数观点



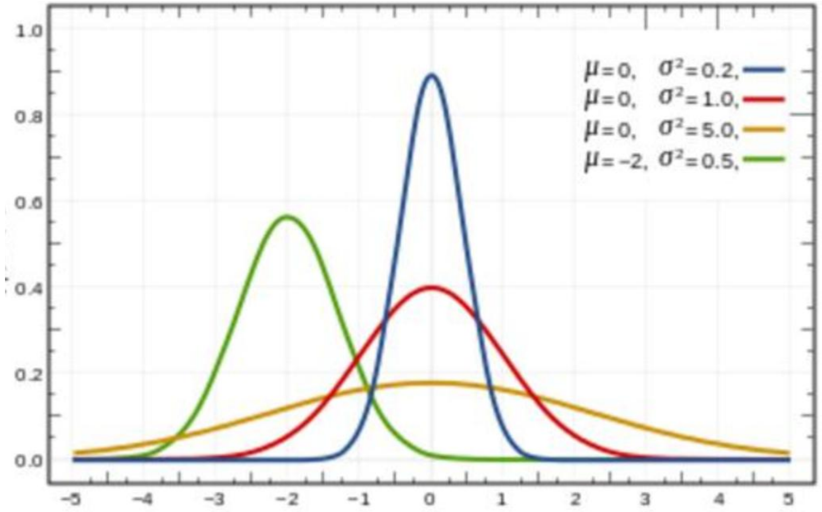
丰富的理论：CAGD课程

4. RBF函数插值/逼近

Gauss函数

- 两个参数：均值μ，方差σ

$$g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- 几何意义：
  - 均值μ：位置
  - 方差σ：支集宽度

- 不同均值和方差的Gauss函数都线性无关
  - 有什么启发？

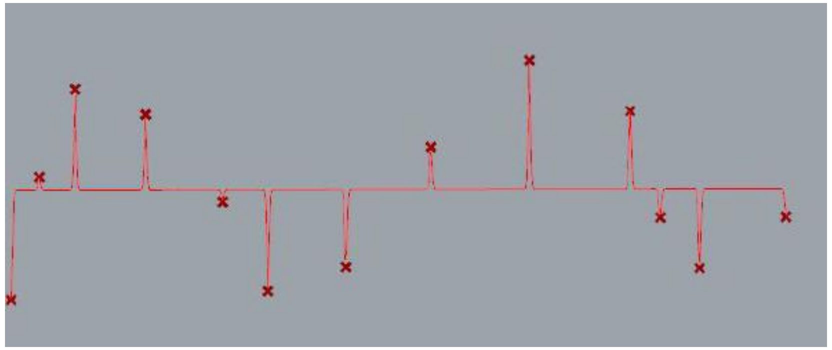
RBF函数拟合

- RBF函数

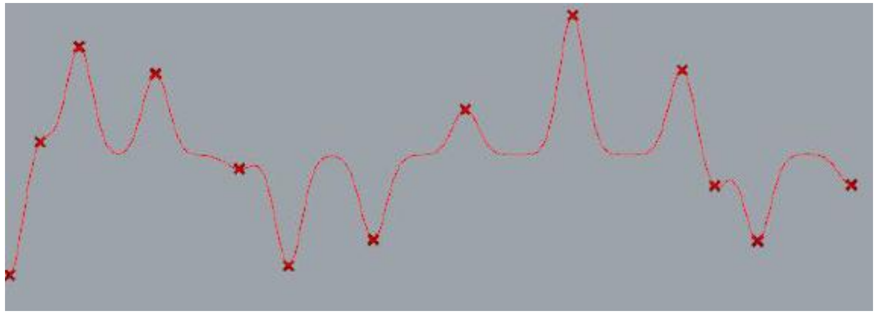
$$f(x) = b_0 + \sum_{i=1}^{n} b_i g_i(x)$$

- 方法
- 原因

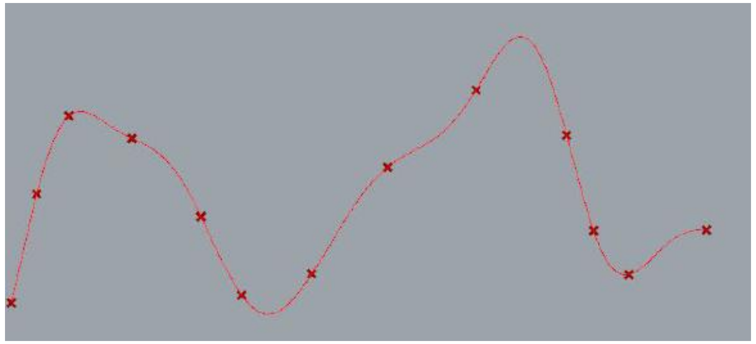
讨论：现象



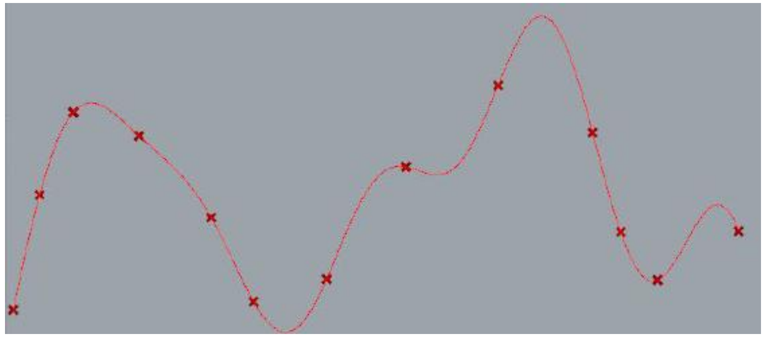
$\sigma = 0.1$



$\sigma = 1$

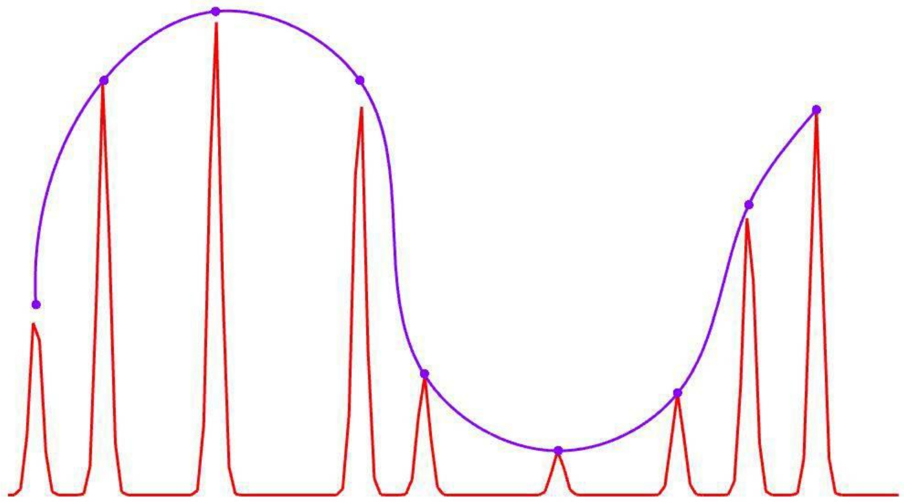


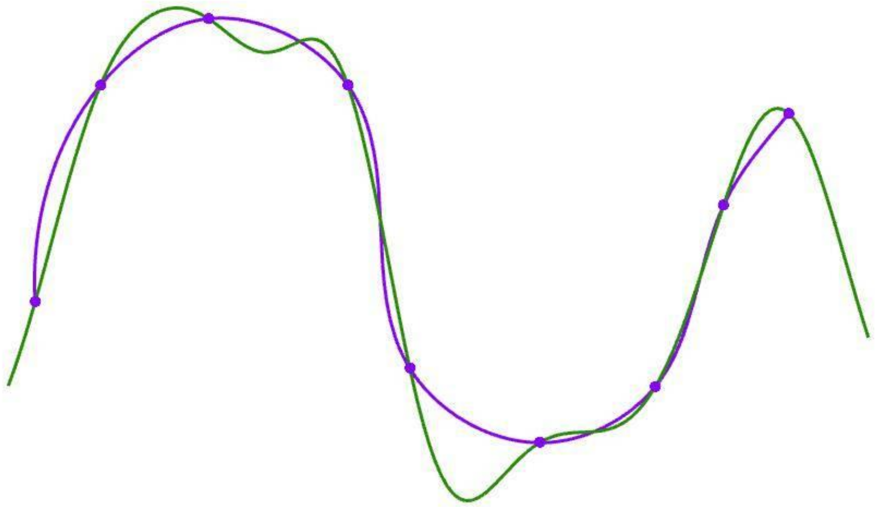
$\sigma = 5$



$\sigma = 10$

讨论：现象



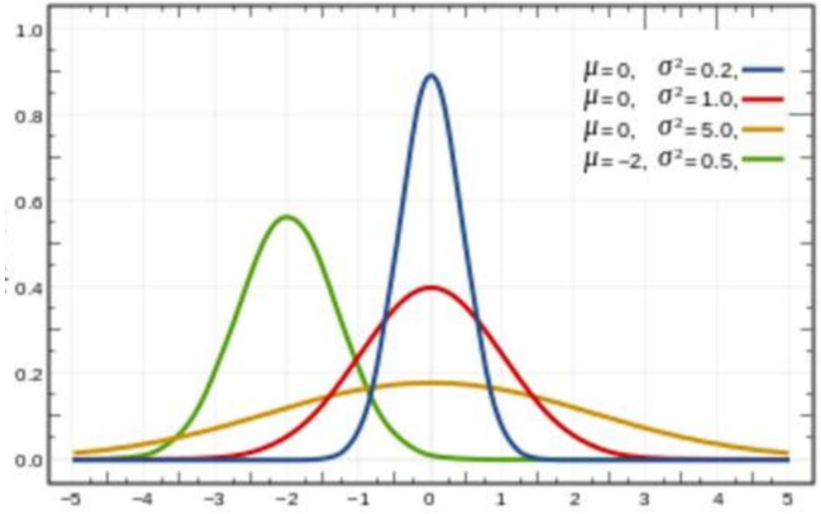


思考:

- 均值μ和方差σ是否可以一起来优化?

$$g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f(x) = b_0 + \sum_{i=1}^{n} b_i g_i(x)$$



- 问题?

5. 从另一个角度来看拟合函数

Gauss拟合函数

- 一般Gauss函数表达为标准Gauss函数的形式

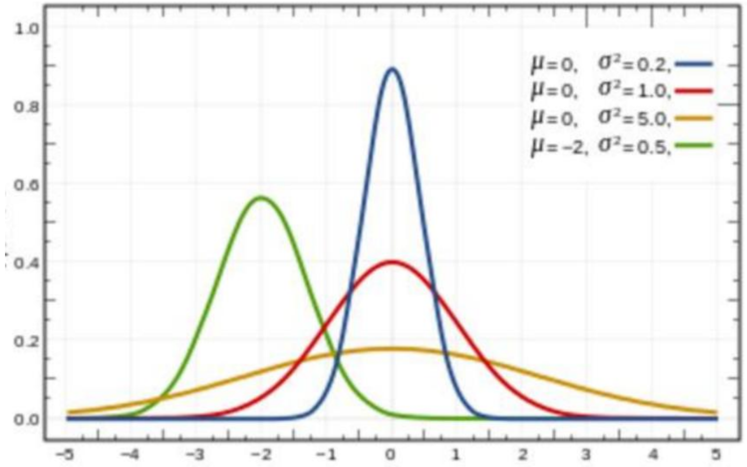
$$g_{\mu,\sigma}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}-\frac{\mu}{\sigma}\right)^2}=g_{0,1}(ax+b)$$

$$a=\frac{x}{\sigma},b=-\frac{\mu}{\sigma}$$

$$f(x)=b_0+\sum\nolimits_{i=1}^{n}b_i g_i(x)$$

$$f(x)=\omega_0+\sum\nolimits_{i=1}^{n}\underline{\omega_i g_{0,1}(a_i x+b_i)}$$

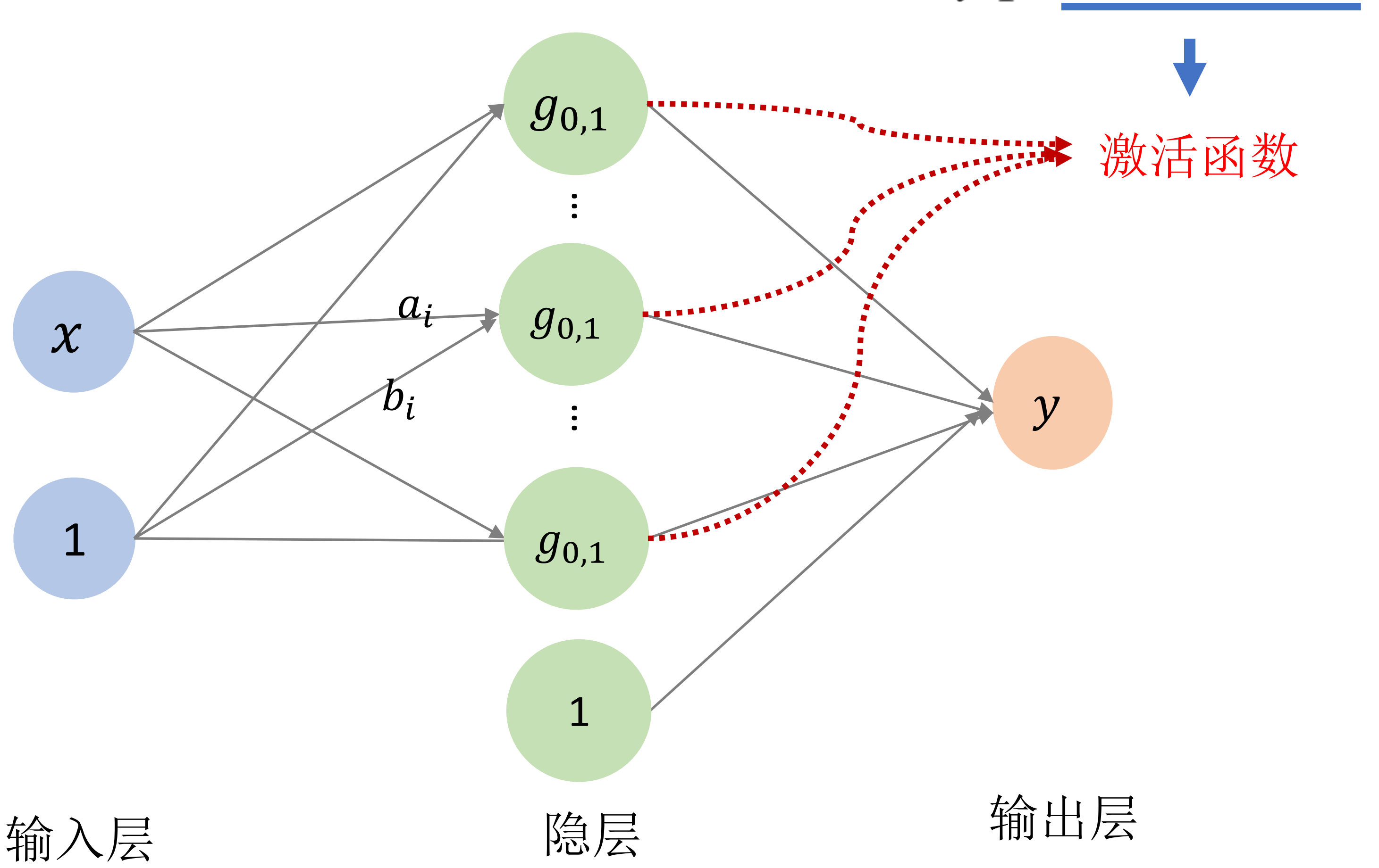
基函数是由一个基本函数通
过平移和伸缩变换而来的



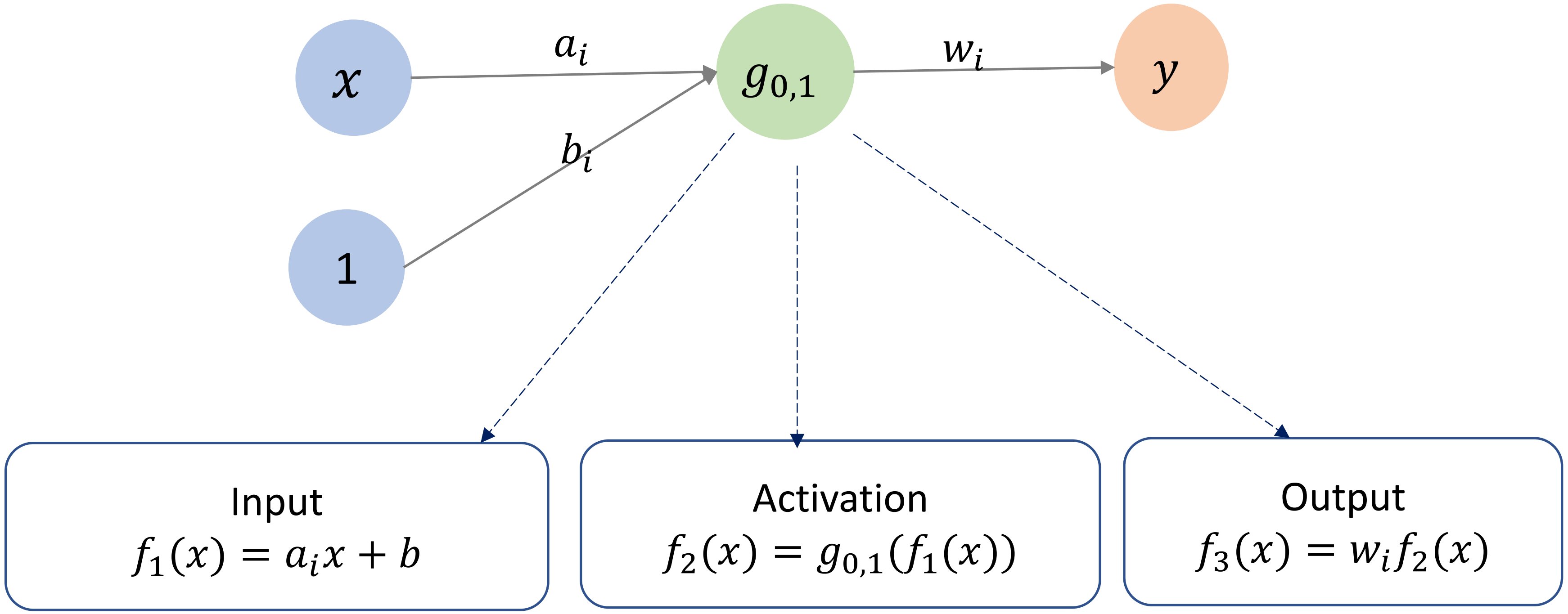
换个方式看函数：神经网络

- 将Gauss函数看成网络

$$f(x) = \omega_0 + \sum_{i=1}^{n} \omega_i g_{0,1}(a_i x + b_i)$$



抽象： 神经元


x
a_i
$g_{0,1}$
w_i
y
b_i
1
Input
$f_1(x) = a_i x + b$
Activation
$f_2(x) = g_{0,1}(f_1(x))$
Output
$f_3(x) = w_i f_2(x)$

RBF 神经网络

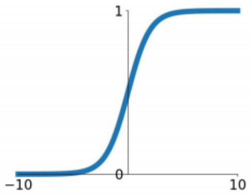
- 高维情形：RBF (Radial Basis Function)，径向基函数
- 一种特殊的BP网络
  - 优化：BP算法
- 核函数思想
- Gauss函数的特性：拟局部性

思考：激活函数的选择？

- 启发：由一个简单的函数通过（仿射）变换构造出一组基函数，张成一个函数空间
- 表达能力是否足够强：是否完备/稠密的？

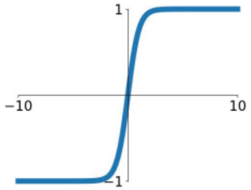
Sigmoid

$\sigma(x) = \frac{1}{1+e^{-x}}$



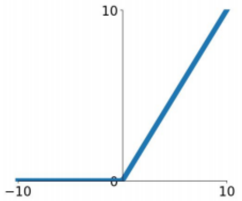
tanh

$\tanh(x)$



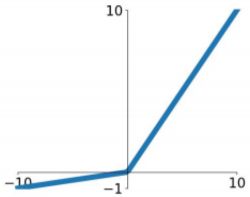
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

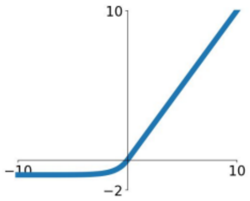


Maxout

$\max(w_1^T x + b_1, w_2^T x + b_2)$

ELU

$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$



高维情形：多元函数
（后面的课程再展开解释）

- 变量的多个分量的线性组合

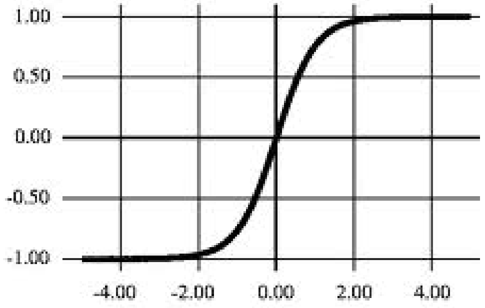
$$(x_1, x_2, \dots, x_n) \longrightarrow g_{0,1}(a_1^i x_1 + a_2^i x_2 + \cdots + a_n^i x_n + b_i)$$

- 单隐层神经网络函数：

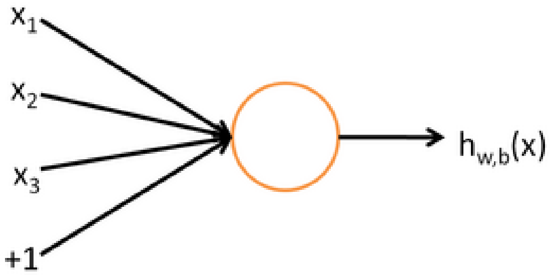
$$f(x_1, x_2, \dots, x_n) = \omega_0 + \sum_{i=1}^{n} \omega_i g_{0,1}\left(a_1^i x_1 + a_2^i x_2 + \cdots + a_n^i x_n + b_i\right)$$

多层神经网络：多重复合的函数

- 线性函数和非线性函数的多重复合

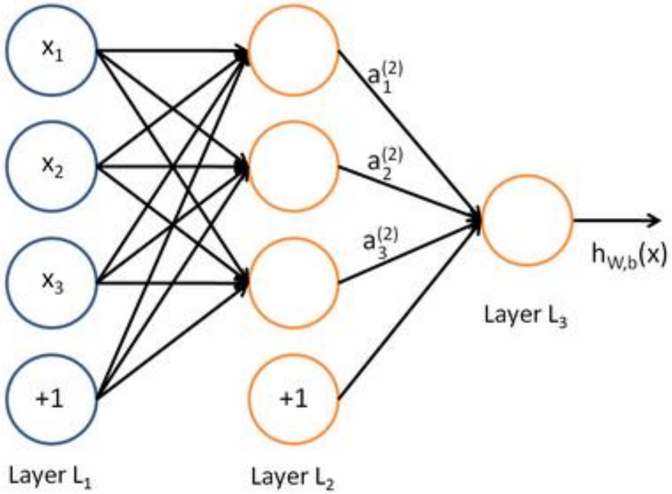


f(*net*)



Neuron

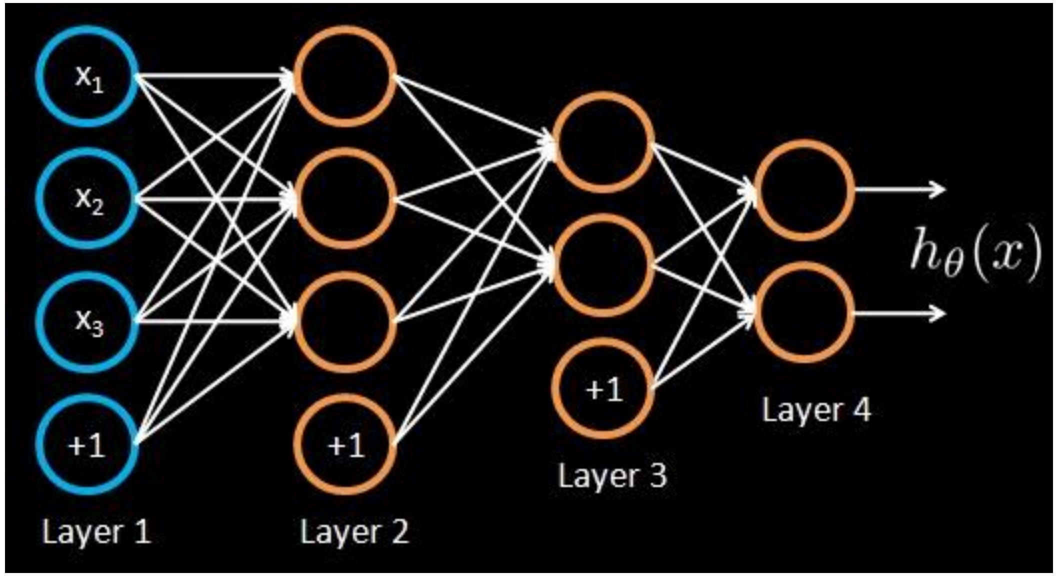
$$h_{W,b}(x) = f(W^T x) = f(\sum_{i=1}^{3} W_i x_i + b)$$



Neural network

$$a_1^{(2)} = f(W_{11}^{(1)} x_1 + W_{12}^{(1)} x_2 + W_{13}^{(1)} x_3 + b_1^{(1)})$$
$$a_2^{(2)} = f(W_{21}^{(1)} x_1 + W_{22}^{(1)} x_2 + W_{23}^{(1)} x_3 + b_2^{(1)})$$
$$a_3^{(2)} = f(W_{31}^{(1)} x_1 + W_{32}^{(1)} x_2 + W_{33}^{(1)} x_3 + b_3^{(1)})$$
$$h_{W,b}(x) = a_1^{(3)} = f(W_{11}^{(2)} a_1^{(2)} + W_{12}^{(2)} a_2^{(2)} + W_{13}^{(2)} a_3^{(2)} + b_1^{(2)})$$

用神经网络函数来拟合数据



Regression problem:

Input: Given training set $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots.$

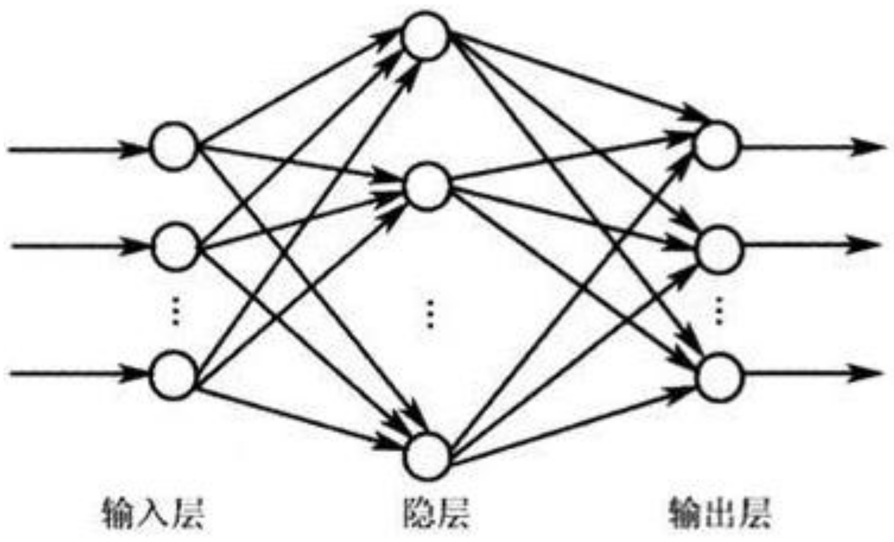
Output: Adjust parameters θ (for every node) to make:

$$h(x_i) \approx y_i$$

Why it works?

- 万能逼近定理：自由度足够多！

$$F(\mathrm{x}) = \sum_{i=1}^{N} v_i \varphi\left(\mathrm{w}_i^{\mathrm{T}} \mathrm{x} + b_i\right)$$



与传统拟合一样存在同样的问题：
函数个数如何选？！

调参！

Deep Learning Frameworks

TensorFlow

DL4J Deep Learning for Java

neon

PaddlePaddle

Keras

MATLAB®

theano

torch

Chainer

dmlc mxnet

Caffe

Microsoft CNTK

cuDNN

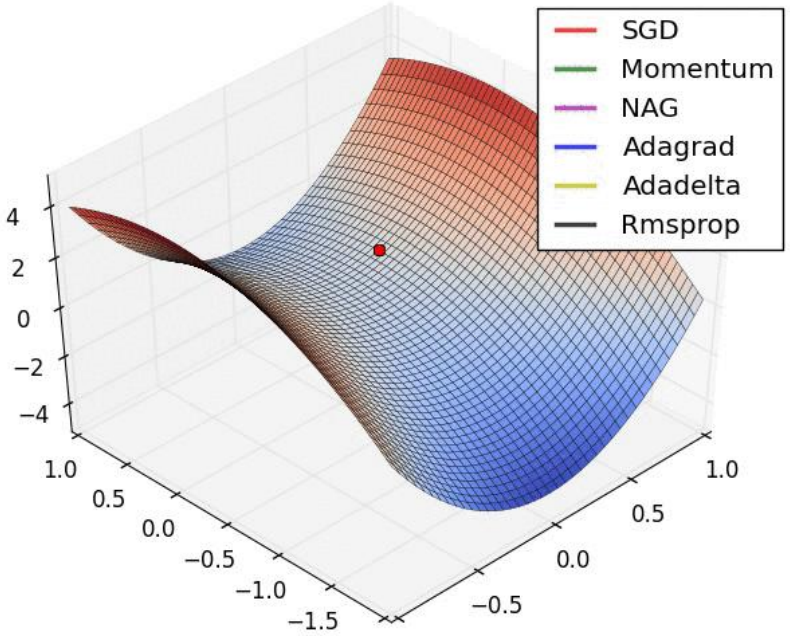
NVIDIA DIGITS

Lasagne

使用深度学习的方法

- 问题建模
  - 理解问题、问题分解（多个映射级联）…
- 找哪个?
  - 损失函数、各种Penalty、正则项…
- 到哪找?
  - 神经网络函数、网络简化…
- 怎么找?
  - 优化方法（BP方法）
  - 初始值、参数…

调参：有耐心、有直觉…



作业3： 数据拟合模型

- 任务： 以下2选1
  - 平面点列的曲线拟合
  - 使用神经网络进行昆虫分类
- 目标
  - 学习和实现数据拟合的数学模型并进行可视化
- 提交时间：
  - 2025年4月12日星期六

E. T. Y. Lee, “Choosing nodes in parametric curve interpolation,” Comput. Aided Design, vol. 21, p. 363–370, July 1989.
Zhang C M, Cheng F H and Miura K, A method for determining knots in parametric curve interpolation, CAGD, 1998, 15: 399-416.
M. S. Floater and T. Surazhsky, “Parameterization for curve interpolation,” in Topics in Multivariate Approximation and Interpolation (K. Jetter, M. D. Buhmann, W. Haussmann, R. Schaback, and J. Stöckler, eds.), vol. 12 of Studies in Computational Mathematics, pp. 39–54, Elsevier, 2006.

谢 谢 ！