中国传媒大学

深度学习与类脑计算

（二）

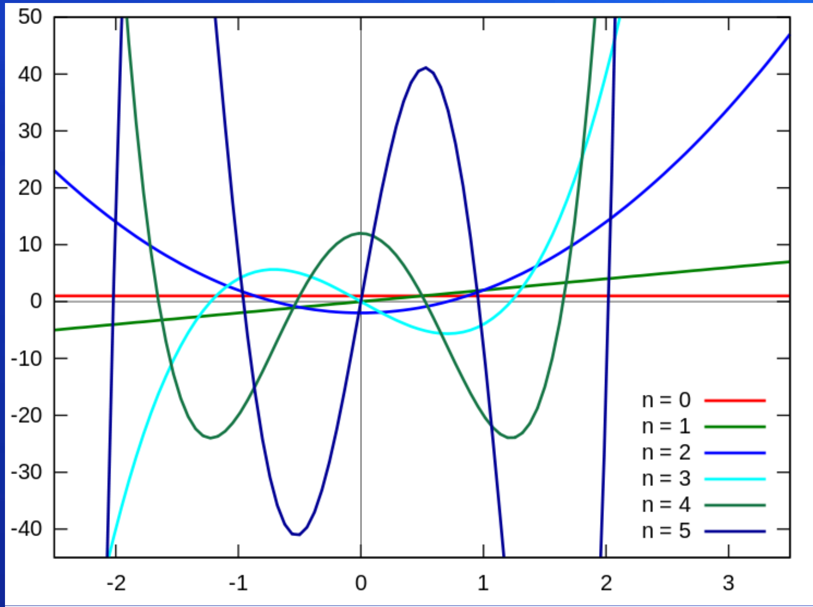
曹立宏

脑科学与智能媒体研究院

一些基础的基础的基础…

- 导数-单变量
  - 极限
  - 连锁法则
  - 高阶
  - Taylor展开
  - 多项式逼近

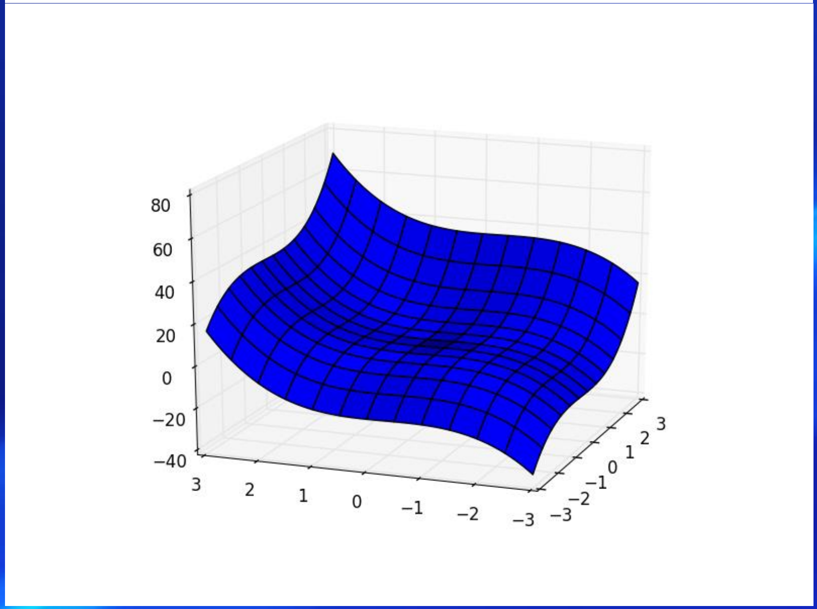
$$f(x)=\frac{f(a)}{0!}+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots+\frac{f^{(n)}(a)}{n!}(x-a)^n+R_n(x)$$



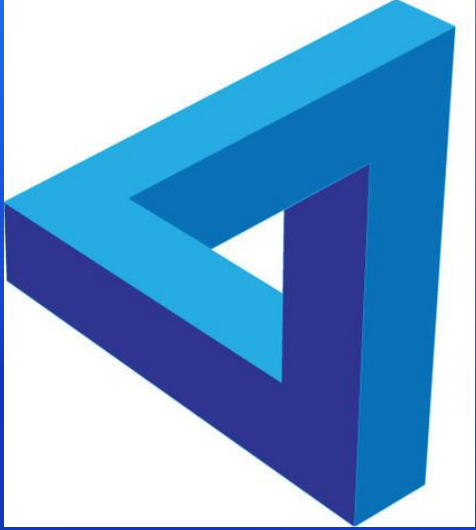
- 多变量 f: $R^n \rightarrow R^1$
  - 偏导
  - Jacobian矩阵 1xn
  - Hessian矩阵 nxn

$$\nabla f=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix}$$

$$\mathrm{H}(f)=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1\partial x_1} & \frac{\partial^2 f}{\partial x_1\partial x_n}\\ \frac{\partial^2 f}{\partial x_n\partial x_1} & \frac{\partial^2 f}{\partial x_n\partial x_n}\end{bmatrix}$$



距离



- 任何两个Objects的相似性（例如马和牛）
- 数学上的定义：
  - D(x,y)=0 ⟺ x=y
  - D(x,y)=D(y,x)
  - D(x,y)+D(y,z)>=D(x,z)

对于R^n，闵可夫斯基距离（Minkowski distance）

$$\vec{x}=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \vec{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad D_p(\vec{x},\vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}$$

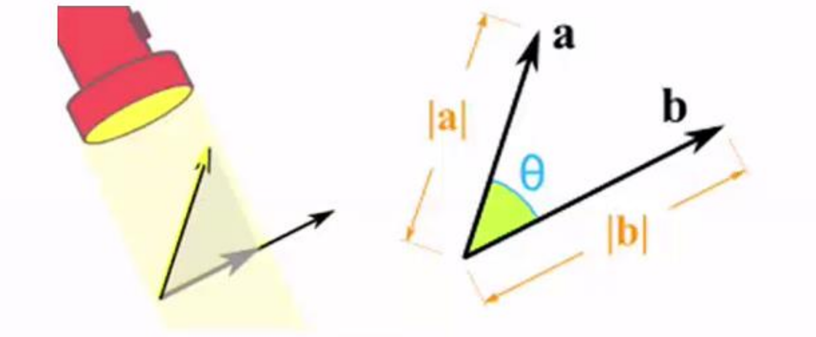
当 p=2, 欧氏距离

当 p 趋近于无穷大时，闵可夫斯基距离转化成切比雪夫距离（Chebyshev distance）

$$\lim_{p\to\infty}\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p} = \max_i |x_i - y_i|^1$$

函数与函数之间的距离？

线性代数

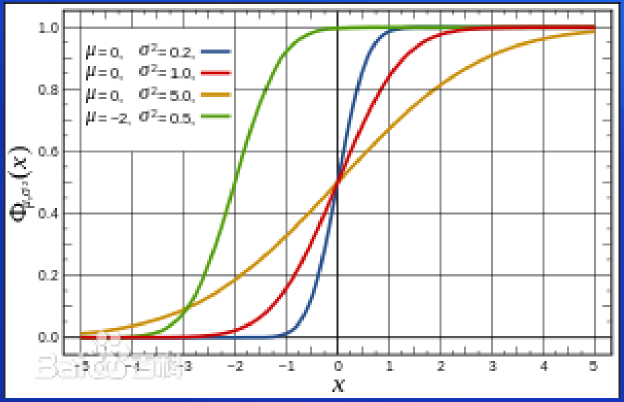


$$|a| \times |b| \times \cos(\theta) = |a| \times \cos(\theta) \times |b|$$

- Vector
  - *Norm - Distance*
  - *Dot product - Projection*
- Matrix
  - *Transformer on vector* T: $R^n \rightarrow R^{\mathrm{m}}$ $\vec{y}=\mathrm{T}_{\mathrm{mxn}}\vec{x}$ Linear
  - *Product – Inverse*
  - *Decomposition*
    - *Eigendecomposition – Real Symmetric Matrix*
    - *Singular Value Decomposition – Real Matrix*
    - *Moore-Penrose pseudoinverse*
- Tensor

概率统计

- 确定性 vs 随机性
- 离散与连续
- 随机性的描述
  - 分布函数-累积分布
  - 随机变量的转换 – 伪随机数-样本生成
  - 均值、方差 –
  - 高阶矩
  - 特征函数



切比雪夫定理：只要均值、方差存在，

则对任意常数 ε>0，有P(| X - E(X) | ≥ ε) ≤ D(X) / ε²

- 多个随机变量，他们之间的关系
  - 联合分布、边际分布
  - 条件概率 – Bayes
  - 独立
  - 特征函数

大数定律 (The Law of Large Numbers)

- 辛钦大数定律：常用的大数定律

设 $\{a_i, i \geq 1\}$为独立同分布的随机变量序列，若a_i的数学期望存在为μ，则对任意的$\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{N}\sum_{i=1}^{N} a_i - \mu\right| < \varepsilon\right) = 1$$

独立同分布 – i.i.d.

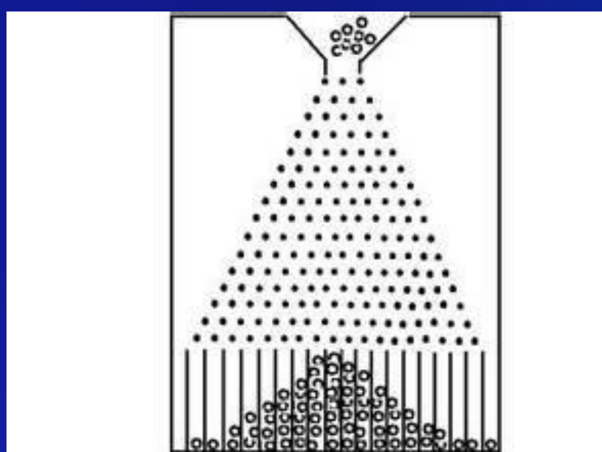
作业：用模拟的方法验证大数定律

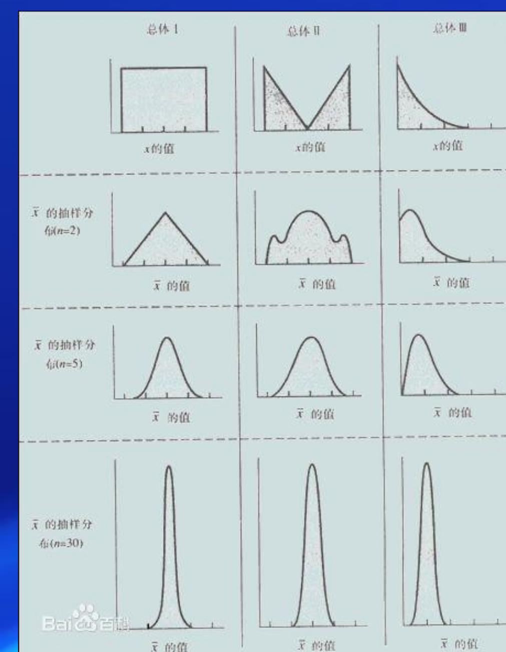
中心极限定理–Central Limit Theorem

设 $\{a_i, i \geq 1\}$ i.i.d.，若a_i的数学期望和方差存在，分别为μ和σ^2，令$Y_n = (a_1 + a_2 + \cdots + a_n)/n$，

则,

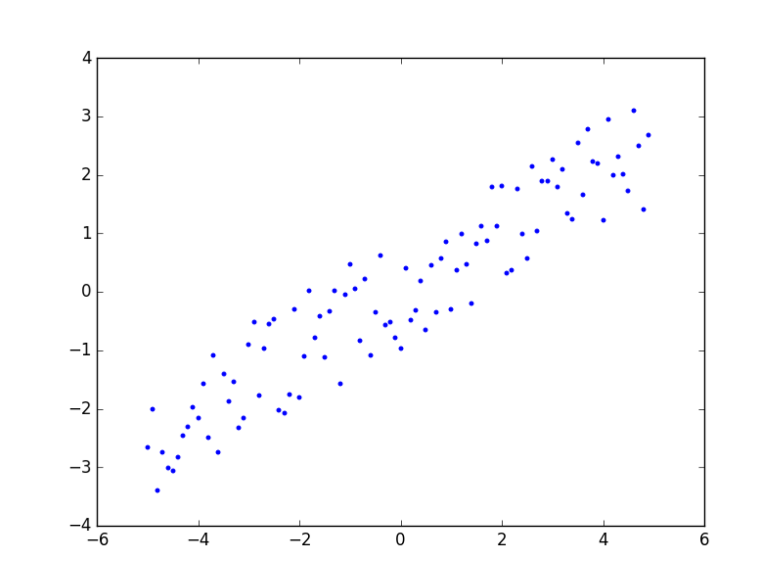
$$\lim_{n\to\infty} Y_n \sim N(\mu, \sigma^2/n)$$

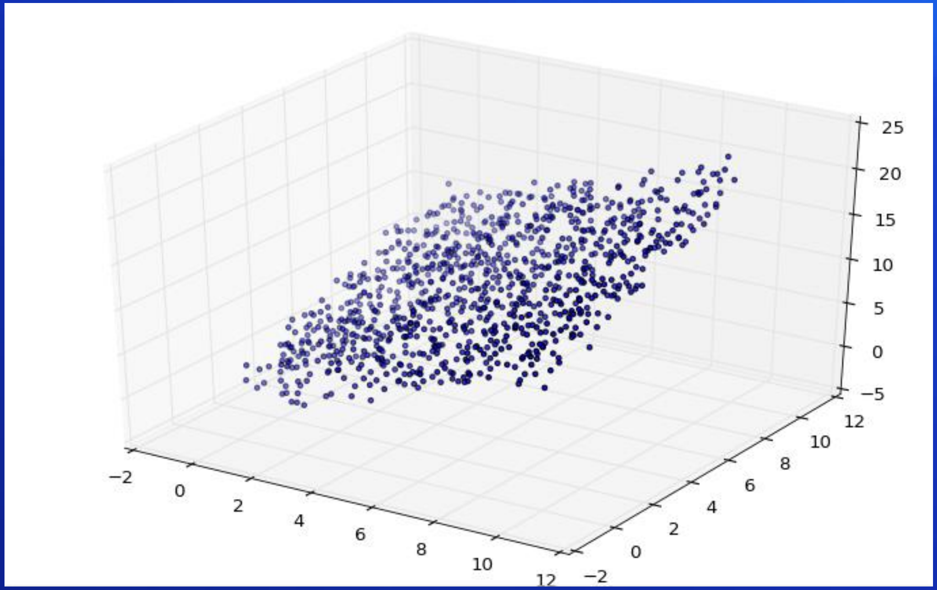




作业：用模拟的方法验证中心极限定理

Regression





$\{x_i, y_i |\ _{i=1,\ \ldots,\ \mathrm{N}}\}$

$\{x_i, y_i, z_i |\ _{i=1,\ \ldots,\ \mathrm{N}}\}$

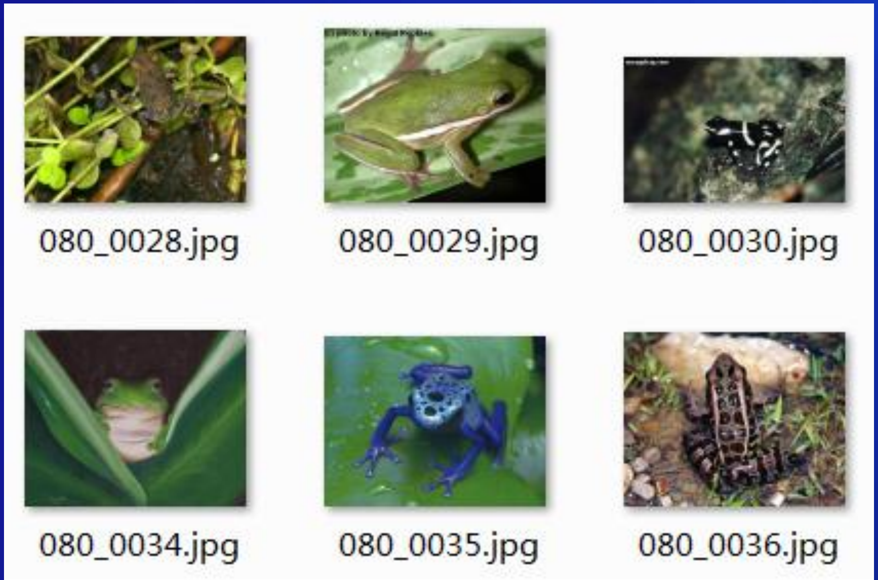
f: $R^n \rightarrow R^1$

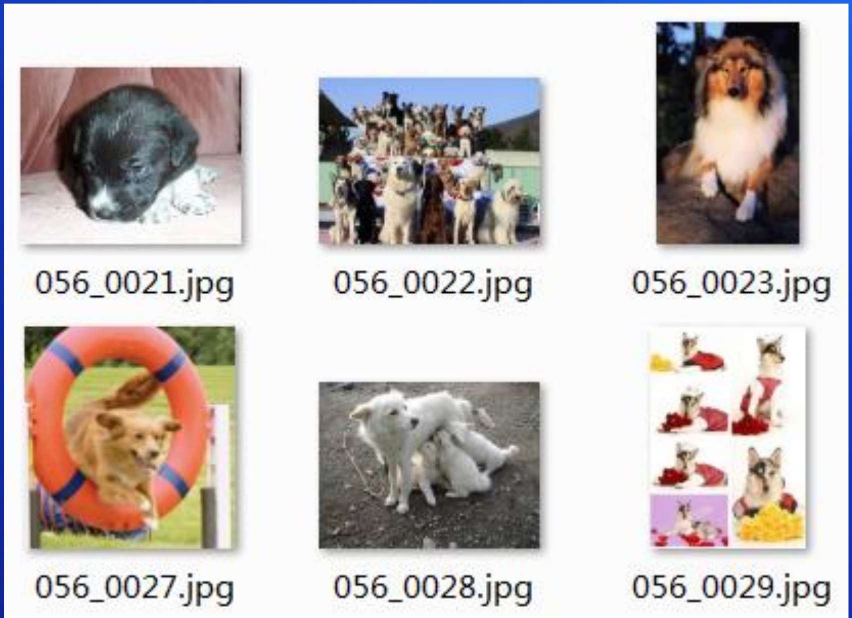
$\{\vec{x}_i, y_i, |\ _{i=1,\ \ldots,\ \mathrm{N}}\}$

f: $R^n \rightarrow R^m$

$\{\vec{x}_i, \vec{y}_i, |\ _{i=1,\ \ldots,\ \mathrm{N}}\}$

Classification





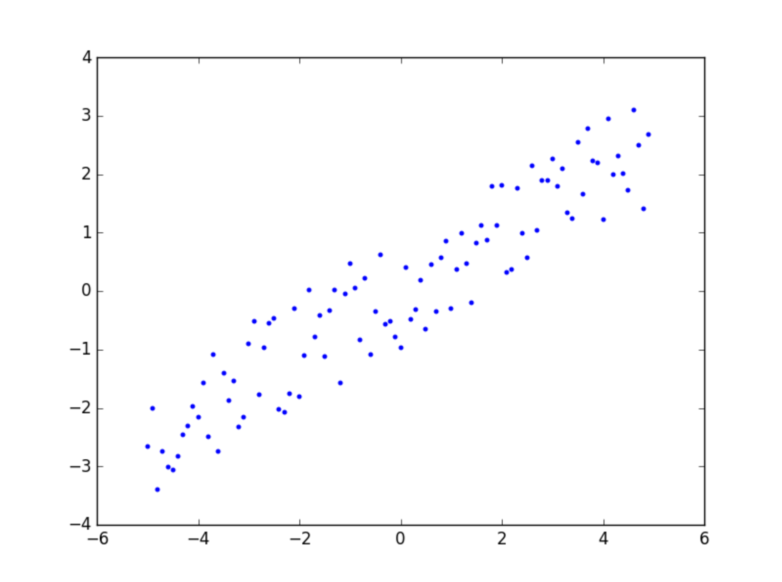
图像的表征：WxH RGB Pixels

$$C: \quad R^n \rightarrow \{1, \ldots, K\}$$

n 可能非常大！1百万像素的图片，n=3百万

找规律+找方法

- Why？
  - Estimation
  - Forecasting
- How？
  - 目测
    - 2-3-4D
    - Limited
    - Dimension Reduction
  - 建模
    - 参数法 – Pre knowledge - Regression
    - 非参数法 – Intuition and Try and Try and Try…

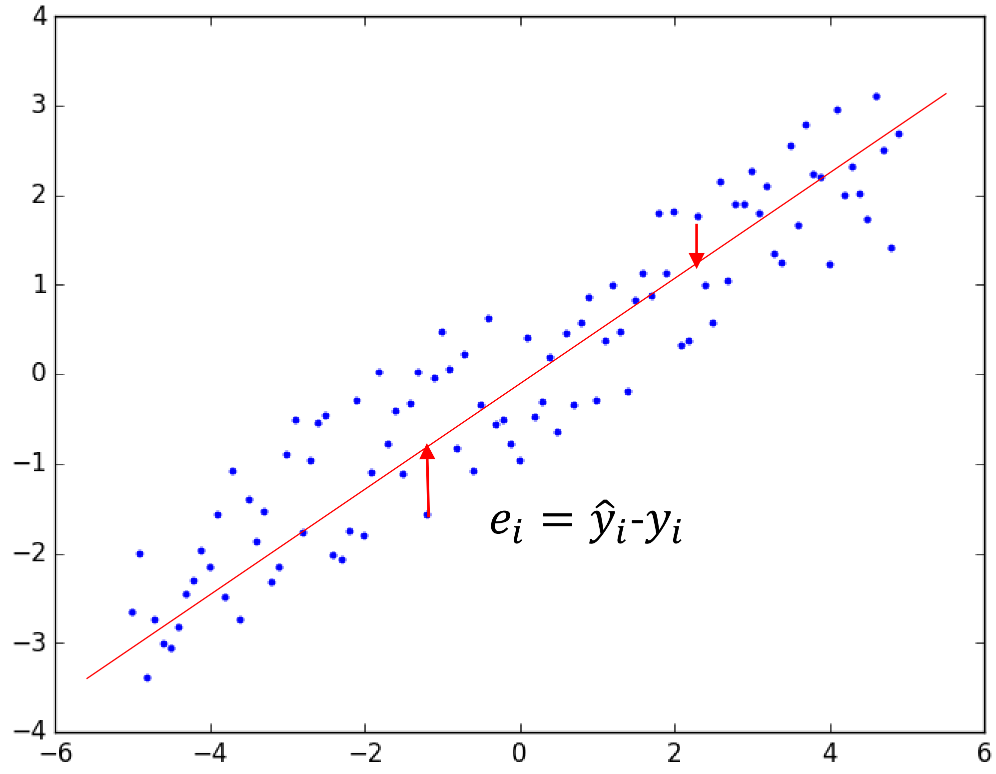


最小二乘法

- 假设

$$y = f(x)$$

$$\hat{y}_i = f(x_i)$$



想法1

$$\mathcal{E} = \sum_{i=1}^{N} {e_i}^2$$

想法2

$$\mathcal{E} = \sum_{i=1}^{N} |e_i|$$

想法n

假设1 $y = f(x)$= $ax + b$

假设2 $y = f(x)$= $ax^2 + bx + c$

如果X的观察也不准确呢？
如果我们知道Y的观察比X的准1倍呢？

如何知道因该用几阶呢？

Maximum Likelihood

- 假设$\vec{\theta}$，使得

$$\hat{y}_i = f(x_i, \vec{\theta}) \sim N(y_i, \sigma)$$

$$\mathrm{P}(\vec{\mathrm{Y}} \mid \vec{X}, \vec{\theta})$$

实践是检验真理的唯一标准

- …坚持实践是检验真理的唯一标准，就是坚持马克思主义，坚持辨证唯物主义…详见《毛泽东选集》第5卷291页
- 理论具有重要指导意义
  - 上下限 － 少走弯路
  - 极限 － 理想与现实
  - 敏感性 － 抓住要矛盾


光明日報

GUANG MING RIBAO

华主席给金主席的感谢电

华主席离平壤回国金主席到车站热烈欢送

金主席向华主席赠送礼品

实践是检验真理的唯一标准

Homework

- 随机样本生成：针对以下3种不同的分布函数分别生成随机样本各N个，计算并画出直方图（用一幅图展示）。设置5个不同的N=10，30，100，300，1000。
  - 均匀分布[-1，1]
  - 正态分布N（0，1）
  - 伯努利分布p=0.3
- 用模拟的方法验证大数定律：对以上不同分布生成的随机样本分别求平均E(N)，画出E(N)（用一幅图）
- 用模拟的方法验证中心极限定律：**自我设计**

Homework

- 玩转最小二乘法：针对以下3种真实模型（噪音$\varepsilon \sim N(0,1)$，x在[-3，3]均匀取N=100个值，），采用3种最小二乘法模型(1次、2次和3次），分别拟合，画出散点和拟合曲线（共9张图）
  - $Y = 2x + 1 + \varepsilon$
  - $Y = 0.01x^2 + 2x + 1 + \varepsilon$
  - $Y = 0.1e^{0.1x} + 2x + 1 + \varepsilon$
- 对X添加噪音$\varepsilon \sim N(0,0.2)$,重复以上,观察结果变化。
- 添加（0，3）作为观察样本，重复以上,观察结果变化。
- 添加（4，0）作为观察样本，重复以上,观察结果变化。