信息学中的概率统计

王若松

前沿计算研究中心
北京大学

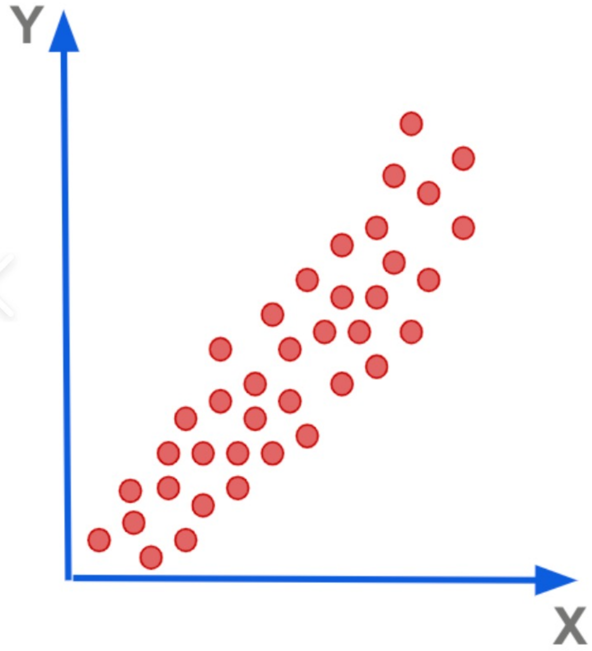
1. 回归分析

- **点估计**：估计分布中所含有的未知参数
- **回归分析**：估计变量之间的关系

- 例1：给定同一个电阻不同电流下电压的测量数据，估计电压和电流间的关系
- 例2：估计广告投入和利润回报间的关系

- 给定数据$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，估计$y$与$x$的关系

- **线性相关关系**： $y = \alpha + \beta x + \epsilon$



1. 回归分析

- **回归分析**：估计变量之间的关系
- 给定数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，估计$y$与$x$的关系

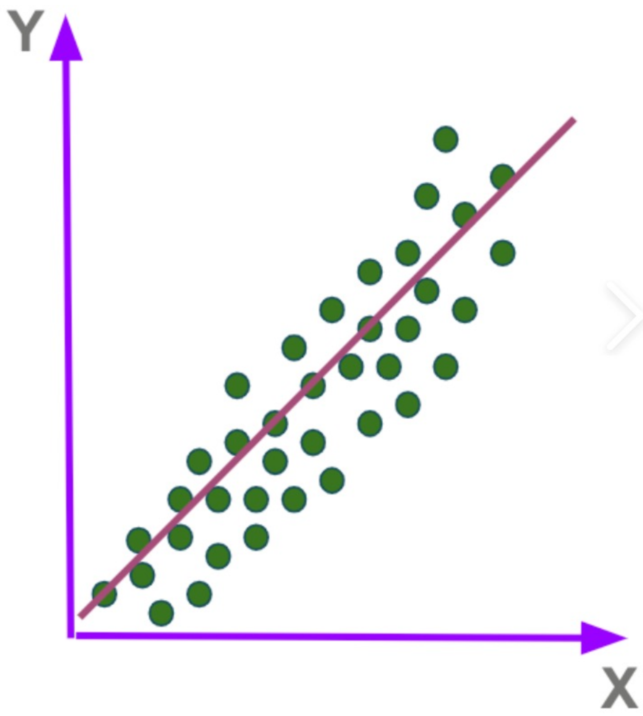
- **线性相关关系**： $y = \alpha + \beta x + \epsilon$
  - α与β为需要估计的未知参数
  - ϵ为误差，$E(\epsilon) = 0$， $\mathrm{Var}(\epsilon) = \sigma^2$， σ^2为未知参数
  - x可以精确测量或严格控制

- 目标：利用数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，给出$\alpha$和$\beta$的估计$\hat{\alpha}$和$\hat{\beta}$
- 假设：$y_i = \alpha + \beta x_i + \epsilon_i$
  - $E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$， σ^2为未知参数，且ϵ_i相互独立

北京大学 PEKING UNIVERSITY

1. 回归分析

- 目标：利用数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，给出$\alpha$和$\beta$的估计$\hat{\alpha}$和$\hat{\beta}$
- 假设：$y_i = \alpha + \beta x_i + \epsilon_i$， $E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立
- 给定估计$\hat{\alpha}$和$\hat{\beta}$
- **经验回归函数**：$\hat{y} = \hat{\alpha} + \hat{\beta} x$
- 给定$x = x_0$，$\hat{y}_0 = \hat{\alpha} + \hat{\beta} x_0$为**预测值或拟合值**



2. 最小二乘估计

- 目标：利用数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，给出$\alpha$和$\beta$的估计$\hat{\alpha}$和$\hat{\beta}$

- 定义：$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \beta x_i - \alpha)^2$
- 选择α和β，最小化$Q(\alpha, \beta)$。称得到的$\hat{\alpha}$和$\hat{\beta}$为**最小二乘估计**

- 计算问题：给定数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，如何计算最小二乘估计
- 统计性质：最小二乘估计有哪些统计性质
- 预测：给定新数据x_0，如何估计y_0

2. 最小二乘估计

- 目标：利用数据$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$，给出$\alpha$与$\beta$的估计$\hat{\alpha}$和$\hat{\beta}$
- 定义：$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \beta x_i - \alpha)^2$
- 选择α和β，最小化$Q(\alpha, \beta)$。称得到的$\hat{\alpha}$和$\hat{\beta}$为**最小二乘估计**

- **正规方程：**
- $\frac{\partial Q}{\partial \beta} = -2\sum x_i \cdot (y_i - \beta x_i - \alpha) = 0$
- $\frac{\partial Q}{\partial \alpha} = -2\sum (y_i - \beta x_i - \alpha) = 0$

2. 最小二乘估计

- **正规方程：**
- $\frac{\partial Q}{\partial \beta} = -2\sum x_i \cdot (y_i - \beta x_i - \alpha) = 0$
- $\frac{\partial Q}{\partial \alpha} = -2\sum (y_i - \beta x_i - \alpha) = 0$

- $\frac{\partial Q}{\partial \alpha} = 0 \Rightarrow \alpha = \frac{1}{n}(\sum y_i - \beta \sum x_i)$
- $\frac{\partial Q}{\partial \beta} = 0 \Rightarrow \sum x_i \cdot (y_i - \beta x_i - \alpha) = 0 \Rightarrow \sum x_i y_i - \beta \sum x_i^2 - \frac{1}{n}\sum x_i \cdot (\sum y_i - \beta \sum x_i) = 0$

- $\hat{\beta} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \cdot \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}, \hat{\alpha} = \frac{1}{n}\sum y_i - \hat{\beta} \cdot \frac{1}{n}\sum x_i$

2. 最小二乘估计

- 目标：利用数据$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$，给出$\alpha$与$\beta$的估计$\hat{\alpha}$和$\hat{\beta}$
- 定义：$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \beta x_i - \alpha)^2$
- 选择α和β，最小化$Q(\alpha, \beta)$。称得到的$\hat{\alpha}$和$\hat{\beta}$为**最小二乘估计**
- $\hat{\beta} = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \cdot \sum y_i}{\sum x_i^2 - \frac{1}{n}(\sum x_i)^2}, \hat{\alpha} = \frac{1}{n}\sum y_i - \hat{\beta} \cdot \frac{1}{n}\sum x_i$

- 定义 $\bar{x} = \frac{1}{n}\sum x_i, \bar{y} = \frac{1}{n}\sum y_i$
- $s_{xx} = \sum(x_i - \bar{x})(x_i - \bar{x}) = \sum x_i^2 - n \cdot (\bar{x})^2$
- $s_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \sum x_i \cdot \bar{y} - \sum y_i \cdot \bar{x} + n \cdot \bar{x} \cdot \bar{y} = \sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}$

- $\hat{\beta} = \frac{s_{xy}}{s_{xx}}, \hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$

PEKING UNIVERSITY 北京大学

2. 最小二乘估计

- 假设：$y_i=\alpha+\beta x_i+\epsilon_i$，$E(\epsilon_i)=0$，$\mathrm{Var}(\epsilon_i)=\sigma^2$，且$\epsilon_i$相互独立

- 注意到$\sum(x_i-\overline{x})=0$
- $s_{xy}=\sum(x_i-\overline{x})(y_i-\overline{y})=\sum(x_i-\overline{x})(\alpha+\beta x_i+\epsilon_i-\overline{y})=\sum(x_i-\overline{x})(\beta x_i+\epsilon_i)$

- $\hat{\beta}=\frac{\sum(x_i-\overline{x})(\beta x_i+\epsilon_i)}{s_{xx}}=\sum\epsilon_i\cdot\frac{(x_i-\overline{x})}{s_{xx}}+\frac{\sum(x_i-\overline{x})\cdot\beta x_i}{s_{xx}}$
- $\quad=\sum\epsilon_i\cdot\frac{(x_i-\overline{x})}{s_{xx}}+\frac{\sum(x_i-\overline{x})\cdot\beta(x_i-\overline{x})}{s_{xx}}=\sum\epsilon_i\cdot\frac{(x_i-\overline{x})}{s_{xx}}+\beta$

- $\hat{\alpha}=\overline{y}-\hat{\beta}\cdot\overline{x}=\beta\cdot\overline{x}+\alpha+\frac{1}{n}\sum\epsilon_i-\hat{\beta}\cdot\overline{x}=\alpha+\frac{1}{n}\sum\epsilon_i+\left(\beta-\hat{\beta}\right)\cdot\overline{x}$
- $\quad=\alpha+\sum\epsilon_i\cdot\left(\frac{1}{n}-\frac{x_i-\overline{x}}{s_{xx}}\cdot\overline{x}\right)$

2. 最小二乘估计

- 假设：$y_i=\alpha+\beta x_i+\epsilon_i$，$E(\epsilon_i)=0$， $\mathrm{Var}(\epsilon_i)=\sigma^2$，且$\epsilon_i$相互独立
- $\hat{\beta}=\beta+\sum\epsilon_i\cdot\frac{(x_i-\bar{x})}{s_{xx}}, \hat{\alpha}=\alpha+\sum\epsilon_i\cdot\left(\frac{1}{n}-\frac{(x_i-\bar{x})}{s_{xx}}\cdot\bar{x}\right)$
- $E(\hat{\beta})=\beta, E(\hat{\alpha})=\alpha$
- $s_{xx}=\sum(x_i-\bar{x})(x_i-\bar{x})$
- $\mathrm{MSE}(\hat{\beta})=\mathrm{Var}(\hat{\beta})=\sum\mathrm{Var}(\epsilon_i)\cdot\left(\frac{(x_i-\bar{x})}{s_{xx}}\right)^2=\frac{\sigma^2}{s_{xx}}$
- $\mathrm{MSE}(\hat{\alpha})=\mathrm{Var}(\hat{\alpha})=\sigma^2\cdot\sum\left(\frac{1}{n}-\frac{(x_i-\bar{x})}{s_{xx}}\cdot\bar{x}\right)^2=\sigma^2\sum\left(\frac{1}{n^2}-\frac{2}{n}\cdot\frac{(x_i-\bar{x})}{s_{xx}}\cdot\bar{x}+\left(\frac{(x_i-\bar{x})}{s_{xx}}\cdot\bar{x}\right)^2\right)$
- $\qquad\qquad\qquad=\sigma^2\left(\frac{1}{n}+\frac{(\bar{x})^2}{s_{xx}}\right)$

2. 最小二乘估计

- 假设：$y_i = \alpha + \beta x_i + \epsilon_i$，$E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立

- $\hat{\beta} = \sum \epsilon_i \cdot \frac{(x_i - \bar{x})}{s_{xx}} + \beta, \hat{\alpha} = \alpha + \sum \epsilon_i \cdot \left(\frac{1}{n} - \frac{(x_i - \bar{x})}{s_{xx}} \cdot \bar{x}\right)$
- $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}, \mathrm{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x})^2}{s_{xx}}\right)$

- $\mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = E\left(\sum_i \epsilon_i \cdot \left(\frac{1}{n} - \frac{(x_i - \bar{x})}{s_{xx}} \cdot \bar{x}\right) \cdot \sum_j \epsilon_j \cdot \frac{(x_j - \bar{x})}{s_{xx}}\right)$
- $= E\left(\sum_i \epsilon_i^2 \cdot \left(\frac{1}{n} - \frac{(x_i - \bar{x})}{s_{xx}} \cdot \bar{x}\right) \cdot \frac{(x_i - \bar{x})}{s_{xx}}\right)$
- $= E\left(\sum_i \epsilon_i^2 \cdot \frac{1}{n} \cdot \frac{(x_i - \bar{x})}{s_{xx}}\right) - E\left(\sum_i \epsilon_i^2 \cdot \bar{x} \cdot \left(\frac{(x_i - \bar{x})}{s_{xx}}\right)^2\right) = -\sigma^2 \cdot \frac{\bar{x}}{s_{xx}}$

2. 最小二乘估计

- 假设：$y_i=\alpha+\beta x_i+\epsilon_i$，$E(\epsilon_i)=0$， $\mathrm{Var}(\epsilon_i)=\sigma^2$，且$\epsilon_i$相互独立
- $\hat{\beta}=\sum\epsilon_i\cdot\frac{(x_i-\bar{x})}{s_{xx}}+\beta,\hat{\alpha}=\alpha+\sum\epsilon_i\cdot\left(\frac{1}{n}-\frac{(x_i-\bar{x})}{s_{xx}}\cdot\bar{x}\right)$
- $\mathrm{Var}(\hat{\beta})=\frac{\sigma^2}{s_{xx}},\mathrm{Var}(\hat{\alpha})=\sigma^2\left(\frac{1}{n}+\frac{(\bar{x})^2}{s_{xx}}\right),\mathrm{Cov}(\hat{\alpha},\hat{\beta})=-\sigma^2\cdot\frac{\bar{x}}{s_{xx}}$
- 令 $\hat{y}_i=\hat{\alpha}+\hat{\beta}x_i$为 x_i的预测值，$E(\hat{y}_i)=\alpha+\beta x\Rightarrow E(\hat{y}_i-y_i)=0$
- $\mathrm{Var}(\hat{y}_i)=\mathrm{Var}(\hat{\alpha}+\hat{\beta}x_i)=\mathrm{Var}(\hat{\alpha})+x_i^2\cdot\mathrm{Var}(\hat{\beta})+2x_i\cdot\mathrm{Cov}(\hat{\alpha},\hat{\beta})$
- $=\sigma^2\left(\frac{1}{n}+\frac{(\bar{x})^2}{s_{xx}}\right)+x_i^2\cdot\frac{\sigma^2}{s_{xx}}-2x_i\cdot\sigma^2\cdot\frac{\bar{x}}{s_{xx}}$
- $=\sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_i)^2}{s_{xx}}\right)$

2. 最小二乘估计

- 假设：$y_i = \alpha + \beta x_i + \epsilon_i$，$E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立
- $\hat{\beta} = \sum \epsilon_i \cdot \frac{(x_i - \bar{x})}{s_{xx}} + \beta, \hat{\alpha} = \alpha + \sum \epsilon_i \cdot \left(\frac{1}{n} - \frac{(x_i - \bar{x})}{s_{xx}} \cdot \bar{x}\right)$
- $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}, \mathrm{Var}(\hat{\alpha}) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x})^2}{s_{xx}}\right), \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \cdot \frac{\bar{x}}{s_{xx}}$
- 令 $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$为 x_i的预测值，$E(\hat{y}_i) = \alpha + \beta x$， $\mathrm{Var}(\hat{y}_i) = \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x} - x_i)^2}{s_{xx}}\right)$

- $\mathrm{Var}(\hat{y}_i - y_i) = \mathrm{Var}(\hat{y}_i) + \mathrm{Var}(y_i) - 2\mathrm{Cov}(\hat{y}_i, y_i)$
- $= \mathrm{Var}(\hat{y}_i) + \sigma^2 - 2\mathrm{Cov}(\hat{\alpha} + \hat{\beta} x_i, \epsilon_i)$
- $= \sigma^2 \left(\frac{1}{n} + \frac{(\bar{x} - x_i)^2}{s_{xx}}\right) + \sigma^2 - 2\sigma^2 \left(\frac{1}{n} - \frac{(x_i - \bar{x})}{s_{xx}} \cdot \bar{x}\right) - 2\sigma^2 \frac{(x_i - \bar{x})}{s_{xx}} \cdot x_i$
- $= \frac{n-1}{n}\sigma^2 + \frac{(\bar{x} - x_i)^2}{s_{xx}}\sigma^2 - 2\sigma^2 \frac{(x_i - \bar{x})}{s_{xx}} \cdot (x_i - \bar{x}) = \frac{n-1}{n}\sigma^2 - \frac{(\bar{x} - x_i)^2}{s_{xx}}\sigma^2$

2. 最小二乘估计

- 假设：$y_i=\alpha+\beta x_i+\epsilon_i$，$E(\epsilon_i)=0$， $\mathrm{Var}(\epsilon_i)=\sigma^2$，且$\epsilon_i$相互独立

- 令 $\hat{y}_i=\hat{\alpha}+\hat{\beta}x_i$为 x_i的预测值，$E(\hat{y}_i)=\alpha+\beta x_i \Rightarrow E(\hat{y}_i-y_i)=0$
- $\mathrm{Var}(\hat{y}_i-y_i)=\frac{n-1}{n}\sigma^2-\frac{(\overline{x}-x_i)^2}{s_{xx}}\sigma^2$
- $E(\sum(\hat{y}_i-y_i)^2)=\sum\mathrm{Var}(\hat{y}_i-y_i)=(n-1)\sigma^2-\frac{\sum(\overline{x}-x_i)^2}{s_{xx}}\sigma^2=(n-2)\sigma^2$

- $E\left(\frac{1}{n-2}\sum(\hat{y}_i-y_i)^2\right)=\sigma^2$，也即$s^2=\frac{1}{n-2}\sum(\hat{y}_i-y_i)^2$为 σ^2的无偏估计量

2. 最小二乘估计

- 假设：$y_i = \alpha + \beta x_i + \epsilon_i$，$E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立
- 更强的假设： $y_i = \alpha + \beta x_i + \epsilon_i$， $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立

- 在新假设下
  - α, β, σ^2的最大似然估计?
  - 给定新数据x_0，给出y_0的置信区间?

- 似然函数$L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}}$
- 对数似然函数$\ln L(\alpha, \beta, \sigma^2) = \sum_{i=1}^{n} -\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2} - n \cdot \frac{\ln 2\pi}{2} - n \cdot \frac{\ln \sigma^2}{2}$

2. 最小二乘估计

PEKING UNIVERSITY 北京大学

- 假设： $y_i = \alpha + \beta x_i + \epsilon_i$， $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立
- 对数似然函数$\ln L(\alpha, \beta, \sigma^2) = \sum_{i=1}^{n} -\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma^2} - n \cdot \frac{\ln(2\pi)}{2} - n \cdot \frac{\ln \sigma^2}{2}$

- 对于固定的σ^2，最大化似然函数等价于最大化$-\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$
- 等价于最小二乘估计$\hat{\alpha}$和$\hat{\beta}$

- σ^2的最大似然估计?
- $\frac{\partial L}{\partial \sigma^2} = \frac{(y_i-\alpha-\beta x_i)^2}{2(\sigma^2)^2} - \frac{n}{2\sigma^2} = 0$
- $\widehat{\sigma^2}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\alpha} - \hat{\beta} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
- $\widehat{\sigma^2}_{\mathrm{MLE}}$为有偏估计

2. 最小二乘估计

- 假设： $y_i = \alpha + \beta x_i + \epsilon_i$， $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立
- 最小二乘估计$\hat{\alpha}$和$\hat{\beta}$服从何种分布?

- 回顾:
  - $\hat{\beta} = \sum \epsilon_i \cdot \frac{(x_i - \overline{x})}{s_{xx}} + \beta, \hat{\alpha} = \alpha + \sum \epsilon_i \cdot \left(\frac{1}{n} - \frac{(x_i - \overline{x})}{s_{xx}} \cdot \overline{x}\right)$
  - $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}, \mathrm{Var}(\hat{\alpha}) = \sigma^2\left(\frac{1}{n} + \frac{(\overline{x})^2}{s_{xx}}\right), \mathrm{Cov}(\hat{\alpha}, \hat{\beta}) = -\sigma^2 \cdot \frac{\overline{x}}{s_{xx}}$

- 若 $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立，则$\hat{\alpha}$和$\hat{\beta}$服从二维高斯分布
- 作业：令 $s^2 = \frac{1}{n-2} \sum (\hat{y}_i - y_i)^2$，则有$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$，且$s^2$与$\hat{\alpha}$和$\hat{\beta}$独立

2. 最小二乘估计

- 假设： $y_i = \alpha + \beta x_i + \epsilon_i$， $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立
- 给定新数据x_0，给出$E(y_0) = \alpha + \beta x_0$的估计
- 点估计： $\hat{\alpha} + \hat{\beta} x_0$
  - $E\left(\hat{\alpha} + \hat{\beta} x_0\right) = \alpha + \beta x_0, \mathrm{Var}\left(\hat{\alpha} + \hat{\beta} x_0\right) = \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x} - x_0)^2}{s_{xx}}\right)$
  - $\hat{\alpha} + \hat{\beta} x_0 \sim N\left(\alpha + \beta x_0, \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x} - x_0)^2}{s_{xx}}\right)\right)$
- 给出 $\alpha + \beta x_0$的置信区间
  - $\frac{\hat{\alpha} + \hat{\beta} x_0 - (\alpha + \beta x_0)}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(\overline{x} - x_0)^2}{s_{xx}}\right)}} \sim N(0,1)$， $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$，且$s^2$与$\hat{\alpha}$和$\hat{\beta}$独立
  - 枢轴量$G = \frac{\hat{\alpha} + \hat{\beta} x_0 - (\alpha + \beta x_0)}{\sqrt{\frac{s^2}{\sigma^2}} \cdot \sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(\overline{x} - x_0)^2}{s_{xx}}\right)}} = \frac{\hat{\alpha} + \hat{\beta} x_0 - (\alpha + \beta x_0)}{\sqrt{s^2} \cdot \sqrt{\frac{1}{n} + \frac{(\overline{x} - x_0)^2}{s_{xx}}}} \sim t(n-2)$

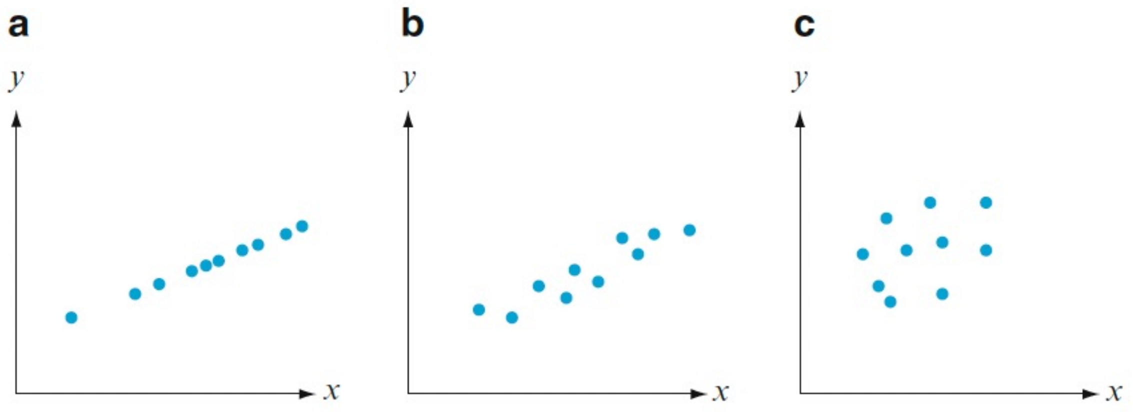
2. 最小二乘估计

- 假设： $y_i = \alpha + \beta x_i + \epsilon_i$， $\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立
- 给定新数据x_0，给出$y_0 = \alpha + \beta x_0 + \epsilon_0$的区间估计
  - $\epsilon_0 \sim N(0, \sigma^2)$，且与$\epsilon_i$相互独立
- $\hat{\alpha} + \hat{\beta} x_0 - y_0 \sim N\left(0, \sigma^2\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{s_{xx}} + 1\right)\right)$
- $\frac{\hat{\alpha}+\hat{\beta}x_0-y_0}{\sqrt{\sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{s_{xx}}+1\right)}} \sim N(0,1)$， $\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2)$，且$s^2$与$\hat{\alpha}$和$\hat{\beta}$独立
- 枢轴量$G = \frac{\hat{\alpha}+\hat{\beta}x_0-y_0}{\sqrt{\frac{s^2}{\sigma^2}} \cdot \sqrt{\sigma^2\left(\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{s_{xx}}+1\right)}} = \frac{\hat{\alpha}+\hat{\beta}x_0-y_0}{\sqrt{s^2} \cdot \sqrt{\frac{1}{n}+\frac{(\bar{x}-x_0)^2}{s_{xx}}+1}} \sim t(n-2)$

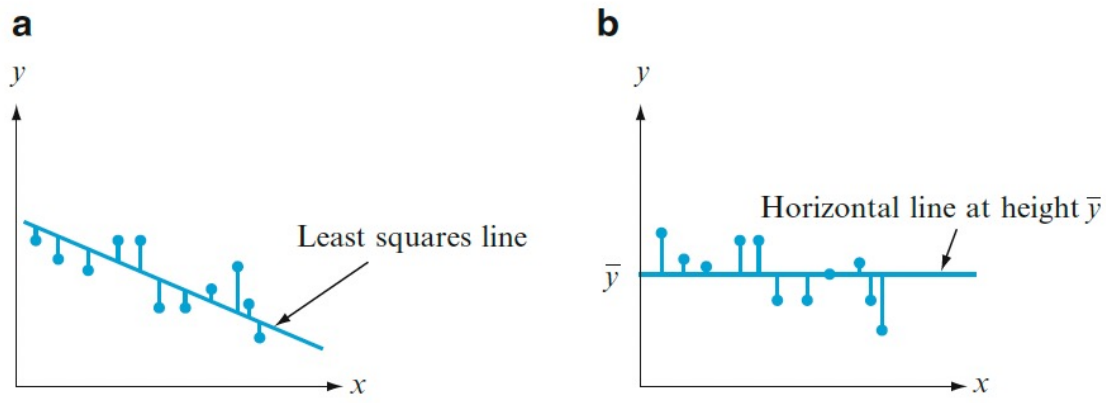
PEKING UNIVERSITY

2. 最小二乘估计

- **确定系数**$r^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$
- 常用于评估模型的有效性



- $\mathrm{SST} = \sum(y_i - \bar{y})^2$：总平方和
- $\mathrm{SSE} = \sum(y_i - \hat{y}_i)^2$：残差平方和
  - $y_i - \hat{y}_i$：残差
- $\mathrm{SSR} = \sum(\hat{y}_i - \bar{y})^2$：回归平方和



- 作业：$\mathrm{SST} = \mathrm{SSE} + \mathrm{SSR}$
- $r^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = \frac{\mathrm{SSR}}{\mathrm{SST}}$，有$0 \le r^2 \le 1$
- r^2描述了总平方和中，回归平方和所占的比例，也即可被模型解释的比例

3. 多元线性回归

- 多元线性回归
  - 给定$(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots, (\boldsymbol{x_n}, y_n)$， $\boldsymbol{x_i} \in \mathbb{R}^d$， $y_i \in \mathbb{R}$
  - 定义$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{\beta}^T \boldsymbol{x_i})^2$
  - 选择$\boldsymbol{\beta}$，最小化$Q(\boldsymbol{\beta})$。称得到的$\hat{\boldsymbol{\beta}}$为最小二乘估计

- 回顾：对于一元线性回归，$Q(\alpha, \beta) = \sum_{i=1}^{n}(y_i - \beta x_i - \alpha)^2$
  - 如何处理α?

- 假设：$y_i = \boldsymbol{\beta}^T \boldsymbol{x_i} + \epsilon_i$， $E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立

3. 多元线性回归

- 给定$(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)$， $\boldsymbol{x}_i \in \mathbb{R}^d$， $y_i \in \mathbb{R}$
- 定义$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)^2$
- 选择$\boldsymbol{\beta}$，最小化$Q(\boldsymbol{\beta})$。称得到的$\hat{\boldsymbol{\beta}}$为最小二乘估计

- 令矩阵$X \in \mathbb{R}^{n \times d}$， X的第i行为$\boldsymbol{x}_i$，$\boldsymbol{y} \in \mathbb{R}^n$
- $Q(\boldsymbol{\beta}) = |\boldsymbol{y} - X\boldsymbol{\beta}|^2$

- **正规方程：**
- $\nabla Q = -2X^T(\boldsymbol{y} - X\boldsymbol{\beta}) = 2X^T X\boldsymbol{\beta} - 2X^T\boldsymbol{y} = 0$

- $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}$

3. 多元线性回归

- 假设：$y_i = \boldsymbol{\beta}^T \boldsymbol{x_i} + \epsilon_i$， $E(\epsilon_i) = 0$， $\mathrm{Var}(\epsilon_i) = \sigma^2$，且$\epsilon_i$相互独立
- 令矩阵$X \in \mathbb{R}^{n \times d}$， X的第i行为$\boldsymbol{x_i}$，$\boldsymbol{y} \in \mathbb{R}^n$，$\boldsymbol{\epsilon} \in \mathbb{R}^n$
  - $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\hat{\beta} = (X^T X)^{-1} X^T \boldsymbol{y} = (X^T X)^{-1} X^T (X\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \boldsymbol{\beta} + (X^T X)^{-1} X^T \boldsymbol{\epsilon}$
  - $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
  - $E(\boldsymbol{\epsilon} \cdot \boldsymbol{\epsilon}^T) = \sigma^2 I$
  - $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}) = E\left((X^T X)^{-1} X^T \boldsymbol{\epsilon} \cdot \left((X^T X)^{-1} X^T \boldsymbol{\epsilon}\right)^T\right) = \sigma^2 (X^T X)^{-1}$
  - $\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{\beta}} = X(X^T X)^{-1} X^T \boldsymbol{y}$
- 如何计算$E(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|^2)$?
  - $E(|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|^2) = \sigma^2 \mathrm{tr}((X^T X)^{-1})$

3. 多元线性回归

- 假设：$y_i = \boldsymbol{\beta}^T \boldsymbol{x_i} + \epsilon_i$，$\epsilon_i \sim N(0, \sigma^2)$，且$\epsilon_i$相互独立
- 对数似然函数$\ln L(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{n} -\frac{(y_i - \boldsymbol{\beta}^T \boldsymbol{x_i})^2}{2\sigma^2} - n \cdot \frac{\ln(2\pi)}{2} - n \cdot \frac{\ln \sigma^2}{2}$

- 对于固定的σ^2，最大化似然函数等价于最大化$-Q(\boldsymbol{\beta}) = -|\boldsymbol{y} - X\boldsymbol{\beta}|^2$
- 等价于最小二乘估计$\widehat{\boldsymbol{\beta}}$

- σ^2的最大似然估计?
- $\frac{\partial L}{\partial \sigma^2} = \frac{(y_i - \boldsymbol{\beta}^T \boldsymbol{x_i})^2}{2(\sigma^2)^2} - \frac{n}{2\sigma^2} = 0$
- $\widehat{\sigma^2}_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{\boldsymbol{\beta}}^T \boldsymbol{x_i})^2$

3. 多元线性回归

- $Q(\boldsymbol{\beta}) = |\boldsymbol{y} - X\boldsymbol{\beta}|^2$

- **正规方程：** $\nabla Q = -2X^T(\boldsymbol{y} - X\boldsymbol{\beta}) = 2X^TX\boldsymbol{\beta} - 2X^T\boldsymbol{y} = 0$
- $\widehat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\boldsymbol{y}$

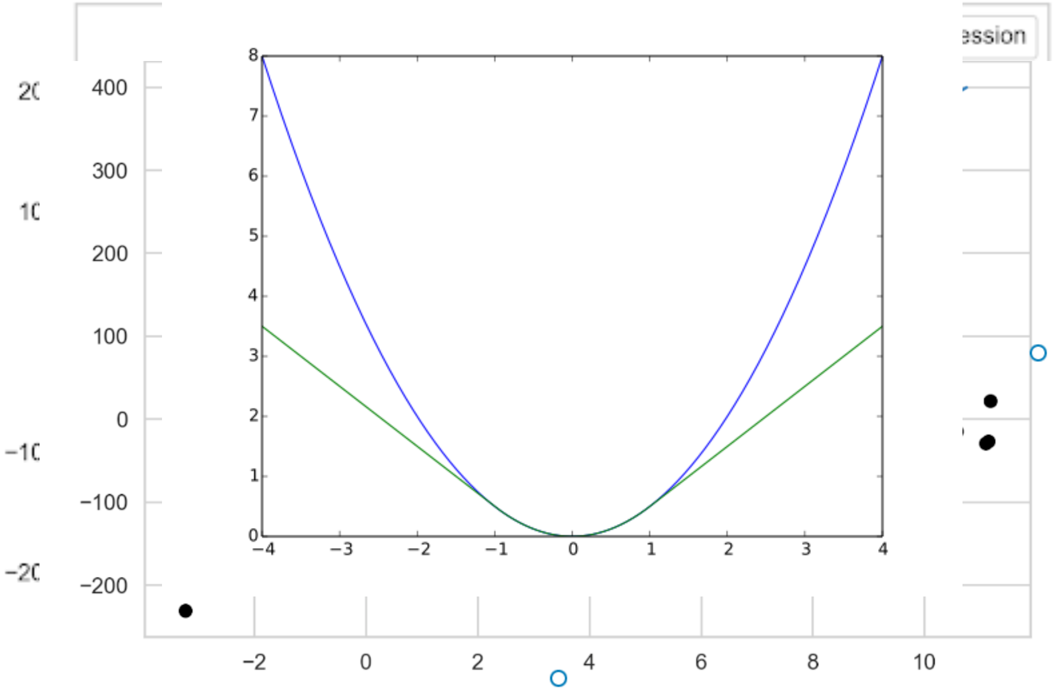
- X^TX不可逆的情况?
- X^TX何时不可逆?
- $\widehat{\boldsymbol{\beta}}$是否为无偏估计量?

- **多重共线性问题**

3. 多元线性回归

- 最小二乘估计对异常值敏感

- 鲁棒回归： $Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \rho(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)$
  - $\rho(e_i) = e_i^2$： 最小二乘估计
  - $\rho(e_i) = |e_i|$： 最小绝对偏差
  - $\rho_\epsilon(e_i) = \begin{cases} e_i^2, & |e_i| \le \epsilon \\ (2|e_i| - \epsilon)\epsilon, & |e_i| \ge \epsilon \end{cases}$： Huber回归



3. 多元线性回归

- 给定$(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \dots, (\boldsymbol{x_n}, y_n)$，$\boldsymbol{x_i} \in \mathbb{R}^d$，$y_i \in \mathbb{R}$
- 令矩阵$X \in \mathbb{R}^{n\times d}$，$X$的第$i$行为$\boldsymbol{x_i}$，$\boldsymbol{y} \in \mathbb{R}^n$
- X中某些列可能是多余的，如何进行选择?

- 方案1：枚举列的全部子集，选择效果最好的子集
  - 通常使用确定系数r^2作为评估标准

- 方案2：LASSO：$Q(\boldsymbol{\beta}) = \sum_{i=1}^n (y_i - \boldsymbol{\beta}^T \boldsymbol{x_i})^2 + \lambda \cdot \sum_{i=1}^n |\beta_i|$
  - λ为超参数，控制选取列数的数量