机器学习与数据科学

Machine Learning and Data Science

主讲: 李春光

www.pris.net.cn/teacher/lichunguang

模式识别与智能系统实验室

信息与通信工程学院 网络搜索教研中心

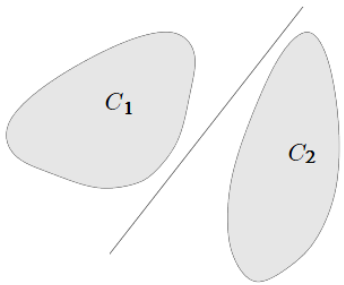
北京邮电大学

专题二：线性模型
数学基础知识补充-I

- **内容提要**
  - 凸集分离定理
  - 无约束优化问题

凸集分离定理

- 如果两个凸集分离，则可以用一张超平面将两者隔在两边
  - 也被称为超平面分离定理 (**Separating Hyperplane Theorem**)
  - 直观地看，凸集分离是指两个凸集没有交叉和重合的部分



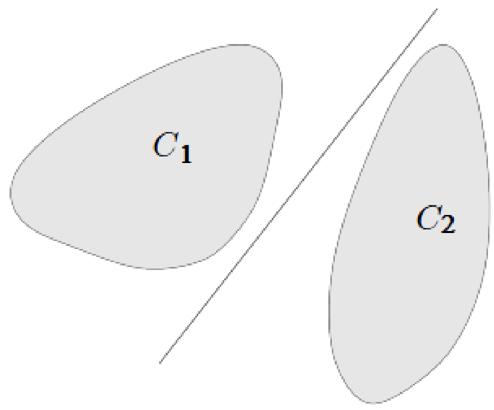
- 应用:
  - **Farkas**引理——最优性条件中最重要的基础
  - 基于**Farkas**引理，可以证明**Gordan**定理
    - Gorden定理在证明最优性条件(Kuhn-Tucker条件)起关键作用

线性可分性的等价条件

- 线性可分等价于两个凸集可分
  - 借助**Farkas**定理可以检测是否线性可分

Farkas 定理：不等式方程组 $\begin{cases} A\mathbf{w} < 0 \\ \mathbf{c}^T\mathbf{w} > 0 \end{cases}$

有解的充要条件是方程 $\begin{cases} A^T\mathbf{y} = \mathbf{c} \\ \mathbf{y} \geq 0 \end{cases}$ 无解.

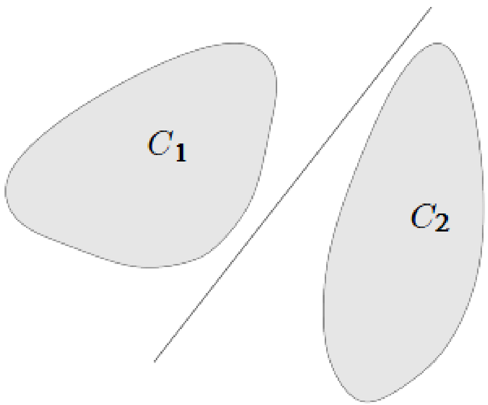


Gordan定理

- Gordan定理

设A为 $m \times n$ 矩阵，那么 $A\mathbf{x} < 0$ 有解的充要条件是：
不存在非零向量 $\mathbf{y} \geq 0$, 使得 $A^T\mathbf{y} = 0$

- 原不等式组无解 →另一组不等式组有解
  - 可用于建立最优性条件的代数表示



梯度与下降方向

- 梯度
  - 梯度：增长最快的方向
  - 方向导数：梯度与方向的内积

- 下降方向

设 $f: R^n \mapsto R$ 在点 $\bar{x} \in R^n$ 处可微。若存在 $p \in R^n$，使 $\nabla f(\bar{x})^T p < 0$，则向量 p 是 $f(x)$ 在点 $\bar{x}$ 处的下降方向

- 最速下降方向
  - 最快下降方向：梯度的反方向

有约束优化问题的最优性条件

- 几何条件：
  - 可行方向的集合与下降方向的集合相交为空集
    - 在局部最优点处，算法在寻找下降方向时发现已无路可走，即算法无法从可行集中找到下降方向
    - x* 是一个最优点，当且仅当它是一个可行点且对于所有可行向量 u 如下条件成立：

$$\nabla f\left(\mathbf{x}^*\right)^T\left(\mathbf{u}-\mathbf{x}^*\right)\geq 0$$

- 为了在计算中检验最优性条件，需要建立最优性条件的代数表示
  - KKT 条件
    - Lagrangian函数关于原变量的一阶条件
    - 原变量的可行条件
    - 对偶变量的可行条件
    - 针对不等式约束的互补松弛条件

有约束最优化问题 举例

- 无约束最优化问题: $\min\limits_{\mathbf{x}\in R^n} f(\mathbf{x})$
  - 由于 **u – x*** 是任意向量，因此得到: $\nabla f(\mathbf{x}^*)^T = 0$
- 线性等式约束最优化问题： $\min\limits_{\mathbf{x}\in R^n} f(\mathbf{x}) \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}$
  - 设**x*** 是可行解**(i.e. Ax*=b)** 且存在对偶证书**(dual certificate)**向量**v**满足：

$$A\mathbf{x}^* = \mathbf{b}, \quad \nabla f(\mathbf{x}^*) + A^T\mathbf{v} = 0$$

- 非负约束最优化问题： $\min\limits_{\mathbf{x}\in R^n} f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \geq \mathbf{0}$
  - 令**u =2x*, u=0,** 则得出最优性条件

$$\nabla f(\mathbf{x}^*)^T \mathbf{x}^* = 0, \qquad \nabla f(\mathbf{x}^*) \geq 0, \qquad \mathbf{x}^* \geq 0$$

专题二：线性模型
数学基础知识补充

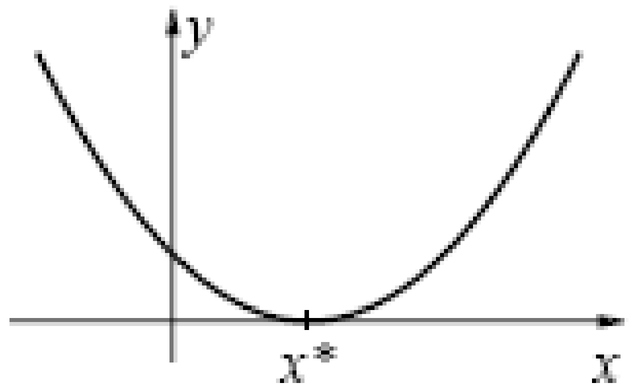
- **内容提要**
  - 凸集分离定理
  - 无约束优化问题

无约束最优化问题

- 问题定义

$$\min_{\mathbf{x}} f(\mathbf{x})$$

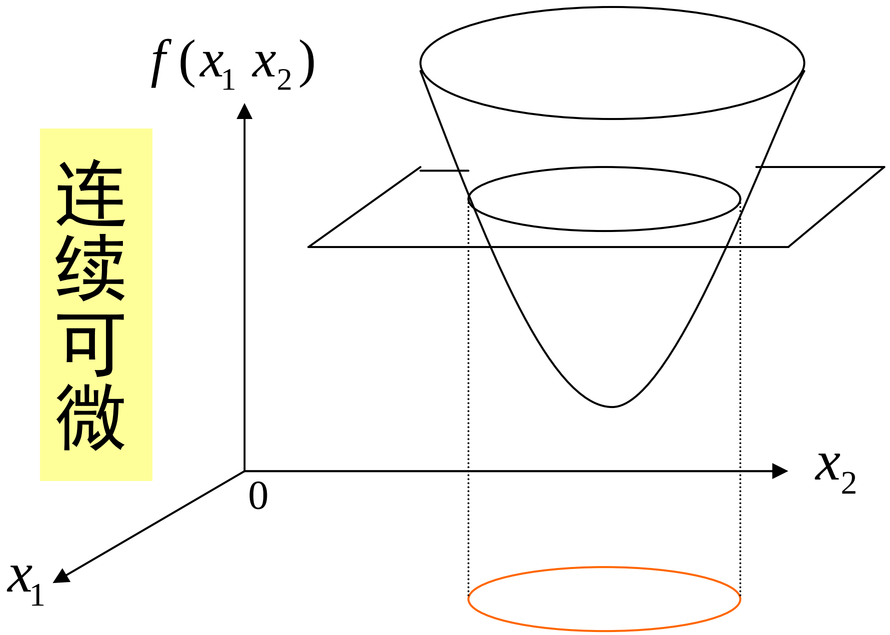
其中 $\mathbf{x} = (x_1, x_2, ..., x_n)^T \in R^n, \qquad f : R^n \to R$

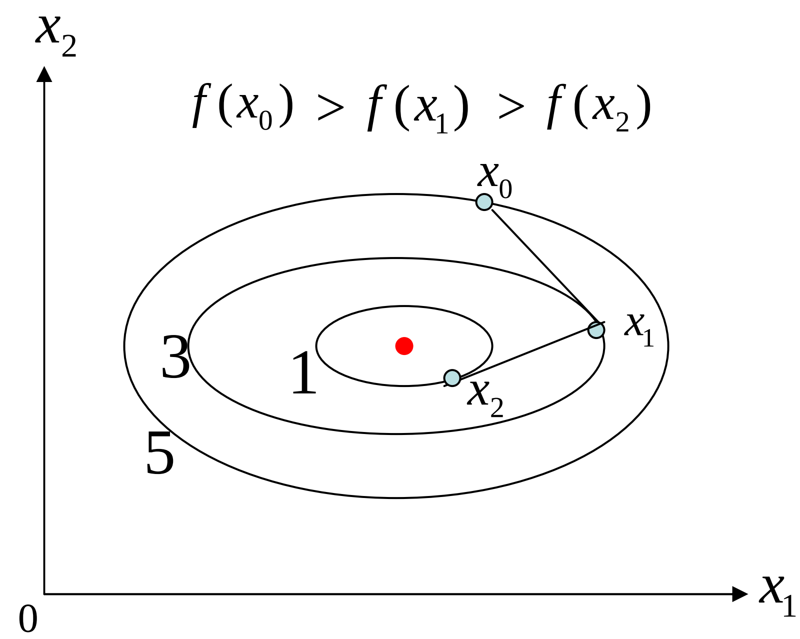


迭代下降算法基本思想

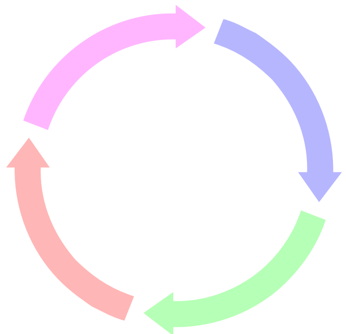
- 迭代下降算法：
  - 寻找一个搜索方向 $\mathbf{d}^{(k)}$ ，使得每次迭代时函数值减小，即 $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$ ，有 $f\left(\mathbf{x}^{(k+1)}\right) \leq f\left(\mathbf{x}^{(k)}\right)$

连续可微





迭代下降算法基本步骤



第 1 步　　选取初始点 $x^{(0)}$，k:=0;

第 2 步　　构造搜索方向 $d^{(k)}$；

第 3 步　　根据 $d^{(k)}$，确定步长 λ_k；

第 4 步　　令 $x^{(k+1)} = x^{(k)} + \lambda_k d^{(k)}$，

若 $x^{(k+1)}$ 已满足某种终止条件，停止迭代，输出近似解 $x^{(k+1)}$；否则令 $k:=k+1$，转回第 2 步。

– 初始点、搜索方向和步长参数

最速下降法

要求：目标函数 $f(x)$ 一阶连续可微

- 由柯西(Cauchy)在1847年提出的，是求无约束极值的最早的数值算法

步骤：

第 1 步　选取初始点 $x^{(0)}$，给定终止误差 $\boldsymbol{\varepsilon > 0}$，令 $\boldsymbol{k := 0}$；

第 2 步　计算 $\nabla f\left(x^{(k)}\right)$，若 $\left\|\nabla f\left(x^{(k)}\right)\right\| < \varepsilon$，停止迭代，输出 $x^{(k)}$。

否则进行第 3 步；

第 3 步　取 $d^{(k)} = -\nabla f\left(x^{(k)}\right)$

第 4 步　进行一维搜索，求 λ_k，使得

$$f\left(x^{(k)} + \lambda_k d^{(k)}\right) = \min_{\lambda \geq 0} f\left(x^{(k)} + \lambda d^{(k)}\right)$$

令 $x^{(x+1)} = x^{(k)} + \lambda_k d^{(k)}$，$\boldsymbol{k := k + 1}$，转第 2 步。

典型的无约束优化算法

- 根据搜索方向的不同，分为:
  - 最速下降法
  - 牛顿法
    - 阻尼牛顿法
    - 修正牛顿法
    - 伪(Pseudo)牛顿法
  - 共轭梯度法
  - 最小二乘法
    - 线性最小二乘
    - 非线性最小二乘
    - 修正最小二乘法