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-II

- **内容提要**
  - 矩阵分析基础
    - 迹
    - 范数
    - 矩阵的实值函数求微分
  - 凸(convex)优化
    - 凸集(convex set)
    - 凸函数(convex function)

矩阵的迹(Trace)

- 对于一个 $m\times m$ 方阵A，它的迹等于其对角线元素的和：

$$\operatorname{trace}(A)=\sum_{i=1}^{m}a_{ii}$$

- 一个$m\times n$ 的矩阵A的 Frobenius 范数(Norm):

$$\|A\|_F^2=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2$$

- 矩阵A的 Frobenius 范数与迹的关系:

$$\|A\|_F^2=\operatorname{trace}\left(A^TA\right)=\operatorname{trace}\left(AA^T\right)$$

- 矩阵迹的运算性质:

$$\operatorname{trace}(AB)=\operatorname{trace}(BA),\quad \operatorname{trace}(A)=\operatorname{trace}\left(A^T\right)$$
$$\operatorname{trace}(ABC)=\operatorname{trace}(CAB)=\operatorname{trace}(BCA)$$

矩阵的实值函数求导数

- 考虑矩阵X的实值函数

$$\nabla f\left(X\right)=\nabla trace\left(X^{T}A\right)=A$$

$$\nabla f\left(X\right)=\nabla trace\left(AX\right)=A^{T}$$

$$\nabla f\left(X\right)=\nabla trace\left(X^{T}AX\right)=AX+A^{T}X$$

$$\nabla\left\|X\right\|_{F}^{2}=\nabla trace\left(X^{T}X\right)=2X$$

参考资料

- 线性代数回顾
  - **http://www.deeplearningbook.org/contents/linear_algebra.html**

- 迹的性质和矩阵的微分：
  - **http://web.stanford.edu/~jduchi/projects/matrix_prop.pdf**

专题 二：线性模型
数学基础知识补充-II

- **内容提要**
  - 矩阵分析基础
    - 迹
    - 范数
    - 矩阵的实值函数求微分
  - 凸**(convex)**优化
    - 凸集(convex set)
    - 凸函数(convex function)

凸集和凸函数

- 凸集:
  - 给定集合**S**，对于**S**中的两个元素x_1和x_2，如果对于任意$0 \le \alpha \le 1$，有：$\alpha x_1 + \left(1-\alpha\right) x_2 \in S$
    则称**S**为凸集
- 凸函数
  - 在最优化中，即**convex** 函数，指下凸函数
  - 函数$f(x)$，对于任意$0 \le \alpha \le 1$，有：

$$f\left(\alpha x_1 + \left(1-\alpha\right) x_2\right) \le \alpha f\left(x_1\right) + \left(1-\alpha\right) f\left(x_2\right)$$

    - 称之为凸函数(convex function)

凸(convex)优化

- 对于一个最优化问题，如果目标函数为凸函数，可行集为凸集，即为凸优化问题
  - 等式约束为只能为线性约束
  - 不等式约束的可行集为凸集

- 判定条件：
  - 一阶条件

$$f(\mathbf{x}_2) \geq f(\mathbf{x}_1) + \nabla f(\mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1)$$

  - 二阶条件
    - 假设函数2阶可微. 在每一点处，Hessian矩阵为半正定

- 局部极小值点即为全局极小值点