计算机91刘青帅第一次作业

证明：范数$\|\cdot\|$的对偶范数满足范数的定义

$$\|z\|_* = \sup\{z^T x: \|x\| \le 1\} = \sup\{z^T x: \|x\| = 1\}$$

\

1.非负性

$\|z\|_* = \max_{\|x\|=1} z^T \cdot x = \max_{\|x\|=1} \|z\|_2 \cdot \|x\|_2 \cdot \cos\angle(x,z)$

由于$\|x\|_2 \ge 0, \|z\|_2 \ge 0$，并且一定存在$x$使得$\cos\angle(x,z) > 0$ 因此$\|z\|_* \ge 0$

2.正定性

$\|z\|_* = \max_{\|x\|=1} \|z\|_2 \cdot \|x\|_2 \cdot \cos\angle(x,z)$

若$\|z\|_* = 0$，假设$z \ne 0$，则$\|z\|_2 \ne 0$且$\|x\|_2 \ne 0$.则一定存在x使得$\angle(x,z) \in [0, \frac{\pi}{2})$，

即$\cos\angle(x,z) > 0$，由此可以推出$\|z\|_* \ne 0$，故假设矛盾.

因此$\|z\|_* = 0$时$z = 0$

3.齐次性

$\forall z \in R^n$和$t \in R$，$\|tz\|_* = \max_{\|x\|=1} (tz)^T \cdot x = \max_{\|x\|=1} t \cdot z^T \cdot x = t \cdot \max_{\|x\|=1} z^T \cdot x = t\|z\|_*$

因此$\forall z \in R^n$和$t \in R$，$\|tz\|_* = t\|z\|_*$

4.三角不等式

$\|z_1 + z_2\|_* = \max_{\|x\|=1} (z_1 + z_2)^T x = \max_{\|x\|=1} z_1^T x + z_2^T x \le \max_{\|x\|=1} z_1^T x + \max_{\|x\|=1} z_2^T x = \|z_1\|_* + \|z_2\|_*$

因此$\|z_1 + z_2\|_* \le \|z_1\|_* + \|z_2\|_*$

由以上四条可知，范数的对偶范数满足范数的定义

证明：$z^T x \le \|x\| \|z\|_*$

$\|z\|_* = \max_{\|x\|\le 1} z^T x = \max_{\|x\|=1} z^T x = \max_{x \ne 0} \frac{z^T x}{\|x\|} \Rightarrow \|z\|_* \ge \frac{z^T x}{\|x\|} \Rightarrow z^T x \le \|x\| \|z\|_*$

$f: \mathbb{R}^n \to \mathbb{R}$，$f(x) = log \sum_{i=1}^m exp(a_i^T x + b_i)$，其中$a_1, a_2, \cdots, a_m \in \mathbb{R}^n$，$b_1, b_2, \cdots, b_m \in \mathbb{R}$

1、求函数f的梯度$\nabla f(x)$

2、求函数f的二阶导数$\nabla^2 f(x)$

$$Df(x)=\frac{\sum_{i=1}^{m}a_i^Te^{a_i^Tx+b_i}}{\sum_{i=1}^{m}e^{a_i^Tx+b_i}}\Rightarrow\nabla f(x)=Df(x)^T=\frac{\sum_{i=1}^{m}a_ie^{a_i^Tx+b_i}}{\sum_{i=1}^{m}e^{a_i^Tx+b_i}}$$

$$\nabla^2f(x)=\frac{d}{dx}(\nabla f(x))=\frac{\sum_{i=1}^{m}a_i^Ta_ie^{a_i^Tx+b_i}\cdot\sum_{i=1}^{m}e^{a_i^Tx+b_i}-\sum_{i=1}^{m}a_ie^{a_i^Tx+b_i}\cdot\sum_{i=1}^{m}a_i^Te^{a_i^Tx+b_i}}{(\sum_{i=1}^{m}e^{a_i^Tx+b_i})^2}$$

- 令$A\in\mathbb{R}^{m\times n},B\in\mathbb{R}^{p\times n},b\in\mathbb{R}^m,d\in\mathbb{R}^p$，集合

$$P=\{x\in\mathbb{R}^n:Ax\le b\ 且Bx=d\}$$

是凸集吗？为什么？

$\forall x,y\in P,\forall\theta\in[0,1]$

$A(\theta x+(1-\theta)y)=A(\theta x)+A((1-\theta)y)=\theta Ax+(1-\theta)Ay\le\theta b+(1-\theta)b=b$

$B(\theta x+(1-\theta)y)=B(\theta x)+B((1-\theta)y)=\theta Bx+(1-\theta)By=\theta d+(1-\theta)d=d$

故$\forall x,y\in P,\forall\theta\in[0,1],\theta x+(1-\theta)y\in P$

因此P是凸集

证明：最大值函数$f(x)=\max\{x_1,x_2,\cdots,x_n\},x=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}\in\mathbb{R}^n$为凸函数

$\forall x,y\in R^n,\forall\lambda\in[0,1],f(\lambda x+(1-\lambda y))=max\lambda x_1+(1-\lambda)y_1,\lambda x_2+(1-\lambda)y_2,\cdots,\lambda x_n$

故$\forall x,y\in R^n,\forall\lambda\in[0,1]$，$f(\lambda x+(1-\lambda y))\le\lambda f(x)+(1-\lambda)f(y)$

因此最大值函数$f(x)$是凸函数

（1）函数f是凸函数当且仅当$\forall x\in C,\nabla^2f(x)$为对称半正定矩阵 作业：证明！

已知$f(y)=f(x)+\nabla^Tf(x)(y-x)+\frac{1}{2}(y-x)^T\nabla^2f(x)(y-x)$

由上式可得$f(y)-f(x)-\nabla^Tf(x)(y-x)=\frac{1}{2}(y-x)^T\nabla^2f(x)(y-x)$记为式①

“$\Leftarrow$”

$\nabla^2f(x)$为对阵半正定矩阵$\Rightarrow\forall x,y\in R^n$，$(y-x)^T\nabla^2f(x)(y-x)\ge0\overset{由①}{\Rightarrow}$

$f(y)-f(x)-\nabla^Tf(x)(y-x)\ge0\Rightarrow f(y)\ge f(x)+\nabla^Tf(x)(y-x)\Rightarrow f$是凸函数

“$\Rightarrow$”

f是凸函数 $\Rightarrow \forall x, y \in R^n, f(y) \geq f(x) + \nabla^T f(x)(y-x) \Rightarrow$

$f(y) - f(x) - \nabla^T f(x)(y-x) \geq 0 \overset{由①}{\Rightarrow} (y-x)^T \nabla^2 f(x)(y-x) \geq 0$
$\Rightarrow \nabla^2 f(x)$是对称半正定矩阵