Manifole Learning Homework 8

安捷 1601210097

2017 年 4 月 26 日

习题 (145.1a)**.** 为了证明 f_1 是凸函数, 我们只需证明 f_1 满足如下凸集定义

$$f_1(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leq \alpha f_1(\mathbf{x}_1) + (1-\alpha) f_1(\mathbf{x}_2), \tag{1}$$

其中 $\alpha \in [0,1]$. 显然，因为 $\alpha \geq 0$ 并且 $(1-\alpha) \geq 0$, 如果存在元素 $\mathbf{x}_1$ *or* $\mathbf{x}_2$ 小于 0, 那么如果 $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 > 0$, 定义 *1* 满足; 如果 $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \leq 0$, 那么 $+\infty \leq +\infty$ 也满足定义 *1*.

上面讨论了变量小于 *0* 的情况, f_1 满足凸函数的定义. 如果 $\mathbf{x}_1 > \mathbf{0}$ 并且 $\mathbf{x}_2 > \mathbf{0}$, 我们有如下推导

$$\begin{aligned}
f_1(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) &= -\left[\prod_i (\alpha x_{1i} + (1-\alpha)x_{2i})\right]^{\frac{1}{n}} \\
&= -\left[\prod_i \alpha x_{1i} + \prod_i (1-\alpha)x_{2i} + other_non_negtive_term\right]^{\frac{1}{n}} \\
&\leq -\left[\prod_i \alpha x_{1i} + \prod_i (1-\alpha)x_{2i}\right]^{\frac{1}{n}} \\
&\leq -\left[\prod_i \alpha x_{1i}\right]^{\frac{1}{n}} + \left[\prod_i (1-\alpha)x_{2i}\right]^{\frac{1}{n}} \\
&= \alpha f_1(\mathbf{x}_1) + (1-\alpha) f_1(\mathbf{x}_2)
\end{aligned}$$

习题 (145.1b)**.** 类似 *a* 我们证明 f_2 满足定义 *1*. 那么我们有如下推导

$$\begin{aligned}
\alpha f_2(\mathbf{x}_1) + (1-\alpha) f_2(\mathbf{x}_2) =& \alpha \ln(e^{x_{11}} + \ldots + e^{x_{1n}}) + (1-\alpha)\ln(e^{x_{21}} + \ldots + e^{x_{2n}}) \\
=& \ln\left[(e^{x_{11}} + \ldots + e^{x_{1n}})^{\alpha}(e^{x_{21}} + \ldots + e^{x_{2n}})^{1-\alpha}\right] \\
\geq& \ln(e^{\alpha x_{11} + (1-\alpha)x_{21}} + \ldots + e^{\alpha x_{1n} + (1-\alpha)x_{2n}}) \\
=& f_2(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2)
\end{aligned}$$

习题 (145.1c). 类似 a，我们有如下推导

$$
\begin{aligned}
f_3(\alpha\mathbf{x}_1+(1-\alpha)\mathbf{x}_2)=&((\alpha x_{11}+(1-\alpha)x_{21})^2+\ldots+(\alpha x_{1n}+(1-\alpha)x_{2n})^2)^{\frac{p}{2}}\\
\leq&((\alpha x_{11})^2+((1-\alpha)x_{21})^2+\ldots+(\alpha x_{1n})^2+((1-\alpha)x_{2n})^2)^{\frac{p}{2}}\\
=&((\alpha x_{11})^2+\ldots+(\alpha x_{1n})^2+((1-\alpha)x_{21})^2+\ldots+((1-\alpha)x_{2n})^2)^{\frac{p}{2}}\\
\leq&((\alpha x_{11})^2+\ldots+(\alpha x_{1n})^2)^{\frac{p}{2}}+(((1-\alpha)x_{21})^2+\ldots+((1-\alpha)x_{2n})^2)^{\frac{p}{2}}\\
=&\alpha f_3(\mathbf{x}_1)+(1-\alpha)f_3(\mathbf{x}_2)
\end{aligned}
$$

习题 (145.1d). 对于凸函数，我们知道，如果 $f(x)$ 和 $h(x)$ 都是单调递增的凸函数，那么 $f(h(x))$ 是凸函数. 对于 $f_4(\mathbf{x})=e^{\beta\mathbf{x}^T\mathbf{A}\mathbf{x}}$，我们假设有 $f(x)=e^{\beta x}$ 并且 $h(\mathbf{x})=\mathbf{x}^T\mathbf{A}\mathbf{x}$. 显然，$f(x)$ 单调递增的凸函数. 对于 $h(\mathbf{x})=\mathbf{x}^T\mathbf{A}\mathbf{x}$，因为 $\mathbf{A}$ 是对称半正定矩阵，$\mathbf{x}^T\mathbf{A}\mathbf{x}$ 可以写为 $\mathbf{x}^T\mathbf{A}\mathbf{x}=\lambda_1p_1^2+\ldots+\lambda_np_n^2$. *So* $h(\mathbf{x})$ 也是一个凸函数. 并且 $f_4(x)$ 是个凸函数.

习题 (145.2). 考虑凸函数 $f(x)=-\ln(x)$. 使用 *Jasen* 不等式，有

$$
\begin{aligned}
f(\alpha_1x_1+\ldots+\alpha_nx_n)\leq&\alpha_1f(x_1)+\ldots+\alpha_nf(x_n)\Leftrightarrow\\
-\ln(\alpha_1x_1+\ldots+\alpha_nx_n)\leq&-(\alpha_1\ln(x_1)+\ldots+\alpha_n\ln(x_n))\Leftrightarrow\\
x_1^{\alpha_1}\ldots x_n^{\alpha_n}\leq&\alpha_1x_1+\ldots+\alpha_nx_n
\end{aligned}
$$

习题 (145.3). 由于每一个 $f_i(x)$ 都是凸函数，我们可以将 $f(x)$ 的次梯度写成如下形式

$$
\partial f(x)=\{g=\sum_i^m g_i|f(y)\geq\sum_i^m f_i(x)+\sum_i^m <g_i,y-x>,\forall y\in C\},
$$

其中 g_i 是 $f_i(x)$ 的次梯度. 我们定义 $g=\sum_i^m g_i$，因此我们可以得到

$$
\partial f(x)=\sum_i^m\partial f_i(x) \tag{2}
$$

习题 (145.4). $f'(x;y)=\max_{g\in\partial f(x)}<g,y>$. 有

$$
\begin{aligned}
f'(x;\alpha\mathbf{y}_1+(1-\alpha)\mathbf{y}_2)=&\max_{g\in\partial f(x)}<g,\alpha\mathbf{y}_1+(1-\alpha)\mathbf{y}_2>\\
\leq&\alpha\max_{g\in\partial f(x)}<g,\mathbf{y}_1>+(1-\alpha)\max_{g\in\partial f(x)}<g,\mathbf{y}_2>\\
=&\alpha f'(x;\mathbf{y}_1)+(1-\alpha)f'(x;\mathbf{y}_2)
\end{aligned}
$$

习题 (145.5). *l2* 范数的次梯度是

$$
\partial||x||_2=\{\begin{array}{ll}\frac{x}{||x||_2} & if\ x\neq0\\ \{g|||g||_2\leq1\} & if\ x=0\end{array} \tag{3}
$$

习题 (145.6)**.** 为证明 $f^*(y)$ 是凸函数，有

$$\begin{aligned}
f^*(\alpha\mathbf{y}_1+(1-\alpha)\mathbf{y}_2) =& \sup_{\mathbf{x}\in C}(<\mathbf{x},\alpha\mathbf{y}_1+(1-\alpha)\mathbf{y}_2>-f(\mathbf{x}))\\
=& \sup_{\mathbf{x}\in C}(<\mathbf{x},\alpha\mathbf{y}_1+(1-\alpha)\mathbf{y}_2>-(\alpha+1-\alpha)f(\mathbf{x}))\\
\leq& \sup_{\mathbf{x}\in C}(<\mathbf{x},\alpha\mathbf{y}_1-\alpha f(\mathbf{x}))+\sup_{\mathbf{x}\in C}(<\mathbf{x},(1-\alpha)\mathbf{y}_2>-(1-\alpha)f(\mathbf{x}))\\
=& \alpha f^*(\mathbf{y}_1)+(1-\alpha)f^*(\mathbf{y}_2)
\end{aligned}$$

习题 (145.8)**.** 首先，有共轭函数

$$\begin{aligned}
f^*(Y) &= \sup_{||X||_{2,1}\leq 1}\{\sum_i^n <X_i,Y_i>-||X||_{2,0}\}\\
&= \sup_{||X||_{2,1}\leq 1}\{\sum_i^n ||X_i||||Y_i||-||X||_{2,0}\}
\end{aligned}$$

我们假设 X 有 k 列并且 $l2$ 范数等于 0. 因此

$$\begin{aligned}
f^*(Y) &= \sup_{||X||_{2,1}\leq 1}\{\sum_{i\in||X_i||\neq 0}||X_i||||Y_i||-||X||_{2,0}\}\\
&= \sup_{||X||_{2,1}\leq 1}\{\sum_{i\in||X_i||\neq 0}||Y_i||-k\}
\end{aligned}$$

其中 i 是满足 $||X_i||\neq 0$ 的下标，$||Y_i||$ 也是下标. 那么 $f^*(Y)$ 可以按如下方式计算

$$f^*(Y)=\sum_{i\in||X_i||\neq 0}(||Y_i||-1)$$

我们按照如下方式计算 $f^*(Y)$ 的共轭函数

$$\begin{aligned}
f^{**}(Z) =& \sup_Y\{\sum_i^n <Y_i,Z_i>-\sum_{i\in||X_i||\neq 0}(||Y_i||-1)\}\\
=& \sup_Y\{\sum_i^n ||Y_i||||Z_i||-\sum_{i\in||X_i||\neq 0}(||Y_i||-1)\}
\end{aligned}$$

其中 $||Z_i||$ 按如下方式取到最大值

$$f^{**}(Z)=\sum_{i\in||X_i||\neq 0}||Z_i||$$

习题 (146.1)**.** 我们知道 $W=\alpha\alpha^T$, 所以 $Tr(W)=\lambda_1=1$. 因为 $\lambda_1\geq 0$, 我摸可以知道 W 是半正定矩阵.

对于 $\mathbf{X}=\mathbf{0}$ 的情形, $W=uv^T$ 的核范数可以按照下式计算 $||W||_*=Tr(W^TW)=Tr(vu^Tuv^T)\leq 1$.

习题 (146.2)**.** *l2* 范数的凸包是

$$E_c(x)=\inf_w\{||w||_2+\frac{1}{2c}||w-x||^2\}$$

l2 范数的 *proximal mapping* 是

$$P_c(x)=\arg\min_w\{||w||_2+\frac{1}{2c}||w-x||^2\}$$

因此有

$$P_c(x)=\{\begin{array}{cc} x-\frac{cx}{||x||_2} & if\ c<||x||_2 \\ 0 & if\ c\geq ||x||_2 \end{array}$$

习题 (146.3)**.** 我们知道 $u=P_cf(x)\Leftrightarrow x-u\in\partial f(u)$, 那么有如下证明

$$\begin{aligned}
&x-\frac{1}{\lambda}[P_c(\lambda^2 g)(\lambda x+a)-a]\in\lambda\partial g(\lambda u+a)\Leftrightarrow\\
&\lambda x+a-P_c(\lambda^2 g)(\lambda x+a)\in\lambda^2\partial g(P_c(\lambda^2 g)(\lambda x+a))\Leftrightarrow\\
&x'-u'\in\partial(\lambda^2 g)(u')
\end{aligned}$$

习题 (146.4)**.** 类似与 *3*，我们有如下证明

$$\begin{aligned}
&x-\lambda[P_c(\lambda^{-1}g)(x/\lambda)]\in\partial g(u/\lambda)\Leftrightarrow\\
&x/\lambda-P_c(\lambda^{-1}g)(x/\lambda)\in\partial g(P_c(\lambda^{-1}g)(x/\lambda))\Leftrightarrow\\
&x'-u'\in\partial(\lambda^{-1}g)(u')
\end{aligned}$$

习题 (146.5)**.** *Proof.* S_+ 是凸集，而 procimal operator of indicator/charateristic function of a convex set is projection，于是

$$P_cf\left(X\right)=\Pi_{S_+}X=\arg\min_{S_+}\|W-X\|_2^2 \tag{4}$$

其中，Π 是投影算子。 □

习题 (146.6)**.** *Proof.* 令 $h\left(u\right)=I_L\left(u\right),h^*\left(u\right)=I_{L^\perp}\left(u\right)$

$$I_L\left(u\right)=\begin{cases} 0, & x\in L\\ +\infty, & x\notin L\end{cases} \tag{5}$$

那么，$P_L\left(x\right)$ 就是 x 在 L 上的映射，$P_{L_\perp}\left(x\right)$ 就是 x 在 $L_\perp$ 上的映射。于是有 $x=P_L\left(x\right)=P_{L_\perp}\left(x\right)$ □