Manifold Learning and Sparse Representation: Assignment 8

ZHANG Yuan, 1601111332

Exercise 145

1.(a)

Proof. Prove by definition, i.e., $f_1(\alpha x+(1-\alpha)y) \leq \alpha f_1(x)+(1-\alpha)f_1(y)$. If there exists $x_i \leq 0$ or $y_i \leq 0$, the proof is trivial. Therefore, we only focus on the case where $x \succ 0$ and $y \succ 0$.

$$\begin{aligned} f_1(\alpha x+(1-\alpha)y) &= -\prod_i(\alpha x_i+(1-\alpha)y_i)^{\frac{1}{n}} \\ &\leq -\prod_i(\alpha x_i)^{\frac{1}{n}} - \prod_i((1-\alpha)y_i)^{\frac{1}{n}} \\ &= -\alpha\prod_i(x_i)^{\frac{1}{n}} - (1-\alpha)\prod_i(y_i)^{\frac{1}{n}} \\ &= \alpha f_1(x)+(1-\alpha)f_1(y) \end{aligned}$$

□

1.(b)

Proof.

$$\begin{aligned} \alpha f_2(x)+(1-\alpha)f_2(y) =& \alpha\ln(e^{x_1}+\ldots+e^{x_n})+(1-\alpha)\ln(e^{y_1}+\ldots+e^{y_n}) \\ =& \ln(e^{x_1}+\ldots+e^{x_n})^{\alpha}(e^{y_1}+\ldots+e^{y_n})^{1-\alpha} \\ \geq& \ln(e^{\alpha x_1+(1-\alpha)y_1}+\ldots+e^{\alpha x_n+(1-\alpha)y_n}) \\ =& f_2(\alpha x+(1-\alpha)y) \end{aligned}$$

□

1.(c)

Proof. Since Hessian matrix of $g(x)=x^T Ax$ is p.s.d., it is a convex function. Besides, $h(y)=e^{\beta y}$ is a convex function and monotonically increasing as well. We can conclude that $f_4(x)=h(g(x))$ is convex. □

2.

Proof.

$$LHS = \exp(\sum_i \alpha_i ln x_i) \leq \exp(ln\sum_i \alpha_i x_i) = RHS$$

where the inequality holds thanks to Jasen inequality and therefore the equality holds iff. $x_1 = x_2 = ... = X_n$. □

3.

We can easily verify that

$$f(y) = \sum_i f_i(y) \geq \sum_i^m f_i(x) + \sum_i^m < g_i, y - x >$$

where g_i is the sub-gradient of f_i. Therefore, the following holds (while I cannot figure out the other way around)

$$\sum_i \partial f_i \subset \partial \sum_i f_i = \partial f.$$

4.

Proof. The proposition can be proved by definition as follows

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

□

5.

The sub-gradient of l2-norm is

$$\partial ||x||_2 = \{ \begin{array}{cc} \frac{x}{||x||_2} & if \ x \neq 0 \\ \{g | \|g\|_2 \leq 1\} & if \ x = 0 \end{array} \tag{1}$$

6.

Again, the proposition can be proved by definition as follows

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

7.

We can first verify that the conjugate function of $f(x) = \frac{1}{2}x^T Ax$ is $f^*(y) = \frac{1}{2}y^T A^{-1} y$ as follows. By definition, $f^*(y) = sup\{\langle x, y \rangle - \frac{1}{2}x^T Ax\}$ with the following fact

$$\frac{\partial \langle x, y \rangle - \frac{1}{2}x^T Ax}{\partial x} = y - \frac{1}{2}(A + A^T)x = y - Ax \triangleq 0$$

which implies that it reaches supreme when $x = A^{-1}y$. Then, $f^*(y) = \frac{1}{2}(A^{-1}y)^T A(A^{-1}y) = \frac{1}{2}y^T A^{-1}y$.

Therefore, thanks to Fenchel inequality,

$$
\begin{aligned}
f(x) + f^*(y-b) &\geq \langle x, y-b \rangle \\
\frac{1}{2}x^T Ax + \frac{1}{2}(y-b)^T A^{-1}(y-b) &\geq \langle x, y \rangle - b^T x \\
\frac{1}{2}x^T Ax + b^T x + \frac{1}{2}(y-b)^T A^{-1}(y-b) &\geq \langle x, y \rangle
\end{aligned}
$$

Exercise 148

2.

We first note that

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

The critic point of $\parallel w \parallel + \frac{1}{2c} \parallel w - x \parallel^2$ satisfies $\frac{w}{\|w\|} + \frac{1}{c}(w - x) = 0$.

Thus,

$$w^* = P_c f(x) = (\parallel x \parallel - c)\frac{x}{\parallel x \parallel} = x - \frac{cx}{\parallel x \parallel}$$

$$E_c f(x) = \parallel x \parallel - \frac{c}{2}$$

3.

Assuming $f(x) = g(\lambda x + a)$, we have

$$
\begin{aligned}
&P_c f(x) \\
=&argmin_w\{f(w) + \frac{1}{2c} \parallel w - x \parallel^2\} \\
=&aigmin_w\{g(\lambda w + a) + \frac{1}{2c} \parallel w - x \parallel^2\} \\
=&\frac{1}{\lambda}[-a + argmin_t\{g(t) + \frac{1}{2c} \parallel \frac{t-a}{\lambda} - x \parallel^2\}] \\
=&\frac{1}{\lambda}[-a + argmin_t\{g(t) + \frac{1}{2c\lambda^2} \parallel t - (a + \lambda x) \parallel^2\}] \\
=&\frac{1}{\lambda}[-a + P_c(\lambda^2 g)(\lambda x + a)]
\end{aligned}
$$

4.

Proof. Similarly, assuming $f(x) = \lambda g(x/\lambda)$, we have the following

$$\begin{aligned}
&P_c f(x) \\
=&arg\min_w \{f(w) + \frac{1}{2c} \parallel w - x \parallel^2\} \\
=&arg\min_w \{\lambda g(w/\lambda) + \frac{1}{2c} \parallel w - x \parallel^2\} \\
=&\lambda arg\min_t \{\lambda g(t) + \frac{1}{2c} \parallel \lambda t - x \parallel^2\} \\
=&\lambda arg\min_t \{g(t)/\lambda + \frac{1}{2c} \parallel t - x/\lambda \parallel^2\} \\
=&\lambda P_c(g/\lambda)(x/\lambda)
\end{aligned}$$

□