Introduction to Machine Learning

Fall 2022

University of Science and Technology of China

Lecturer: Jie Wang | Homework 3
Posted: Oct. 25, 2022 | Due: Nov. 8, 2022

Notice, to get the full credits, please present your solutions step by step.

Exercise 1: Convex Sets

Let $C \subset \mathbb{R}^n$ be a nonempty convex set. Please show the following statements.

1. Please find the interior and relative interior of the following convex sets (you don't need to prove them).

   (a) $\{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1, x_3 = 0\} \subset \mathbb{R}^3$.

   (b) $\{\mathbf{A} \in S_{++}^n : \mathrm{Tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$.

   (c) $\{\mathbf{A} \in S_{++}^n : \mathrm{Tr}(\mathbf{A}) = 1\} \subset S^n$.

   (d) (Optional) $\{\mathbf{A} \in S_{++}^n : \mathrm{Tr}(\mathbf{A}) \leq 1\} \subset \mathbb{R}^{n\times n}$.

   (e) $\mathbf{conv}\,(\{x, x^2, x^3\}) \subset C[0,1]$ with L^∞ norm, i.e., $\|f\|_\infty = \max_{x\in[0,1]} |f(x)|$ for any $f \in C[0,1]$.

2. Some operations that preserve convexity.

   (a) Both $\mathbf{cl}\, C$ and $\mathbf{int}\, C$ are convex.

   (b) The set $\mathbf{relint}\, C$ is convex.

   (c) The intersection $\bigcap_{i\in I} C_i$ of any collection $\{C_i : i \in \mathcal{I}\}$ of convex sets is convex.

   (d) If C_1 and C_2 are convex sets in $\mathbb{R}^n$, then the set

   $$C_1 - C_2 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \mathbf{x}_1 \in C_1, \mathbf{x}_2 \in C_2\}$$

   is convex.

   (e) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{x} \in C\}$ is convex, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{a} \in \mathbb{R}^m$.

   (f) The set $\{\mathbf{y} \in \mathbb{R}^m : \mathbf{x} = \mathbf{B}\mathbf{y} + \mathbf{b}, \mathbf{x} \in C\}$ is convex, where $\mathbf{B} \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$.

Solution: ■

Exercise 2: Affine Sets

Please show the following statements about affine sets.

1. If $U \subset \mathbb{R}^n$ and $\mathbf{0} \in U$, then U is an affine set if and only if it is a subspace.
2. If $U \subset \mathbb{R}^n$ is an affine set, there is a unique subspace $V \subset \mathbb{R}^n$ such that $U = \mathbf{u} + V$ for any $\mathbf{u} \in U$.
3. Let $U = \mathbf{aff}(\{(1,0,0)^\top, (0,1,0)^\top, (0,0,1)^\top\})$. Please find two vectors $\mathbf{v}_1$ and $\mathbf{v}v_2$ such that we can represent any vectors $\mathbf{v} \in U$ in the form of $\mathbf{v} = (1,0,0)^\top + \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2$ uniquely, where α_1 and α_2 are two real numbers depending on $\mathbf{v}$. Furthermore, given a point $\mathbf{x}_0 \in U$, find two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ such that we can represent any vectors $\mathbf{w} \in U$ in the form of $\mathbf{w} = \mathbf{x}_0 + \alpha_1\mathbf{w}_1 + \alpha_2\mathbf{w}_2$ uniquely.

Solution: ■

Exercise 3: Convex Hull and Affine Hull (Optional)

Let A be a subset of $\mathbb{R}^n$.

1. (a) Please show that the convex hull of A is the smallest convex set containing A, i.e., all the convex sets containing A also contain **conv** A.

   (b) Please find the convex hull of the following sets.

   i. $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \cup \{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) \geq 2\} \subset \mathbb{R}^{n\times n}$.

   ii. $\{f \in C[0,1] : \|f\|_\infty = 1\} \cup \{f \in C[0,1] : \|f\|_\infty = 2\} \subset C[0,1]$.

2. (a) Please show that the affine hull of A is the smallest affine set containing A, i.e., all the affine sets containing A also contain **aff**A.

   (b) Please find the affine hull of the following sets.

   i. $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \subset \mathbb{R}^{n\times n}$.

   ii. $\{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) = 1\} \cup \{\mathbf{A} \in S^n_{++} : \mathrm{Tr}(\mathbf{A}) \geq 2\} \subset \mathbb{R}^{n\times n}$.

Solution: ■

Exercise 4: Relative Interior and Interior

Let $C \subset \mathbb{R}^n$ be a nonempty convex set.

1. Let $\mathbf{x}_0 \in C$. Please show the following statements.
   (a) The point $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $r > 0$ such that $\mathbf{x}_0 + r\mathbf{v} \in C$ for any $\mathbf{v} \in \textbf{aff}\, C - \mathbf{x}_0$ and $\|\mathbf{v}\|_2 \leq 1$.
   (b) Let $\{\mathbf{v}_i\}_{i=1}^m$ be a basis of $\textbf{aff}\, C - \mathbf{x}_0$. Then $\mathbf{x}_0 \in \textbf{relint } C$ if and only if there exists $r > 0$ such that $\mathbf{x}_0 + r\sum_i \alpha_i \mathbf{v}_i \in C$ for any $\{\alpha_i\}_{i=1}^m$ with $\sum_i \alpha_i^2 \leq 1$.
2. (a) We let $x_0 \in \textbf{int } C$, $x_1 \in \textbf{bd } C$ and $x_2 = \lambda(x_1 - x_0) + x_0$.
      i. Please show that if $\lambda > 1$, then $x_2 \notin C$.
      ii. Please show that if $\lambda \in (0, 1)$, then $x_2 \in \textbf{int } C$.
   (b) i. Please show that $x \in \textbf{relint } C$ if and only if for any $y \in C$, there exists $\gamma > 0$ such that $x + \gamma(x - y) \in C$.
      ii. Please show that if $x \in \textbf{relint } C$, $y \in \textbf{cl } C$, then $\lambda x + (1-\lambda)y \in \textbf{relint } C$ for $\lambda \in (0, 1]$.
3. (a) Please show the following statements.
      i. Suppose $\textbf{int } C$ is nonempty, then $\textbf{int } C = \textbf{int}\,(\textbf{cl } C)$ (in fact, the result still holds when $C = \emptyset$).
      ii. $\textbf{cl}\,(\textbf{relint } C) = \textbf{cl } C$.
      iii. $\textbf{relint}\,(\textbf{cl } C) = \textbf{relint } C$.

      [Hint: if C contains more than one point, then $\textbf{relint } C$ is nonempty. You may also use the results in Question 2.]
   (b) Using the results in Question 3(a), please prove the following statement.

      For a convex set $C \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in \textbf{bd } C$, we can find a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n \backslash \textbf{cl } C$ such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

Solution: ■

Exercise 5: Relative Boundary

The relative boundary of a set $S \subset \mathbb{R}^n$ is defined as $\textbf{relbd}\ S = \textbf{cl}\ S \setminus \textbf{relint}\ S$. Please show the following statements **or give counter-examples**.

1. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S \subset \textbf{bd}\ S$.
2. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S = \textbf{bd}\ S$.
3. For a set $S \subset \mathbb{R}^n$, $\textbf{relbd}\ S = \textbf{relbd}\ \textbf{cl}\ S$.
4. (Optional) For a convex set $C \subset \mathbb{R}^n$, $\textbf{relbd}\ C = \textbf{relbd}\ \textbf{cl}\ C$.
5. For a set $S \subset \mathbb{R}^n$ and $\mathbf{x}_0 \in \textbf{cl}\ S$, we can find a sequence $\{\mathbf{x}_k\} \subset \mathbb{R}^n \setminus \textbf{cl}\ S$ such that $\mathbf{x}_k \to \mathbf{x}_0$ as $k \to \infty$.

Solution: ■

Exercise 6: Minkowski Summation of Sets (Optional)

The Minkowski sum of two sets S_1 and S_2 is defined by

$$S_1 + S_2 = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in S_1, \mathbf{y} \in S_2\}.$$

1. Let $S_1 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq 1\}$ and $S_2 = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_\infty \leq 1\}$.
   (a) Please draw the set $S_1 + S_2$.
   (b) How do you tell if a point $\mathbf{x}$ is in the set $S_1 + S_2$?
2. Recall that $\mathbb{R}^n$ can be decomposed as $\mathbb{R}^n = S \oplus S^\perp$, i.e., $\mathbb{R}^n = S + S^\perp$ and $S \cap S^\perp = \emptyset$, where $S \subset \mathbb{R}^n$ is a subspace and $S^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \perp S\}$. Let $C \subset \mathbb{R}^n$ be a convex set. Define $\hat{C} = C + (\mathbf{aff}\, C - \mathbf{x}_0)^\perp$. Please show that:
   (a) $\dim(\mathbf{aff}\, \hat{C}) = n$;
   (b) $\mathbf{relint}\ C + (\mathbf{aff}\, C - \mathbf{x}_0)^\perp = \mathbf{relint}\ \hat{C}$;
   (c) $\mathbf{relbd}\ C + (\mathbf{aff}\, C - \mathbf{x}_0)^\perp = \mathbf{relbd}\ \hat{C}$.

Solution: ■

Exercise 7: Convex Sets and Linear Functions

Let $C \subset \mathbb{R}^n$ be a convex set and $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$ be a linear function on $\mathbb{R}^n$. The linear function is nontrivial if $\mathbf{a} \neq \mathbf{0}$. Suppose $\mathbf{x}_0 \in C$ and denote

$$\mathcal{B}_C(\mathbf{x}_0, r) = \mathcal{B}(\mathbf{x}_0, r) \cap \mathbf{aff}\, C.$$

Please show the following statements.

1. If $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in \mathcal{B}_C(\mathbf{x}_0, r)$, then $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in C$.
2. The linear function $l(\mathbf{x}) = \alpha, \forall\, \mathbf{x} \in \mathcal{B}_C(\mathbf{x}_0, r)$ for some constant α if and only if $\mathbf{a} \perp (\mathbf{aff}\, C - \mathbf{x}_0)$.
3. The linear function $l(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle$ is not constant if and only if $\Pi_{(\mathbf{aff}\, C - \mathbf{x}_0)}(\mathbf{a}) \neq \mathbf{0}$, where Π means the projection.
4. If $\mathbf{relbd}\ C \neq \emptyset$, then there exists a nontrivial linear function l, and a constant α such that $l(\mathbf{x}) \leq \alpha$ for $\forall\, \mathbf{x} \in C$.

Solution: ■

Exercise 8: Separation Theorems

Let $C_1, C_2, C \subset \mathbb{R}^n$ be convex sets. Please show the following statements.

1. If C_1 is compact, C_2 is closed and $C_1 \cap C_2 = \emptyset$, then C_1 and C_2 can be strongly separated.
2. (Optional) The sets C_1 and C_2 can be properly separated if and only if $\textbf{relint}\ C_1 \cap \textbf{relint}\ C_2 = \emptyset$.
3. If $\dim(\textbf{aff}\, C) = n$ and $\mathbf{x} \in \mathbb{R}^n \setminus C$, then $\mathbf{x}$ and C can be properly separated.

Solution: ■

Exercise 9: Farkas' Lemma

Let $\mathbf{A} = (\mathbf{a}_1, \cdots, \mathbf{a}_n) \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Consider a set $A = \{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$. Its conic hall $\textbf{cone}\, A$ is defined as

$$\textbf{cone}\, A = \{\sum_{i=1}^{n} \alpha_i \mathbf{a}_i : \alpha_i \geq 0, \mathbf{a}_i \in A\}.$$

1. Please show that $\textbf{cone}\, A$ is closed and convex.
2. If $\mathbf{b} \in \textbf{cone}\, A$, please show that there exists $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.
3. If $\mathbf{b} \notin \textbf{cone}\, A$, use separation theorems to show that there exists $\mathbf{y} \in \mathbb{R}^m$, such that $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.
4. Now you can prove Farkas' Lemma: for given $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$, one and only one of the two statements hold:
   - $\exists \mathbf{x} \in \mathbb{R}^n$, $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.
   - $\exists \mathbf{y} \in \mathbb{R}^m$, $\mathbf{A}^\top \mathbf{y} \geq \mathbf{0}$ and $\mathbf{b}^\top \mathbf{y} < 0$.

Solution: ■

Exercise 10: Projection to a Polytope

Hint: you may want to read [1, 2].

1. Let C be a nonempty closed convex subset of $\mathbb{R}^n$. Please show the following statements.

   (a) The projection operator on C, i.e., $\mathbf{P}_C$, is continuous and firmly nonexpansive. In other words, for any $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^n$, we have

   $$\|\mathbf{P}_C(\mathbf{w}_1) - \mathbf{P}_C(\mathbf{w}_2)\|_2^2 + \|(\mathrm{Id} - \mathbf{P}_C)(\mathbf{w}_1) - (\mathrm{Id} - \mathbf{P}_C)(\mathbf{w}_2)\|_2^2 \leq \|\mathbf{w}_1 - \mathbf{w}_2\|_2^2,$$

   where Id is the identity operator.

   (b) For a point $\mathbf{w} \in \mathbb{R}^n$, let $\mathbf{w}(t) = \mathbf{P}_C(\mathbf{w}) + t(\mathbf{w} - \mathbf{P}_C(\mathbf{w}))$. Then, the projection of the point $\mathbf{w}(t)$ is $\mathbf{P}_C(\mathbf{w})$ for all $t \geq 0$, i.e.,

   $$\mathbf{P}_C(\mathbf{w}(t)) = \mathbf{P}_C(\mathbf{w}), \forall t \geq 0.$$

2. Let $\mathbf{y}$ be an N-dimensional vector ,$\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ be N-dimension non-zero vectors and $\lambda \geq 0$ is a regularization parameter. Consider the following optimization problem:

   $$\min_{\theta} \left\{ \left\| \theta - \frac{\mathbf{y}}{\lambda} \right\|_2^2 : \left| \mathbf{x}_i^T \theta \right| \leq 1, i = 1, 2, \ldots, p \right\}. \tag{1}$$

   For notational convenience, we denote the optimal solution of (1) by $\theta^*(\lambda)$.

   (a) We let the feasible set of (1) be F. Please give an interpretation of the geometry of F (you don't need to prove it). Then give a close form of the optimal solution $\theta^*(\lambda)$ in the form of projection.

   (b) Let $\lambda, \lambda_0 > 0$ be two regularization parameters. Please show that

   $$\theta^*(\lambda) \in B\left(\theta^*(\lambda_0), \left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right).$$

   (c) Let $\lambda, \lambda_0 > 0$ be two regularization parameters. Please show that

   $$\theta^*(\lambda) \in B\left(\theta^*(\lambda_0) + \frac{1}{2}\left(\frac{1}{\lambda} - \frac{1}{\lambda_0}\right)\mathbf{y}, \frac{1}{2}\left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right).$$

   (You may use the result in Question 1(a).)

   (d) Suppose that $\mathbf{P}_F(\frac{\mathbf{y}}{\lambda_0}) \neq \theta^*(\lambda_0)$. For any $\lambda \in (0, \lambda_0]$, let us define

   $$\mathbf{v}_1(\lambda_0) = \frac{\mathbf{y}}{\lambda_0} - \theta^*(\lambda_0),$$
   $$\mathbf{v}_2(\lambda, \lambda_0) = \frac{\mathbf{y}}{\lambda} - \theta^*(\lambda_0),$$
   $$\mathbf{v}_2^{\perp}(\lambda, \lambda_0) = \mathbf{v}_2(\lambda, \lambda_0) - \frac{\langle \mathbf{v}_1(\lambda_0), \mathbf{v}_2(\lambda, \lambda_0)\rangle}{\|\mathbf{v}_1(\lambda_0)\|_2^2}\mathbf{v}_1(\lambda_0).$$

Then, the dual optimal solution $\theta^*(\lambda)$ can be estimated as follows:

$$\theta^*(\lambda) \in B\left(\theta^*\left(\lambda_0\right), \left\|\mathbf{v}_2^{\perp}\left(\lambda, \lambda_0\right)\right\|_2\right) \subseteq B\left(\theta^*\left(\lambda_0\right), \left|\frac{1}{\lambda} - \frac{1}{\lambda_0}\right| \|\mathbf{y}\|_2\right)$$

(You may use the result in Question 1(b).)

Solution: ■

References

[1] T. Hastie, R. Tibshirani, and M. Wainwright. Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143:143, 2015.

[2] J. Wang, P. Wonka, and J. Ye. Lasso screening rules via dual polytope projection. *Journal of Machine Learning Research*, 16(32):1063–1101, 2015.