SI251 - Convex Optimization homework 1

Deadline: 2022-10-09 23:59:59

1. You can use Word, Latex or handwriting to complete this assignment. If you want to submit a handwritten version, scan it clearly.
2. The **report** has to be submitted as a PDF file to Gradescope, other formats are not accepted.
3. The submitted file name is **student_id+your_student_name.pdf**.
4. Late policy: You have 4 free late days for the quarter and may use up to 2 late days per assignment with no penalty. Once you have exhausted your free late days, we will deduct a late penalty of 25% per additional late day. Note: The timeout period is recorded in days, even if you delay for 1 minute, it will still be counted as a 1 late day.
5. You are required to follow ShanghaiTech's academic honesty policies. You are not allowed to copy materials from other students or from online or published resources. Violating academic honesty can result in serious sanctions.

Any plagiarism will get Zero point.

1 Convex sets

1. (15 pts) Please prove that the following sets are convex:

   1) (**Ellipsoids**) $\left\{x|\sqrt{(x-x_c)^TP(x-x_c)}\le r\right\}$ $\quad (x_c\in\mathbb{R}^n, r\in\mathbb{R}, P\succeq 0)$;

   2) (**Symmetric positive semidefinite matrices**) $S_+^{n\times n}=\left\{P\in S^{n\times n}|P\succeq 0\right\}$;

   3) The set of points closer to a given point than a given set, i.e.,

   $$\left\{x \mid \|x-x_0\|_2\le\|x-y\|_2 \text{ for all } y\in S\right\},$$

   where $S\in R^n$.

2. (15 pts) Let C be a nonempty subset of $\mathbb{R}^n$, and let λ_1 and λ_2 be positive scalars. Show that if C is convex, then $(\lambda_1+\lambda_2)C=\lambda_1C+\lambda_2C$. Show by example that this need not be true when C is not convex.

   Hint: A vector x in $\lambda_1C+\lambda_2C$ is of the form $x=\lambda_1x_1+\lambda_2x_2$, where $x_1,x_2\in C$.

2 Convex functions

3. (12 pts) Let $C\subset\mathbb{R}^n$ be convex and $f:C\to R^\star$. Show that the following statements are equivalent:

   (a) epi(f) is convex.

   (b) For all points $x_i\in C$ and $\{\lambda_i|\lambda_i\ge 0,\sum_{i=1}^n\lambda_i=1, i=1,2,\cdots,n\}$, we have

   $$f\Big(\sum_{i=1}^n\lambda_ix_i\Big)\le\sum_{i=1}^n\lambda_if(x_i).$$

   (c) For $\forall x,y\in C$ and $\lambda\in[0,1]$,

   $$f\Big((1-\lambda)x+\lambda y\Big)\le(1-\lambda)f(x)+\lambda f(y).$$

4. (18 pts) Please determine whether the following functions are convex, concave or none of those, and give a detailed explanation for your choice.

   1)
   $$f_1(x_1,x_2,\cdots,x_n)=\begin{cases}-(x_1x_2\cdots x_n)^{\frac{1}{n}}, & \text{if } x_1,\cdots,x_n>0\\ \infty & \text{otherwise;}\end{cases}$$

   2) $f_2(x_1,x_2)=x_1^\alpha x_2^{1-\alpha}$, where $0\le\alpha\le 1$, on $\mathbb{R}^2_{++}$;

   3) $f_3(x,u,v)=-\log(uv-x^Tx)$ on $\mathbf{dom}f=\{(x,u,v)|uv>x^Tx,\ \ u,v>0\}$.

3 Convex optimization problems

5. (15 pts) **Robust quadratic programming**. In the lecture, we have learned about robust linear programming as an application of second-order cone programming. Now we will consider a similar robust variation of the convex quadratic program

$$\begin{array}{ll} \text{minimize} & (1/2)x^T Px + q^T x + r \\ \text{subject to} & Ax \preceq b. \end{array}$$

For simplicity, we assume that only the matrix P is subject to errors, and the other parameters (q, r, A, b) are exactly known. The robust quadratic program is defined as

$$\begin{array}{ll} \text{minimize} & \sup_{P\in\mathcal{E}} \left((1/2)x^T Px + q^T x + r\right) \\ \text{subject to} & Ax \prec b \end{array}$$

where $\mathcal{E}$ is the set of possible matrices P.

For each of the following sets $\mathcal{E}$, express the robust QP as a convex problem in a standard form (e.g., QP, QCQP, SOCP, SDP).

(a) A finite set of matrices: $\mathcal{E} = \{P_1, ..., P_K\}$, where $P_i \in S_+^n, i = 1, ..., K$.

(b) A set specified by a nominal value $P_0 \in S_+^n$ plus a bound on the eigenvalues of the deviation $P - P_0$:

$$\mathcal{E} = \{P \in \mathbf{S}^n \mid -\gamma I \preceq P - P_0 \preceq \gamma I\}$$

where $\gamma \in \mathbf{R}$ and $P_0 \in \mathbf{S}_+^n$.

(c) An ellipsoid of matrices:

$$\mathcal{E} = \left\{ P_0 + \sum_{i=1}^{K} P_i u_i \mid \|u\|_2 \leq 1 \right\}.$$

You can assume $P_i \in \mathbf{S}_+^n, i = 0, \dots, K$.

6. (25 pts) **Group-SVM Estimator**. SVM (Support Vector Machine) is a well-known classification model in the field of machine learning. The key idea of it is to find a hyperplane that can maximize the margin between points from different classes (a.k.a. support vectors). Mathematically, SVM is to solve the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w}, b, \xi_i} \ & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i \\ \text{s.t.} \ & y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i + b\right) \geqslant 1 - \xi_i \\ & \xi_i \geqslant 0, i = 1, 2, \dots, m. \end{aligned}$$

From another perspective, $\frac{1}{2}\|\boldsymbol{w}\|^2$ can also be viewed as a regularization term. Perhaps, you have come up with some ideas to replace ℓ_2-norm regularization term with some other regularization term, e.g. ℓ_1-norm, ℓ_∞-norm. That is what we want to discuss in this homework.

Now, a variant of SVM called Group SVM will be introduced to you. Group SVM considers that parameters of SVM should be divided into groups, and the maximum of parameters in the same group should not be too large. Here is the formula of Group SVM.

$$\begin{aligned} \min_{\boldsymbol{w}, b, \xi_i} \ & \lambda \sum_{g=1}^{G} \|\boldsymbol{w_g}\|_\infty + \sum_{i=1}^{m} \xi_i \\ \text{s.t.} \ & y_i \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}_i + b \right) \geqslant 1 - \xi_i \\ & \xi_i \geqslant 0, i = 1, 2, \ldots, m. \end{aligned}$$

where $\boldsymbol{w} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_G)$. (Note: This does not mean $\boldsymbol{w} \in \mathbb{R}^G$)

(a) Please reformulate the problem as a convex problem.

(b) Please derive the dual form of this problem.

(c) Use CVX to solve the problem that you have derived and plot the result of it (need to upload your code as well). Hint: you can generate the dataset via using two-dimensional Gaussian distributions which have different means.

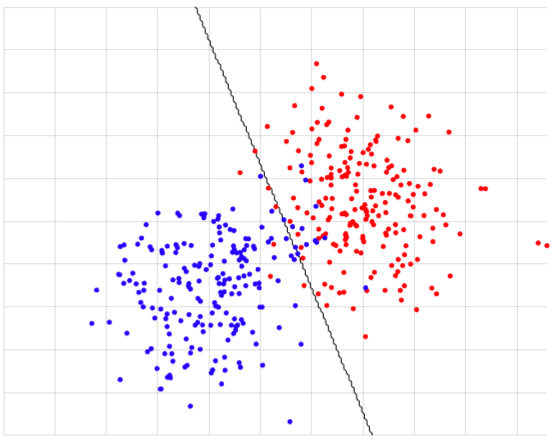

Figure 1: Example

(d) As is known to us, kernel method can be applied to SVM for the extension in nonlinear scenarios. Furthermore, is kernel method available to Group SVM as well? if your answer is yes, please derive the kernelised Group SVM. Otherwise, please illustrate your reason why kernel method cannot be applied to Group SVM.