

CSE 203B WI26 Homework 2 Suggested Solution

Assignments

2.1

```
import numpy as np
import cvxpy as cp
# ---- Data (m=6, n=2) ----
C = np.array([[ -2,  0],
               [ -2,  1],
               [ -2, -1],
               [ -3,  0],
               [ -3,  1],
               [ -3, -1]], dtype=float)

D = np.array([[ 2,  0],
               [ 2,  1],
               [ 2, -1],
               [ 3,  0],
               [ 3,  1],
               [ 3, -1]], dtype=float)

# ---- Variables ----
a = cp.Variable(2)
b = cp.Variable()
u = cp.Variable()
# ---- Constraints ----
constraints = []
for x in C:
    constraints.append(a @ x <= b - u)
for y in D:
    constraints.append(a @ y >= b + u)

constraints.append(cp.norm(a, 2) <= 1)

# ---- Solve ----
problem = cp.Problem(cp.Maximize(u), constraints)
problem.solve(solver=cp.SCS, verbose=False)

# ---- Results ----
print("Status:", problem.status)
print("Optimal margin u:", u.value)
print("Optimal a:", a.value)
print("Optimal b:", b.value)

print("||a||:", np.linalg.norm(a.value))
```

Status: optimal
Optimal margin u: 1.9999937084823782
Optimal a: [9.99995429e-01 9.02212861e-16]
Optimal b: -3.5430041559162176e-15
||a||: 0.9999954286714153

Figure 1: a

```

a = np.array([1.0, 0.0])
b = 0.0
u = 2.0
# ----- Plot points -----
plt.figure(figsize=(7, 4.5))
plt.scatter(C[:, 0], C[:, 1], marker='o', label='Set C')
plt.scatter(D[:, 0], D[:, 1], marker='s', label='Set D')
# ----- Plot separating hyperplane and margins -----
# Separating hyperplane: a^T x = b
# With a=(1,0), this is x1 = b (a vertical line)
plt.axvline(x=b, linestyle='--', label='Separating hyperplane: x1 = b')
# Margins: a^T x = b +/- u -> x1 = b +/- u
plt.axvline(x=b + u, linestyle=':', label='Margins: x1 = b + u')
plt.axvline(x=b - u, linestyle=':')
# ----- Formatting -----
plt.xlabel("x1")
plt.ylabel("x2")
plt.title("2-D SVM Example: Points and Optimal Separating Hyperplane")
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()

```

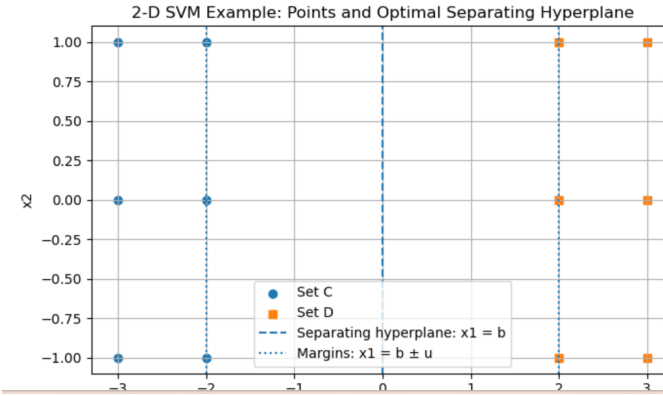


Figure 2: a

2.2:

Proof. Let $x_1, x_2 \in S$ and let $\lambda \in (0, 1)$. By definition of S , we have

$$\|Cx_i + d\| \leq e^T x_i + f, \quad i = 1, 2.$$

We are trying to prove that

$$x_\lambda := \lambda x_1 + (1 - \lambda)x_2 \in S,$$

that is,

$$\|Cx_\lambda + d\| \leq e^T x_\lambda + f.$$

Observe that

$$Cx_\lambda + d = \lambda(Cx_1 + d) + (1 - \lambda)(Cx_2 + d).$$

Using the triangle inequality and positive homogeneity of the norm, we obtain

$$\begin{aligned}
\|Cx_\lambda + d\| &= \|\lambda(Cx_1 + d) + (1 - \lambda)(Cx_2 + d)\| \\
&\leq \lambda\|Cx_1 + d\| + (1 - \lambda)\|Cx_2 + d\| \quad (\text{Triangular Inequality}) \\
&\leq \lambda(e^T x_1 + f) + (1 - \lambda)(e^T x_2 + f) \quad (\text{Dual Cone Definition}) \\
&= e^T (\lambda x_1 + (1 - \lambda)x_2) + f.
\end{aligned}$$

Therefore,

$$\|Cx_\lambda + d\| \leq e^T x_\lambda + f,$$

which implies $x_\lambda \in S$. Hence, S is convex. $\square$

2.3

Theorem. If f_1, f_2 are convex on a convex set D , then $f_1 + f_2$ is convex on D . Moreover, if $\{h_i\}_{i=1}^m$ are convex on D , then $h(x) = \max_{1 \leq i \leq m} h_i(x)$ is convex on D .

Claim. The function

$$f(x, y, z) = \frac{(x - z)^2}{y + 1} + \max\left(1 + |x| - y, \frac{1}{\sqrt{z}}, 0\right)$$

is convex on

$$D := \{(x, y, z) \in \mathbb{R}^3 : y > -1, z > 0\}.$$

Proof. Write $f = g + h$, where

$$g(x, y, z) = \frac{(x - z)^2}{y + 1}, \quad h(x, y, z) = \max(h_1(x, y, z), h_2(x, y, z), h_3(x, y, z)),$$

and

$$h_1(x, y, z) = 1 + |x| - y, \quad h_2(x, y, z) = \frac{1}{\sqrt{z}} = z^{-1/2}, \quad h_3(x, y, z) = 0.$$

We prove g and h are convex on D ; then $f = g + h$ is convex on D .

(1) Convexity of g via the Hessian. On D we have $v := y + 1 > 0$. Let

$$u = x - z, \quad v = y + 1,$$

so that $g(x, y, z) = u^2/v$. A direct computation gives the Hessian of g (with respect to (x, y, z)):

$$\nabla^2 g(x, y, z) = \begin{bmatrix} \frac{2}{v} & -\frac{2u}{v^2} & -\frac{2}{v} \\ -\frac{2u}{v^2} & \frac{2u^2}{v^3} & \frac{2u}{v^2} \\ -\frac{2}{v} & \frac{2u}{v^2} & \frac{2}{v} \end{bmatrix}.$$

Let $a = u/v$. Then $\nabla^2 g$ factors as

$$\nabla^2 g(x, y, z) = \frac{2}{v} \begin{bmatrix} 1 \\ -a \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -a & -1 \end{bmatrix} = \frac{2}{v} ww^\top, \quad w = \begin{bmatrix} 1 \\ -u/v \\ -1 \end{bmatrix}.$$

Since $v > 0$, we have $\frac{2}{v} > 0$, and $ww^\top \succeq 0$ as an outer product. Hence $\nabla^2 g(x, y, z) \succeq 0$ for all $(x, y, z) \in D$. By the second-order characterization of convexity for twice differentiable functions, g is convex on D .

(2) Convexity of the max term h . We show each h_i is convex on D .

(i) *Convexity of $h_1(x, y, z) = 1 + |x| - y$.* The map $x \mapsto |x|$ is convex on $\mathbb{R}$, and $(x, y, z) \mapsto 1 - y$ is affine (hence convex). Therefore h_1 , being the sum of a convex function and an affine function, is convex on D .

(ii) *Convexity of $h_2(x, y, z) = z^{-1/2}$.* Consider $\varphi(z) = z^{-1/2}$ on $(0, \infty)$. We compute

$$\varphi''(z) = \frac{3}{4}z^{-5/2} > 0 \quad (z > 0),$$

so φ is convex on $(0, \infty)$. Since $h_2(x, y, z) = \varphi(z)$ depends only on z , it follows that h_2 is convex on D .

(iii) *Convexity of $h_3(x, y, z) = 0$.* The constant function is affine and hence convex on D .

(iv) *Max of convex functions is convex.* Let $x, x' \in D$ and $\theta \in [0, 1]$. Since each h_i is convex,

$$h_i(\theta x + (1 - \theta)x') \leq \theta h_i(x) + (1 - \theta)h_i(x') \leq \theta \max_j h_j(x) + (1 - \theta) \max_j h_j(x').$$

Taking the maximum over $i \in \{1, 2, 3\}$ on the left-hand side yields

$$h(\theta x + (1 - \theta)x') = \max_i h_i(\theta x + (1 - \theta)x') \leq \theta h(x) + (1 - \theta)h(x').$$

Thus h is convex on D .

(3) Conclusion. We have shown g is convex on D and h is convex on D . Therefore, their sum $f = g + h$ is convex on D . $\square$

2.4 By definition that For any cone $K \subseteq \mathbb{R}^m$, its *dual cone* is defined as

$$K^* := \{z \in \mathbb{R}^m \mid \langle z, u \rangle \geq 0 \quad \forall u \in K\},$$

where $\langle \cdot, \cdot \rangle$ denotes the (Euclidean) inner product. The standard inner product on $\mathbb{R}^n \times \mathbb{R}$ is

$$\langle (y, s), (x, t) \rangle = y^\top x + st.$$

Thus

$$K_p^* = \{(y, s) \in \mathbb{R}^n \times \mathbb{R} \mid y^\top x + st \geq 0 \quad \forall (x, t) \in K_p\}.$$

We want

$$y^\top x + st \geq 0 \quad \forall (x, t) \text{ such that } \|x\|_p \leq t.$$

The definition of the cones implies $t \geq 0$. Among all x with $\|x\|_p \leq t$, the worst case (i.e., the most negative value of the left-hand side) occurs when $y^\top x$ is minimized. Therefore, we require

$$\min_{\|x\|_p \leq t} (y^\top x + st) \geq 0.$$

Since st is constant with respect to x , this is equivalent to

$$\min_{\|x\|_p \leq t} (y^\top x) + st \geq 0.$$

Moreover, because the set $\{x : \|x\|_p \leq t\}$ is centrally symmetric,

$$\min_{\|x\|_p \leq t} y^\top x = - \max_{\|x\|_p \leq t} y^\top x.$$

If $t = 0$, then the feasible set is $\{0\}$. Hence,

$$\max_{\|x\|_p \leq 0} y^\top x = y^\top 0 = 0 = 0 \cdot \|y\|_q.$$

Now assume $t > 0$. Consider the change of variables $x = tu$. Then

$$\|x\|_p \leq t \iff \|tu\|_p \leq t \iff t\|u\|_p \leq t \iff \|u\|_p \leq 1,$$

where we used the positive homogeneity of norms, namely $\|tu\|_p = t\|u\|_p$ for $t \geq 0$. Moreover,

$$y^\top x = y^\top (tu) = t y^\top u.$$

Therefore,

$$\max_{\|x\|_p \leq t} y^\top x = \max_{\|u\|_p \leq 1} t y^\top u = t \max_{\|u\|_p \leq 1} y^\top u.$$

By Lemma 1 (dual norm characterization), we have

$$\max_{\|x\|_p \leq t} y^\top x = t \|y\|_q,$$

where q satisfies $\frac{1}{p} + \frac{1}{q} = 1$.

Substituting back, we obtain

$$(-t\|y\|_q) + st \geq 0 \quad \forall t \geq 0.$$

This is equivalent to

$$t(s - \|y\|_q) \geq 0 \quad \forall t \geq 0,$$

which holds if and only if

$$s \geq \|y\|_q.$$

Therefore,

$$K_p^* = \{(y, s) \in \mathbb{R}^n \times \mathbb{R} \mid \|y\|_q \leq s\}, \quad \text{with } \frac{1}{p} + \frac{1}{q} = 1.$$

Lemma 1 (Dual norm characterization). Let $p \in [1, \infty]$ and let q be its dual exponent, i.e., $\frac{1}{p} + \frac{1}{q} = 1$. Then for any $y \in \mathbb{R}^n$,

$$\max_{\|u\|_p \leq 1} y^\top u = \|y\|_q.$$

Proof. By Hölder's inequality, for any $u \in \mathbb{R}^n$ satisfying $\|u\|_p \leq 1$, we have

$$|y^\top u| \leq \|y\|_q \|u\|_p \leq \|y\|_q.$$

Since $y^\top u \leq |y^\top u|$, it follows that

$$y^\top u \leq \|y\|_q \quad \text{for all } u \text{ with } \|u\|_p \leq 1.$$

Taking the max over all such u yields

$$\max_{\|u\|_p \leq 1} y^\top u \leq \|y\|_q.$$

To show that this bound is tight, we exhibit a vector u^* such that $\|u^*\|_p = 1$ and $y^\top u^* = \|y\|_q$. For $1 < p < \infty$, define

$$u_i^* = \frac{\text{sign}(y_i) |y_i|^{q-1}}{\|y\|_q^{q-1}}, \quad i = 1, \dots, n.$$

A direct calculation shows that $\|u^*\|_p = 1$, and moreover

$$y^\top u^* = \sum_{i=1}^n |y_i|^q \|y\|_q^{-(q-1)} = \|y\|_q.$$

(The cases $p = 1$ and $p = \infty$ follow analogously by choosing u^* supported on the index of maximal magnitude of y .)

Therefore,

$$\max_{\|u\|_p \leq 1} y^\top u = \|y\|_q.$$

□