MATH 173B, HW5

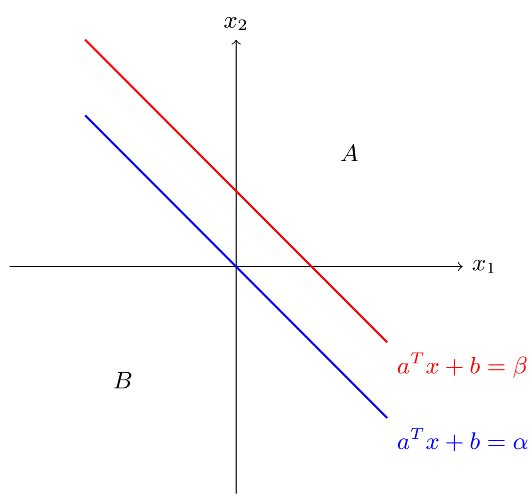
Victor Pekkari — epekkari@ucsd.edu

March 1, 2025

Problem 1

(a)

The A-region is above (including) the red line, and the B-region is beneath (including) the blue line.



(b)

We assume the point $z \in B$ and $z + \lambda a \in A$.

$$a^T z + b = \alpha \tag{1}$$

$$a^T(z + \lambda a) + b = \beta \tag{2}$$

Expanding (2) $\implies$

$$a^T z + \lambda a^T a + b = \beta$$

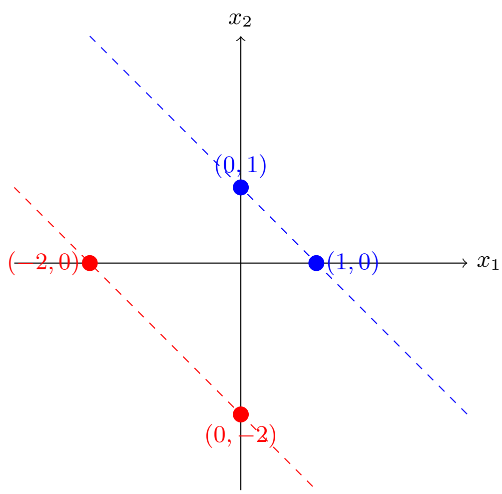
$$\alpha + \lambda \|a\|^2 = \beta \qquad (\alpha = a^T z + b)$$

$$\lambda = \frac{\beta - \alpha}{\|a\|^2}$$

The distance between the two hyperplanes is the length of the array we had to add to point $z \in B$ to get to hyperplane A.

$$w = \|\lambda a\| = \left\| \frac{\beta - \alpha}{\|a\|^2} a \right\| = \frac{\beta - \alpha}{\|a\|}$$

Problem 2



$$A = \{(1,0),(0,1)\}, \quad B = \{(-2,0),(0,-2)\}$$

(a)

Primal SVM optimization problem:

$$\min_{w,b} \frac{\|w\|^2}{\|\beta - \alpha\|^2} \quad s.t. \begin{cases} \beta - w^T x - b \leq 0, & \forall x \in \{(0,1),(1,0)\} \\ w^T x + b - \alpha \leq 0, & \forall x \in \{(-2,0),(0,-2)\} \end{cases}$$

Dual SVM optimization problem:

$$\mathcal{L}(w,b,\lambda,\mu) = \frac{\|w\|^2}{\|\beta - \alpha\|^2} + \lambda_1 \left(\beta - \left(w^T(0,1) + b\right)\right) + \lambda_2 \left(\beta - \left(w^T(1,0) + b\right)\right)$$
$$+\mu_1 \left(w^T(-2,0) + b - \alpha\right) + \mu_2 \left(w^T(0,-2) + b - \alpha\right)$$
$$\mathcal{L}(w,b,\lambda,\mu) = \frac{\|w\|^2}{\|\beta - \alpha\|^2} + \sum_{x \in A} \lambda_i(\beta - w^T x - b) + \sum_{x \in B} \mu_i(w^T + b - \alpha)$$

$$\mathcal{L}(w,b,\lambda,\mu) = \frac{\|w\|^2}{\|\beta-\alpha\|^2} + \langle a, \sum_{x\in B}\mu_x \cdot x - \sum_{x\in A}\lambda_x \cdot x\rangle + \langle b, \sum_{x\in B}\mu_x - \sum_{x\in A}\lambda_x\rangle + \sum_{x\in B}\mu_x - \sum_{x\in A}\lambda_x$$

$$F(\lambda,\mu) = \min_{w,b}\mathcal{L}(w,b,\lambda,\mu) = \mathcal{L}(w^*,b^*,\lambda,\mu)$$

Find $F(\lambda,\mu) = \min_{w,b}\mathcal{L}(w,b,\lambda,\mu)$

Dual optimization problem: $\max_{\lambda,\mu} F(\lambda,\mu)$

$$\max_{\lambda,\mu}\frac{\|w^*\|^2}{\|\beta-\alpha\|^2} + \langle w^{*T}, \sum_{x\in A}\mu_x \cdot x - \sum_{x\in B}\lambda_x \cdot x\rangle + \langle b^*, \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x\rangle + \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x$$

where w^*, b^* are the minimizers for $\mathcal{L}$
Subject to $\lambda, \mu \geq 0$ and $\sum_i \lambda_i = \sum_j \mu_j$

(b)

We have the primal:

$$\mathcal{L}(w,b,\lambda,\mu) = \frac{\|w\|^2}{\|\beta-\alpha\|^2} + \langle a, \sum_{x\in A}\mu_x \cdot x - \sum_{x\in B}\lambda_x \cdot x\rangle + \langle b, \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x\rangle + \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x$$

Since the primal is convex with respect to w, b, taking its gradient with respect to w, b will yield the minimizer.

$$\nabla_w \mathcal{L} = \frac{2w}{\|\beta-\alpha\|^2} + \sum_{x\in A}\mu_x \cdot x - \sum_{x\in B}\lambda_x \cdot x$$

$$\nabla_b \mathcal{L} = \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x$$

Thus, we obtain the system:

$$\begin{cases} \frac{2w}{\|\beta-\alpha\|^2} + \sum_{x\in A}\mu_x \cdot x - \sum_{x\in B}\lambda_x \cdot x & = 0 \\ \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x & = 0 \end{cases}$$

Rearranging,

$$\frac{2w}{\|\beta-\alpha\|^2} + \sum_{x\in A}\mu_x \cdot x - \sum_{x\in B}\lambda_x \cdot x + \sum_{x\in A}\mu_x - \sum_{x\in B}\lambda_x = 0$$

$$\frac{2w}{\|\beta-\alpha\|^2} + \sum_{x\in A}(1+x)\cdot\mu_x - \sum_{x\in B}(1+x)\cdot\lambda_x = 0$$

From this, solving for w leads to:

$$w = \frac{\|\beta - \alpha\|^2}{2}\left(\sum_{x\in B}(1+x)\cdot\lambda_x - \sum_{x\in A}(1+x)\cdot\mu_x\right)$$

Finally, the optimal separating hyperplane equation is:

$$w^T x + b = 0 \implies b = \sum_{x\in B}\lambda_x - \sum_{x\in A}\mu_x$$

Answer: This means that the optimal hyperplane to use as a separator is (put in the points if you want to get the numeric value for w and b):

$$\frac{\|\beta - \alpha\|^2}{2}\left(\sum_{x\in B}(1+x)\cdot\lambda_x - \sum_{x\in A}(1+x)\cdot\mu_x\right)^T x + \sum_{x\in B}\lambda_x - \sum_{x\in A}\mu_x = 0.$$

(c)

Answer: There is no unique pair of closest points between the two convex hulls of A and B.

Dual problem: Since Slater's condition holds for this problem (like we proved in class), the optimal value of the original probelm is the same as the optimal value of the dual problem. This is like said in clase the other day; the closest two point between set A and B, where the points are computed through a weighted average. One of the two points is computed through a weighted average of the points in A, and the other one is computed through a weighted average of the points in B.

$$\min_{\lambda,\nu} \|\sum_{x\in B}\lambda_x\cdot x - \sum_{x\in A}\nu_x\cdot x\|^2 \quad s.t. \begin{cases}\lambda & \geq 0\\ \nu & \geq 0\\ \sum_{x\in B}\lambda_x\cdot x & = \sum_{x\in A}\nu_x\cdot x = 1\end{cases}$$

If we insert our points into the equation above we get that we **can't** find two unique closest points between the convex hulls of A and B. We get all pairwise-closest points on between the red and blue line in the graph above.

Problem 3

(a)

Define the Lagrangian function as:

$$\mathcal{L}(x_1, x_2, \lambda, \nu_1, \nu_2) = f(x) + \sum_{i}^{m}\lambda_i\cdot g_i(x) + \sum_{i}^{P}\nu_i\cdot h_i(x)$$

where:

$$f(x) = 2x_2 + (x_1^2 + x_2^2)$$

$$h(x) = x_1 + x_2 - 1$$

$$g_1(x) = -x_1, \quad g_2(x) = -x_2$$

All functions above are convex hence setting the gradients equal to zero will give us the minimizer:

1. Stationarity:

$$\nabla_{x_1}\mathcal{L} = 2x_1 + \nu - \lambda_1 = 0 \implies x_1^* = \frac{1}{2}(\lambda_1 - \nu)$$

$$\nabla_{x_2}\mathcal{L} = 2x_2 + 2 + \nu - \lambda_2 = 0 \implies x_2^* = \frac{1}{2}(\lambda_2 - 2 - \nu)$$

2. Primal Feasibility:

$$x_1 + x_2 = 1, \quad x_1 \geq 0, \quad x_2 \geq 0$$

3. Dual Feasibility:

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0$$

4. Complementary Slackness:

$$\lambda_1 x_1 = 0, \quad \lambda_2 x_2 = 0$$

(b)

Using complementary slackness tells us that if $x_i > 0 \implies \lambda_i = 0$:
Assume $x_1 > 0$ and $x_2 > 0$, then $\lambda_1 = 0$ and $\lambda_2 = 0$. Using equations from points (1):

$$2x_1 + \nu = 0$$

$$2 + 2x_2 + \nu = 0$$

From (3), $x_1 + x_2 = 1$. Solving for x_1, x_2, ν:

$$\nu = -2x_1, \quad \nu = -2 - 2x_2$$

$$-2x_1 = -2 - 2x_2 \Rightarrow 2x_1 = 2 + 2x_2 \Rightarrow x_1 - x_2 = 1$$

$$\begin{cases} x_1 + x_2 = 1 \\ x_1 - x_2 = 1 \end{cases} \Rightarrow x_1 = 1, \quad x_2 = 0$$

Then, $\nu = -2(1) = -2$

(c)

The found point $(x_1, x_2) = (1, 0)$ satisfies the KKT conditions. This means it does solve the original optimization problem

Problem 4

(a)

$$\mathcal{L}(x, \lambda, \mu) = c^T x + \mu^T (Ax - b) - \lambda^T x,$$

The KKT conditions are:

1. **Stationarity:** The gradient of the Lagrangian must be zero:

$$\nabla_x \mathcal{L} = c + A^T \lambda - \mu = 0.$$

2. **Primal feasibility:** The constraints must hold:

$$Ax - b = 0, \quad x \geq 0.$$

$$-x_i \leq 0 \quad \forall i \in \{1, ..., n\}$$

3. **Dual feasibility:** The multipliers must be non-negative:

$$\lambda \geq 0.$$

4. **Complementary slackness:** Each component must satisfy:

$$-\lambda_i x_i = 0, \quad \forall i \in \{1, \ldots, n\}$$

(b)

Answer: Yes, if a primal feasibe point x^* and a dual feasible (λ^*, ν^*) satisifes the KKT-conditions, x^* will be optimal for the linear optimization problem.

Problem 5

Inital note: The Lagrangian is convex, and Dual feasibility holds since there are no inequality constraints in the problems

$$\mathcal{L}(x, y, \lambda, \nu) = \|x\|^2 + \mu(a^T x - b) \tag{1}$$

The gradient with respect to x must be 0:

$$\nabla_x \mathcal{L} = 2x + \mu a \implies x = -\frac{1}{2}\mu a$$

The constraint must be satisfied for primal feasibilty to hold:

$$a^T \left[-\frac{1}{2}\mu a\right] = b$$

Solving for λ gives:

$$\lambda = -\frac{2b}{\|a\|^2}$$

Computing the minimizer x^*

$$x^* = -\frac{1}{2}\left[-\frac{2b}{\|a\|^2}\right] a \implies x^* = \frac{b}{\|a\|^2} \cdot a$$

Note: The KKT-conditions now hold for the optimization problem

Interpretation: x^* is the closest point in the hyperplane $a^T x = b$ to the origin.