Convex Optimization
Assignment 4

Laura Elisa Celis

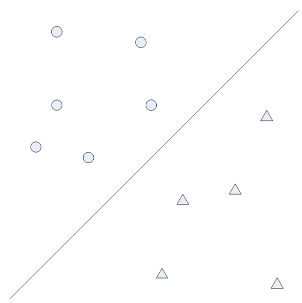
March 14, 2008

8.23 • Linear Discrimination.

In linear discrimination we are given a set of points $x_1, \ldots, x_n, y_1, \ldots, y_m$, and wish to find an affine function $f(x) = a^T x - b$ that classifies the points so that

$$f(x_i) > 0 \text{ and } f(y_j) < 0 \text{ for all } i \in [n], j \in [m].$$

Geometrically, this is equivalent to finding a hyperplane that separates the two sets of points as shown below.

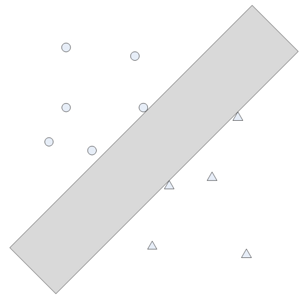


• Robust Linear Discrimination

Given a feasible function f as above, the function αf for some $\alpha \in \mathbb{R}_+$ is also feasible. So we can scale f as desired, and the above is equivalent to finding f such that $f(x_i) \geq t$ and $f(y_j) \leq -t$ for any $t \in \mathbb{R}_+$. This leads to the robust linear discrimination problem, which asks us to maximize some such t given $\|a\|_2 \leq 1$. Written formally:

$$\begin{array}{lll} \text{maximize:} & t & \\ \text{subject to:} & a^T x_i - b \geq t, & i \in [n] \\ & a^T y_j - b \leq -t, & j \in [m] \\ & \|a\|_2 \leq 1 & \end{array}$$

Notice that if $\|a\|_2 = 1$ (which we will later show is true for optimal t), $f(x_i)$ gives the Euclidean distance between x_i and the hyperplane $\{z : a^T z = b\}$, and we can make a similar statement for y_j. Thus, this is geometrically equivalent to finding the thickest *slab* that separates the sets of points as shown below.



- The Problem

  (a) Show that the optimal value $t^\star$ is positive if and only if the two sets of points can be linearly separated. When this holds, show that the linear inequality $\|a\|_2 \leq 1$ is tight.

  Proof. If $t^* > 0$, then $a^T x_i - b \geq t^* > 0$ and $a^T y_j - b \leq -t^* < 0$, so the sets of points can be linearly separated. Conversely, if the sets of points can be linearly separated, then for each i and j we have $a^T x_i - b = \epsilon_i > 0$ and $a^T y_j - b = -\delta_j < 0$. Thus, $t^* \geq t = \min_{i\in[n],j\in[m]}\{\epsilon_i, \delta_j\} > 0$.
  Now, let us assume the points can be linearly separated, and that a^* is the value for a attained for the optimal t^*. Clearly, $\|a^*\|_2 \leq 1$, and choose some $\alpha \geq 1$ so that $\|\alpha a^*\|_2 = 1$. If $a^T x_i - b \geq t^*$, then clearly $(\alpha a)^T x_i - \alpha b \geq \alpha t^*$. Similarly $(\alpha a)^T y_j - \alpha b \leq -\alpha t^*$. Due to the optimality of t^*, we know $\alpha t^* \leq t^*$, and hence $\alpha \leq 1$. Thus, $\alpha = 1$, which means $\|a^*\|_2 = 1$. □

  (b) Assume $t > 0$, and let $\tilde{a} = a/t$ and $\tilde{b} = b/t$. Prove that the robust linear discrimination problem above is equivalent to the quadratic program

$$\begin{aligned} \text{minimize:} \quad & \|\tilde{a}\|_2 \\ \text{subject to:} \quad & \tilde{a}^T x_i - \tilde{b} \geq 1, \quad i \in [n] \\ & \tilde{a}^T y_j - \tilde{b} \leq -1, \quad j \in [m] \end{aligned}$$

  Proof. Consider the original problem. Since $t > 0$, we can divide the constraints by t. This gives the problem:

$$\begin{aligned} \text{maximize:} \quad & t \\ \text{subject to:} \quad & (a/t)^T x_i - b/t \geq 1, \quad i \in [n] \\ & (a/t)^T y_j - b/t \leq -1, \quad j \in [m] \\ & \|a/t\|_2 \leq 1/t \end{aligned}$$

  Now, making the substitution for $\tilde{a}$ and $\tilde{b}$ we get

$$\begin{aligned} \text{maximize:} \quad & t \\ \text{subject to:} \quad & \tilde{a}^T x_i - \tilde{b} \geq 1, \quad i \in [n] \\ & \tilde{a}^T y_j - \tilde{b} \leq -1, \quad j \in [m] \\ & \|\tilde{a}\|_2 \leq 1/t \end{aligned}$$

  Finally, note that maximizing t is equivalent to minimizing $\|\tilde{a}\|$ due to the last constraint. This yields the desired program. Notice that $\|\tilde{a}\| \geq 0$, so we can equivalently minimize $\|\tilde{a}\|^2$. This is indeed a quadratic function, and the reason the above was referred to as a quadratic program. □

- The Dual

We will use the Lagrangian dual method. We let X be the matrix of x_is, and Y be the matrix of y_js, and rewrite the original problem as follows:

$$\begin{array}{ll} \text{maximize:} & t \\ \text{subject to:} & [-X11][a,b,c]^T \preceq 0, \\ & [Y-11][a,b,c]^T \preceq 0, \\ & \|a\|_2^2 - 1 \leq 0 \end{array}$$

where the first constraint corresponds to u, the second constraint corresponds to v, and the last constraint corresponds to λ.

Thus, the Lagrangian is

$$L(a,b,t,u,v,\lambda) = -t + u^T[-X11][a,b,c]^T + v^T[Y-11][a,b,c]^T + \lambda(\|a\|_2^2 - 1) + \delta_{u\in\mathbb{R}_+^n} + \delta_{v\in R_+^m} + \delta_{\lambda\in R_+}.$$

Thus, it is clear that

$$\begin{array}{rcl} \nabla_t L & = & 1^T u + 1^T v - 1, \\ \nabla_b L & = & 1^T u + 1^T v, \\ \nabla_a L & = & Y^T v - X^T u + 2a\lambda. \end{array}$$

Since we are interested in the case where $\nabla_{a,b,t} = 0$, the first two constraints imply $1^T u = 1/2 = 1^T v$. Therefore, the Lagrangian dual function becomes

$$\begin{array}{rcl} g(u,v,\lambda) & = & \inf_{a,b,t}\{L(a,b,t,u,v,\lambda)\} \\ & = & \inf_{\substack{a \\ 1^T u = 1/2 = 1^T v \\ u,v,\lambda\succeq 0}} \{a^T[Y^T v - X^T u] + \lambda(a^T a - 1)\}. \end{array}$$

If we now use the third constraint, we see that

$$\begin{array}{rcl} g(u,v,\lambda) & = & \inf_{\substack{a \\ 1^T u = 1/2 = 1^T v \\ u,v,\lambda\succeq 0}} \{-a^T a\lambda - \lambda\} \\ & = & \begin{cases} -\lambda, & \text{when constraints satisfied;} \\ -\infty, & \text{otherwise.} \end{cases} \end{array}$$

Lastly, we note that the constraints above imply $\|Y^T v - X^T u\|_2 \leq \lambda$. Thus, we can write the dual problem as:

$$\begin{array}{lll} \text{maximize:} & -\|Y^T v - X^T u\|_2 & \\ \text{subject to:} & u \succeq 0 & v \succeq 0 \\ & 1^T u = 1/2 = 1^T v & \end{array}$$

To give a geometric interpretation, we note that $2Y^T v$ is a point in the convex hull of the y_is. Similarly, $2X^T u$ is a point in the convex hull of the x_is. Hence, this problem is minimizing the maximum distance between the convex hulls of x and y as shown below.