Chapter 8 Geometric problems

Last update on 2025-04-24 18:12:06+08:00

Table of contents

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

- the distance of $x_0 \in \mathbb{R}^n$ to a closed set $C \subseteq \mathbb{R}^n$, in the norm $\|\cdot\|$, is defined as

$$\mathbf{dist}(x_0, C) = \mathbf{inf}\{\|x_0 - x\| | x \in C\}.$$

- any point $z \in C$ which is closest to x_0 is referred to as a projection of x_0 on C.
- if for every x_0 there is a unique Euclidean projection of x_0 on C, then C is closed and convex.
- $P_C : \mathbb{R}^n \to \mathbb{R}^n$ denotes any function for which $P_C(x_0)$ is a projection of x_0 on C,

$$P_C(x_0) \in C, \qquad \|x_0 - P_C(x_0)\| = \mathbf{dist}(x_0, C).$$

We refer to P_C as projection on C.

Example

Projection on the unit square in $\mathbb{R}^2$

- consider $x_0 = 0$ and the boundary of the unit square $C = \{x \in \mathbb{R}^2 | \|x\|_\infty = 1\}$
- for both ℓ_1-norm and ℓ_2-norm, the four points $(1,0), (0,-1), (-1,0), (0,1)$ are closest to $x_0 = 0$, with distance 1
- in the ℓ_∞-norm, all points in C lie at a distance 1 from x_0, $\mathbf{dist}(x_0, C) = 1$

Projection onto rank-k matrices

- Let $X_0 \in \mathbb{R}^{m \times n}$ and $k \leq \min\{m, n\}$

$$C = \{X \in \mathbb{R}^{m \times n} | \, \mathbf{rank}\, X \leq k\}$$

- we can find a projection of X_0 on C, in the spectral of maximum singular value norm $\|\cdot\|_2$, via the singular value decomposition.
- $Y = \sum_{i=1}^{\min\{k,r\}} \sigma_i u_i v_i^T$ is a projection of X_0 on C, where $X_0 = \sum_{i=1}^{r} \sigma_i u_i v_i^T$ is a singular value decomposition with $r = \mathbf{rank}\, X_0$.

projection a point on a convex set

the convex C is $Ax = b, \quad f_i(x) \leq 0, i = 1, \ldots, m$, then one can find the projection of x_0 on C by solving with variable x

$$\begin{array}{ll} \text{minimize} & \|x - x_0\| \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

Euclidean projection on a polyhedron

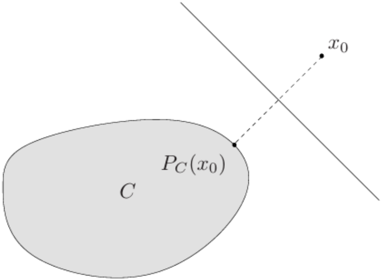
- the polyhedron is $Ax \preceq b$
- the Euclidean projection on a hyperplane $C = \{x | a^T x = b\}$ is

$$P_C(x_0) = x_0 + (b - a^T x_0)a / \|a\|_2^2$$

- the Euclidean projection on a halfspace $C = \{x | a^T x \leq b\}$ is

$$P_C(x_0) = \begin{cases} x_0 + (b - a^T x_0)a / \|a\|_2^2, & a^T x_0 > b \\ x_0, & a^T x_0 \leq b \end{cases}$$

separating a point and a convex set



If $P_C(x_0)$ denotes the Euclidean projection of x_0 on C, then the hyperplane

$$(P_C(x_0) - x_0)^T(x - (x_0 + P_C(x_0))/2) = 0$$

(strictly) separates x_0 from C

in other norms, the link between the projection problem and the separating hyperplane problem is via Lagrange duality

- the primal convex optimization problem is for variables x and y

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m \\ & Ax = b \\ & x - x_0 = y \end{array}$$

- the Lagrangian is

$$L(x, y, \lambda, \mu, \nu) = \|y\| + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T(Ax - b) + \mu^T(x_0 - x - y)$$

and the dual function is

$$g(\lambda, \mu, \nu) = \begin{cases} \mathbf{inf}_x(\sum_{i=1}^m \lambda_i f_i(x) + \nu^T(Ax - b) + \mu^T(x_0 - x)) & \|\mu\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

- the dual problem is for variables λ, μ, ν

$$\begin{array}{ll} \text{maximize} & \mu^T x_0 + \inf_x \left(\sum_{i=1}^m \lambda_i f_i(x) + \nu^T(Ax - b) - \mu^T x\right) \\ \text{subject to} & \lambda \succeq 0 \\ & \|\mu\|_* \leq 1 \end{array}$$

- suppose $x_0 \notin C$ and strong duality holds, then λ, μ, ν are dual feasible with a positive dual objective value,

$$\mu^T(x_0 - x) + \sum_{i=1}^m \lambda_i f_i(x) + \nu^T(Ax - b) > 0$$

  for all feasible x
- this implies

$$\mu^T x_0 > \mu^T x$$

  for $x \in C$, and therefore μ defines a strictly separating hyperplane

separating a point from a polyhedron

the dual problem of

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & Ax \preceq b \\ & x_0 - x = y \end{array}$$

is

$$\begin{array}{ll} \text{maximize} & \mu^T x_0 - b^T \lambda = (Ax_0 - b)^T \lambda \\ \text{subject to} & A^T \lambda = \mu \\ & \|\mu\|_* \leq 1 \\ & \lambda \succeq 0 \end{array}$$

if x is feasible and the dual objective is positive, then

$$(A^T\lambda)^T x = \lambda^T (Ax) \leq \lambda^T b < \lambda^T A x_0,$$

so $\mu = A^T \lambda$ defines a separating hyperplane.

projection and separation via indicator and support functions

the indicator function I_C and the support function S_C of the set C is defined as

$$S_C(x) = \sup_{y \in C} x^T y, \qquad I_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$$

the problem of projecting x_0 on a closed convex set C is

$$\begin{array}{ll} \text{minimize} & \|x - x_0\| \\ \text{subject to} & I_C(x) \le 0 \end{array}$$

the dual problem

$$\begin{array}{ll} \text{maximize} & z^T x_0 - S_C(z) \\ \text{subject to} & \|z\|_* \le 1 \end{array}$$

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

definition

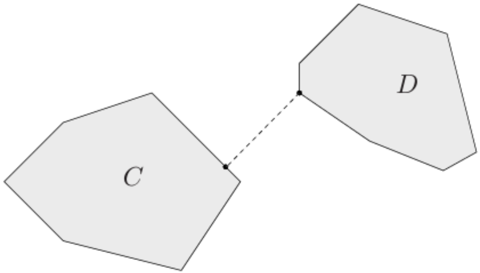
the distance between C and D, in a norm $\|\cdot\|$, is defined as

$$\mathbf{dist}(C, D) = \mathbf{inf}\{\|x - y\| | x \in C, y \in D\}$$

the distance between sets can be expressed in terms of the distance between a point and a set,

$$\mathbf{dist}(C, D) = \mathbf{dist}(0, D - C)$$

computing the distance between convex sets



find $\mathbf{dist}(C, D)$ by solving the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \|x - y\| \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & g_i(y) \leq 0, \quad i = 1, \ldots, p \end{array}$$

where C and D are convex

$$C = \{x | f_i(x) \leq 0, i = 1, \ldots, m\}, \qquad D = \{x | g_i(x) \leq 0, i = 1, \ldots, p\}.$$

separating convex sets

the primal problem is

$$\begin{array}{ll} \text{minimize} & \|w\| \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & g_i(y) \leq 0, \quad i = 1, \ldots, p \\ & x - y = w. \end{array}$$

the dual function is

$$g(\lambda, z, \mu) = \begin{cases} \mathbf{inf}_x(\sum_{i=1}^m \lambda_i f_i(x) + z^T x) + \mathbf{inf}_y(\sum_{i=1}^p \mu_i g_i(y) - z^T y) & \|z\|_* \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

the dual problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{inf}_x(\sum_{i=1}^m \lambda_i f_i(x) + z^T x) + \mathbf{inf}_y(\sum_{i=1}^p \mu_i g_i(y) - z^T y) \\ \text{subject to} & \|z\|_* \leq 1 \\ & \lambda \succeq 0, \quad \mu \succeq 0. \end{array}$$

geometric interpretation: if λ, μ are dual feasible with a positive objective value, then

$$\sum_{i=1}^{m} \lambda_i f_i(x) + z^T x + \sum_{i=1}^{p} \mu_i g_i(y) - z^T y > 0$$

for all x and y.

In particular, for $x \in C$ and $y \in D$, we have $z^T x - z^T y > 0$, so z defines a hyperplane that strictly separates C and D

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

the Löwner-John ellipsoid

suppose $C \subseteq \mathbb{R}^n$ is bounded and has nonempty interior

the minimum volume ellipsoid that contains a set C is called the Löwner-John ellipsoid of C

parametrize the general ellipsoid as

$$\mathcal{E} = \{v | \|av + b\|_2 \leq 1\}$$

since the volume of $\mathcal{E}$ is proportional to $\det A^{-1}$, the convex optimization problem is

$$\begin{array}{ll} \text{minimize} & \log \det A^{-1} \\ \text{subject to} & \sup_{v \in C} \|Av + b\|_2 \leq 1, \end{array}$$

where $A \in \mathbb{S}^n_{++}$ and $b \in \mathbb{R}^n$

minimum volume ellipsoid covering a finite set

we consider the problem of finding the minimum volume ellipsoid that contains the finite set $C = \{x_1, \ldots, x_m\} \subseteq \mathbb{R}^n$

an ellipsoid covers C if and only if it covers its convex hull

the convex optimization problem becomes

$$\begin{array}{ll} \text{minimize} & \log\det A^{-1} \\ \text{subject to} & \|Av + b\|_2 \le 1, \qquad i = 1, \ldots, m \end{array}$$

where $A \in \mathbb{S}^n_{++}$ and $b \in \mathbb{R}^n$

maximum volume inscribed ellipsoid

find the ellipsoid of maximum volume that lies inside a convex set C, which we assume is bounded and has nonempty interior

parametrize the ellipsoid as the image of the unit ball under an affine transformation,

$$\mathcal{E} = \{Bu + d \mid \|u\|_2 \leq 1, B \in \mathbb{S}^n_{++}\}$$

so the volume is proportional to $\det B$

find the maximum volume ellipsoid inside C by solving the convex optimization problem

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \sup\limits_{\|u\|_2 \leq 1} I_C(Bu + d) \leq 0 \end{array}$$

in the variables $B \in \mathbb{S}^n_{++}$ and $d \in \mathbb{R}^n$

maximum volume ellipsoid in a polyhedron

C is a polyhedron,

$$C = \{x | a_i^T x \leq b_i, i = 1, \ldots, m\}$$

$$\sup_{\|u\|_2 \leq 1} I_C(Bu + d) \leq 0 \iff \sup_{\|u\|_2 \leq 1} a_i^T (Bu + d) \leq b_i, \quad i = 1, \ldots, m$$

$$\iff \|Ba_i\|_2 + a_i^T d \leq b_i, \qquad i = 1, \ldots, m$$

the convex optimization problem is

$$\begin{array}{ll} \text{minimize} & \log\det B^{-1} \\ \text{subject to} & \|Ba_i\|_2 + a_i^T d \leq b_i, \qquad i = 1, \ldots, m \end{array}$$

in the variables $B \in \mathbb{S}^n_{++}$ and $d \in \mathbb{R}^n$

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

Chebyshev center

the depth of a point $x \in C$ is defined as

$$\mathbf{depth}(x, C) = \mathbf{dist}(x, \mathbb{R}^n \backslash C).$$

The depth gives the radius of the largest ball, centered at x, that lies in C

A Chebyshev center of the set C is defined as any point of maximum depth in C:

$$x_{\text{cheb}}(C) = \text{argmax}\, \mathbf{depth}(x, C) = \text{argmax}\, \mathbf{dist}(x, \mathbb{R}^n \backslash C)$$

Chebyshev center of a convex set

Let convex $C = \{x | f_1(x) \leq 0, \ldots, f_m(x) \leq 0\}$. A Chebyshev center can be solved by

$$\begin{array}{ll} \text{maximize} & R \\ \text{subject to} & g_i(x, R) \leq 0, \qquad i = 1, \ldots, m \end{array}$$

where $g_i(x, R) = \mathbf{sup}_{\|u\| \leq 1} f_i(x + Ru)$

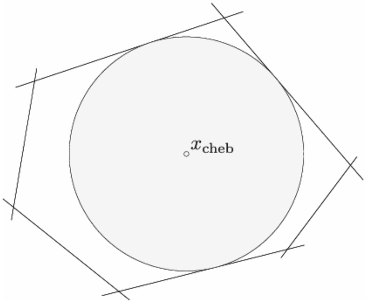
Chebyshev center of a polyhedron

- the polyhedron is characterized by $a_i^T x \le b_i, \ i = 1, \ldots, m$. Then

$$g_i(x, R) = \sup_{\|u\| \le 1} a_i^T (x + Ru) - b_i = a_i^T x + R\|a_i\|_* - b_i$$

- Chebyshev center can be found by solving the LP

$$\begin{array}{ll} \text{maximize} & R \\ \text{subject to} & a_i^T x + R\|a_i\|_* - b_i \le 0, \qquad i = 1, \ldots, m \\ & R \ge 0 \end{array}$$

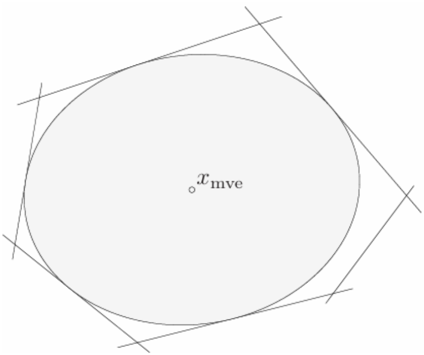


Chebyshev center of a polyhedron C, in the Euclidean norm

maximum volume ellipsoid center

As an extension, the maximum volume ellipsoid center of C, denoted by x_{mve}, as the center of the maximum volume ellipsoid that lies in C

the problem is readily computed as in last section

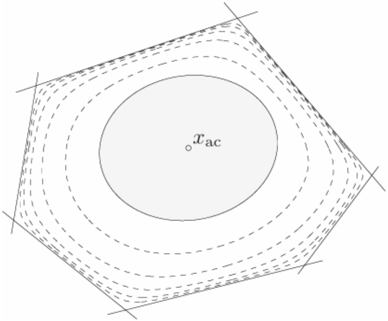


the maximum volume ellipsoid contained in the polyhedron C

analytic center of a set of inequalities

the analytic center x_{ac} of a set of convex inequalities and linear equalities,
$f_i(x) \leq 0,\ i = 1, \ldots, m, \quad Fx = g$ is defined as an optimal point for the convex problem

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^{m} \log(-f_i(x)) \\ \text{subject to} & Fx = g \\ & f_i(x) < 0, \qquad i = 1, \ldots, m \end{array}$$



the dashed lines show five level curves of the logarithmic barrier function

Projection on a set

Distance between sets

Extremal volume ellipsoids

Centering

Classification

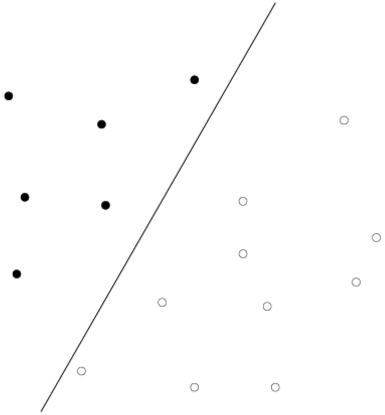
setup of the classification problems

- pattern recognition and classification problems
- given two sets of points in $\mathbb{R}^n$, $\{x_1, \dots, x_N\}$ and $\{y_1, \dots, y_M\}$
- wish to find a function
$$f : \mathbb{R}^n \longrightarrow \mathbb{R}$$
(within a given family of functions) that is positive on the first set and negative on the second, i.e.,
$$f(x_i) > 0, \quad i = 1, \dots, N, \qquad\qquad f(y_i) < 0, \quad i = 1, \dots, M.$$
- If these inequalities hold, we say that f, or its 0-level set $\{x | f(x) = 0\}$, separates, classifies, or discriminates the two sets of points.

Linear discrimination

in linear discrimination, we seek an affine $f(x) = a^T x - b$ that classifies the points,

$$a^T x_i - b > 0, \quad i = 1, \dots, N, \qquad\qquad a^T y_i - b < 0, \quad i = 1, \dots, M.$$



two sets are classified by an affine function f

robust linear discrimination

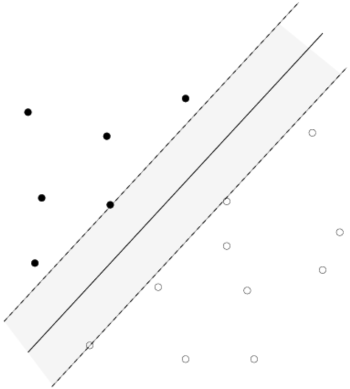
$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & a^T x_i - b \geq t, \qquad i = 1, \ldots, N \\ & a^T y_i - b \leq -t, \qquad i = 1, \ldots, M \\ & \|a\|_2 \leq 1, \end{array}$$

with variables a, b, t

we have to normalize a and b, since otherwise we can scale a and b by a positive constant and make the gap in the values arbitrarily large

the optimal t^* is positive if and only if the two sets of points can be linearly discriminated. In this case, $\|a^*\|_2 = 1$

Geometrically, if $\|a\|_2 = 1$, $a^T x_i - b$ is the Euclidean distance from the point x_i to the separating hyperplane $\mathcal{H} = \{z | a^T z = b\}$. Similarly, $b - a^T y_i$ is the distance from the point y_i to the hyperplane. In other words, it finds the thickest slab that separates the two sets.



- solving the robust linear discrimination problem
- Geometrically, we find the thickest slab that separates the two sets of points
- the optimal value t^* (which is half the slab thickness) turns out to be half the distance between the convex hulls of the two sets of points

- The Lagrangian (for the problem of minimizing $-t$) is

$$-t + \sum_{i=1}^{N} u_i(t + b - a^T x_i) + \sum_{i=1}^{M} v_i(t - b + a^T y_i) + \lambda(\|a\|_2 - 1).$$

- Minimizing over b and t yields the conditions $\mathbf{1}^T u = 1/2, \mathbf{1}^T v = 1/2$. Then

$$g(u, v, \lambda) = \begin{cases} -\lambda & \|\sum_{i=1}^{M} v_i y_i - \sum_{i=1}^{N} u_i x_i\|_2 \le \lambda \\ -\infty & \text{otherwise} \end{cases}$$

- the dual problem becomes

$$\begin{array}{ll} \text{maximize} & -\|\sum_{i=1}^{M} v_i y_i - \sum_{i=1}^{N} u_i x_i\|_2 \\ \text{subject to} & u \succeq 0, \quad \mathbf{1}^T u = 1/2, \\ & v \succeq 0, \quad \mathbf{1}^T v = 1/2. \end{array}$$

- The dual objective is to minimize (half) the distance between these two points, i.e., find (half) the distance between the convex hulls of the two sets.

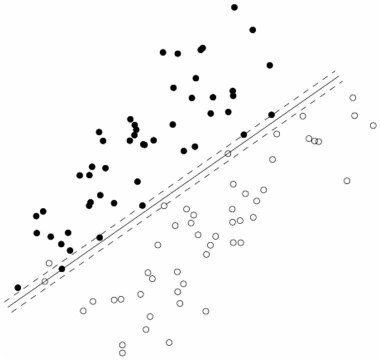
support vector classifier

- When the two sets of points cannot be linearly separated, we might seek an affine function that approximately classifies the points
- we relax the constraints

$$a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, N, \qquad a^T y_i - b \leq -(1 - v_i), \quad i = 1, \ldots, M.$$

- When $u = v = 0$, we recover the original constraints
- by making u and v large enough, these inequalities can always be made feasible
- we minimize by solving the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u + \mathbf{1}^T v \\ \text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, N \\ & a^T y_i - b \leq -(1 - v_i), \quad i = 1, \ldots, M \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$



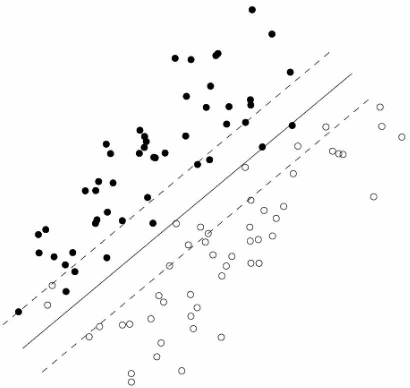
- approximate linear discrimination via linear programming
- This classifier misclassifies one point
- Four points are correctly classified, but lie in the slab defined by the dashed lines
- it is a relaxation of the number of points misclassified by the function $a^T z - b$, plus the number of points that are correctly classified but lie in the slab defined by $-1 < a^T z - b < 1$

trade-off

- we can consider the trade-off between the number of misclassified points, and the width of the slab $\{z | -1 \leq a^T z - b \leq 1\}$, which is given by $2/\|a\|_2$.
- we minimize by solving the LP

$$
\begin{array}{ll}
\text{minimize} & \|a\|_2 + \gamma(\mathbf{1}^T u + \mathbf{1}^T v) \\
\text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, N \\
& a^T y_i - b \leq -(1 - v_i), \quad i = 1, \ldots, M \\
& u \succeq 0, \quad v \succeq 0
\end{array}
$$

- The first term is proportional to the inverse of the width of the slab defined by $-1 \leq a^T z - b \leq 1$
- The second term has the same interpretation as above, i.e., it is a convex relaxation for the number of misclassified points
- The parameter γ, which is positive, gives the relative weight of the number of misclassified points (which we want to minimize), compared to the width of the slab (which we want to maximize).



- support vector classifier
- misclassify three points
- Fifteen points are correctly classified but lie in the slab defined by the dashed lines

linear discrimination via logistic modeling

- Suppose z is a random variable with values 0 or 1, with a distribution that depends on some (deterministic) explanatory variable $u \in \mathbb{R}^n$, via a logistic model of the form

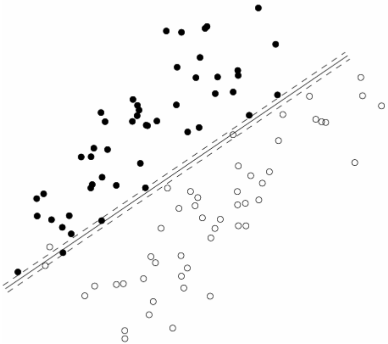
$$\begin{aligned} \mathbf{prob}(z=1) &= (\exp(a^T u - b))/(1+\exp(a^T u - b)) \\ \mathbf{prob}(z=0) &= 1/(1+\exp(a^T u - b)) \end{aligned}$$

- Now we assume that the given sets of points, $\{x_1, ..., x_N\}$ and $\{y_1, ..., y_M\}$, arise as samples from the logistic model
- $\{x_1, ..., x_N\}$ are the values of u for the N samples for which $z = 1$, and $\{y_1, ..., y_M\}$ are the values of u for the M samples for which $z = 0$
- determine a and b by maximum likelihood estimation by solving

$$\text{minimize} \qquad -l(a,b)$$

with variables a, b, where l is the log-likelihood function

$$l(a,b) = \sum_{i=1}^{N}(a^T x_i - b) - \sum_{i=1}^{N}\log(1+\exp(a^T x_i - b)) - \sum_{i=1}^{M}\log(1+\exp(a^T y_i - b))$$



- linear discrimination via logistic modeling
- misclassify two points
- three points are correctly classified, but lie in between the dashed lines
- there is a **Bayesian interpolation**

Nonlinear discrimination

seek a nonlinear function f, from a given subspace of functions, that is positive on one set and negative on another

$$f(x_i) > 0, \quad i = 1, \ldots, N, \qquad\qquad f(y_i) < 0, \quad i = 1, \ldots, M.$$

Provided f is linear (or affine) in the parameters that define it, these inequalities can be solved in exactly the same way as in linear discrimination.

- quadratic discrimination
- polynomial discrimination

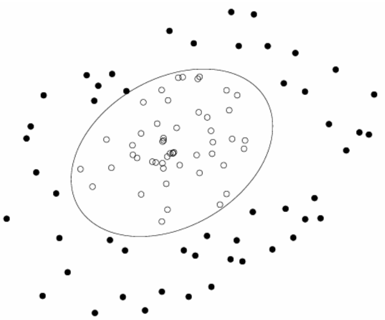
quadratic discrimination

- Suppose we take f to be quadratic: $f(x) = x^T Px + q^T x + r$, where $P \in \mathbb{S}^n, q \in \mathbb{R}^n, r \in \mathbb{R}$
- find a solution to the strict inequalities by solving the nonstrict feasibility problem

$$
\begin{aligned}
x_i^T Px_i + q^T x_i + r \geq 1, &\quad i = 1, \ldots, N, \\
y_i^T Py_i + q^T y_i + r \leq -1, &\quad i = 1, \ldots, M.
\end{aligned}
$$

- we can require that $P \prec 0$,which means the separating surface is ellipsoidal. This quadratic discrimination problem can be solved as an SDP feasibility problem

$$
\begin{array}{rl}
\text{find} & P,\ q,\ r \\
\text{subject to} & x_i^T Px_i + q^T x_i + r \geq 1, \quad i = 1, \ldots, N, \\
& y_i^T Py_i + q^T y_i + r \leq -1, \quad i = 1, \ldots, M, \\
& P \preceq -I.
\end{array}
$$



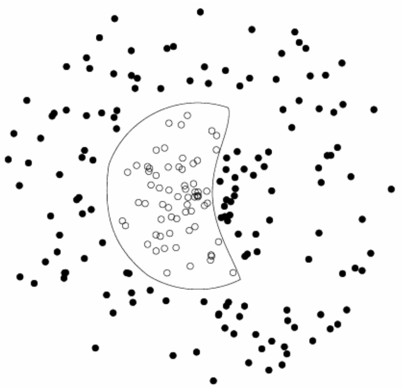
- quadratic discrimination
- $P \prec 0$
- SDP feasibility problem

polynomial discrimination

- consider the set of polynomials on $\mathbb{R}^n$ with degree less than or equal to d:

$$f(x) = \sum_{i_1+\ldots+i_d \leq d} a_{i_1 \cdots i_d} x_1^{i_1} \cdots x_n^{i_n}$$

- We can determine whether or not two sets $\{x_1, \ldots, x_N\}$ and $\{y_1, \ldots, y_M\}$ can be separated by such a polynomial by solving a set of linear inequalities in the variables $a_{i_1 \cdots i_d}$
- As an extension, the problem of determining the minimum degree polynomial on $\mathbb{R}^n$ that separates two sets of points can be solved via quasiconvex programming, since the degree of a polynomial is a quasiconvex function of the coefficients.
- This can be carried out by bisection on d, solving a feasibility linear program at each step.



- minimum degree polynomial discrimination in $\mathbb{R}^2$
- there exists no cubic polynomial that separates two sets of points
- they can be separated by fourth-degree polynomial, the zero level set of which is shown