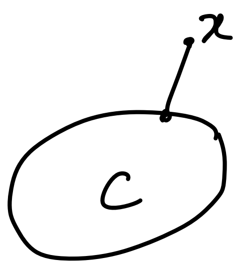
8. Geometric problems

- projection on a set
- extremal volume ellipsoids
- centering
- classification

Projection on convex set



projection of point x on set C defined as

$$P_C(x) = \operatorname{argmin}_{z \in C} \|x - z\|_2$$

i.e., point in C closest to x. suppose C has form

$$C = \{\, x \mid Ax = b,\ f_i(x) \le 0,\ i = 1, \dots, m \,\}$$

$f_i : \mathbf{R}^n \to \mathbf{R}$ convex

$$P_C(x) = \begin{array}{ll} \underset{z}{\operatorname{argmin}} & \|x - z\|_2 \\ \text{subject to} & Az = b \\ & f_i(z) \le 0,\ i = 1, \dots, m \end{array}$$

computing $P_C(x)$ is cvx opt problem

Distance between convex sets

distance between sets C, $\tilde{C}$ defined as

$$\mathrm{dist}(C, \tilde{C}) = \min_{z \in C,\ \tilde{z} \in \tilde{C}} \|z - \tilde{z}\|_2$$

suppose C, $\tilde{C}$ are convex, with form

$$C = \{ x \mid Ax = b,\ f_i(x) \leq 0,\ i = 1, \ldots, m \}$$

$$\tilde{C} = \{ x \mid \tilde{A}x = \tilde{b},\ \tilde{f}_i(x) \leq 0,\ i = 1, \ldots, \tilde{m} \}$$

$f_i, \tilde{f}_i : \mathbf{R}^n \to \mathbf{R}$ convex

$\mathrm{dist}(C, \tilde{C})$ is optimal value of cvx problem

$$\begin{array}{ll} \underset{z,\tilde{z}}{\text{minimize}} & \|z - \tilde{z}\| \\ \text{subject to} & Az = b,\ \tilde{A}\tilde{z} = \tilde{b} \\ & f_i(z) \leq 0,\ i = 1, \ldots, m \\ & \tilde{f}_i(\tilde{z}) \leq 0,\ i = 1, \ldots, \tilde{m} \end{array}$$

Intersection & containment of polyhedra

inequality description

$$\begin{aligned} \mathcal{P}_1 &= \{x \mid \underline{a_i^T x \leq b_i}, i = 1, \ldots, m\} = \{\underline{x \mid Ax \preceq b}\} \\ \mathcal{P}_2 &= \{x \mid \underline{f_i^T x \leq g_i}, i = 1, \ldots, l\} = \{\underline{x \mid Fx \preceq g}\} \end{aligned}$$

- $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$? solve feasibility problem

$$\underline{Ax \preceq b}, \quad \underline{Fx \preceq g}$$

- $\underline{\mathcal{P}_1 \subseteq \mathcal{P}_2}$? for $k = 1, \ldots, l$, check

$$\sup\{\underline{f_k^T x} \mid \underbrace{Ax \preceq b}\} \overset{?}{\leq} \underline{g_k}, \quad , \forall k$$

$i.e.$, solve LPs

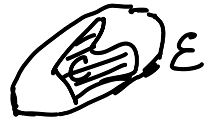
$$\begin{array}{ll} \text{maximize} & \underline{f_k^T x} \\ \text{subject to} & \underline{Ax \preceq b} \end{array} \overset{?}{\leq} g_k$$

$\to \mathcal{E}=\{Bu+d \mid \|u\|_2 \le 1\}$ $\quad$ $\mathcal{E}=\{x \mid x^T P x + 2q^T x + r \le 1\}$

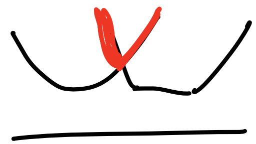
Minimum volume ellipsoid around a set

Löwner-John ellipsoid of a set C: minimum volume ellipsoid $\mathcal{E}$ s.t. $C \subseteq \mathcal{E}$

- parametrize $\mathcal{E}$ as $\mathcal{E} = \{v \mid \|Av+b\|_2 \le 1\}$; w.l.o.g. assume $A \in \mathbf{S}^n_{++}$
- $\mathbf{vol}\,\mathcal{E}$ is proportional to $\det A^{-1}$; to compute minimum volume ellipsoid,

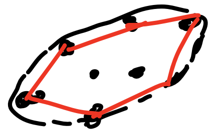


$$\begin{array}{ll} \text{minimize (over } A,\, b) & \log\det A^{-1} \\ \text{subject to} & \to \sup_{v\in C} \|Av+b\|_2 \le 1 \end{array}$$



convex, but evaluating the constraint can be hard (for general C)
(but can be infinite-dim — cannot solve)

finite set $C = \{x_1, \ldots, x_m\}$:



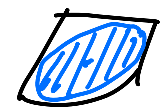
$$\begin{array}{ll} \text{minimize (over } A,\, b) & \log\det A^{-1} \\ \text{subject to} & \|Ax_i+b\|_2 \le 1, \quad i=1,\ldots,m \end{array}$$

also gives Löwner-John ellipsoid for polyhedron $\mathbf{conv}\{x_1,\ldots,x_m\}$

Maximum volume inscribed ellipsoid

maximum volume ellipsoid $\mathcal{E}$ inside a convex set $C \subseteq \mathbf{R}^n$

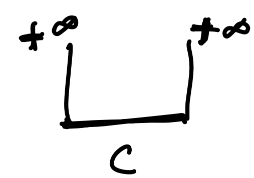
- parametrize $\mathcal{E}$ as $\mathcal{E} = \{Bu + d \mid \|u\|_2 \leq 1\}$; w.l.o.g. assume $B \in \mathbf{S}^n_{++}$
- $\mathbf{vol}\,\mathcal{E}$ is proportional to $\det B$; can compute $\mathcal{E}$ by solving



$$\begin{array}{ll} \underset{B,d}{\text{maximize}} & \log\det B \\ \text{subject to} & \sup_{\|u\|_2 \leq 1} \overset{\text{convex}}{I_C}(\underbrace{Bu + d}_{\text{lin. in } B,d}) \leq 0 \end{array}$$

$$I_c(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$$

(where $I_C(x) = 0$ for $x \in C$ and $I_C(x) = \infty$ for $x \notin C$)



convex, but evaluating the constraint can be hard (for general C)

polyhedron $\{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$:

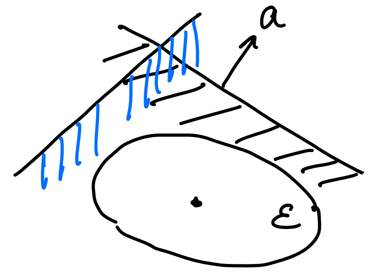
- extends Chebychev-center problem

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

$$\begin{array}{ll} \underset{c,r}{\max} & r \\ & B(c,r) \subset P \end{array}$$

(constraint follows from $\sup_{\|u\|_2 \leq 1} a_i^T(Bu + d) = \|Ba_i\|_2 + a_i^T d$)

$\mathcal{H} = \{x \mid a^T x \le b\}$ variables: B, d



$Bu+d \overset{?}{\in} \mathcal{H} \quad \forall \; \|u\|_2 \le 1$

$a^T(Bu+d) \overset{?}{\le} c \quad \forall \; \|u\|_2 \le 1$

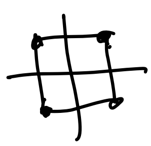
$\sup_{\|u\|_2 \le 1} (B^T a)^T u + a^T d \overset{?}{\le} b$ $\qquad (B^T a)^T u \le \underbrace{\|B^T a\|_2}\,\underbrace{\|u\|_2}_{\le 1}$ (C-S)

$\|B^T a\|_2 \le b - a^T d$ SOC in (B, d)

---

ellipsoidal fitting:

- statistics – ellipsoidal peeling ⟷ remove outliers
- dynamical systems in state-space – reachable sets
- box / rectangle / polyhedron can have large representations:



$2n$ ineq.

2^n vertices

norm balls are not flexible enough

ellipsoids: compact rep & flexible

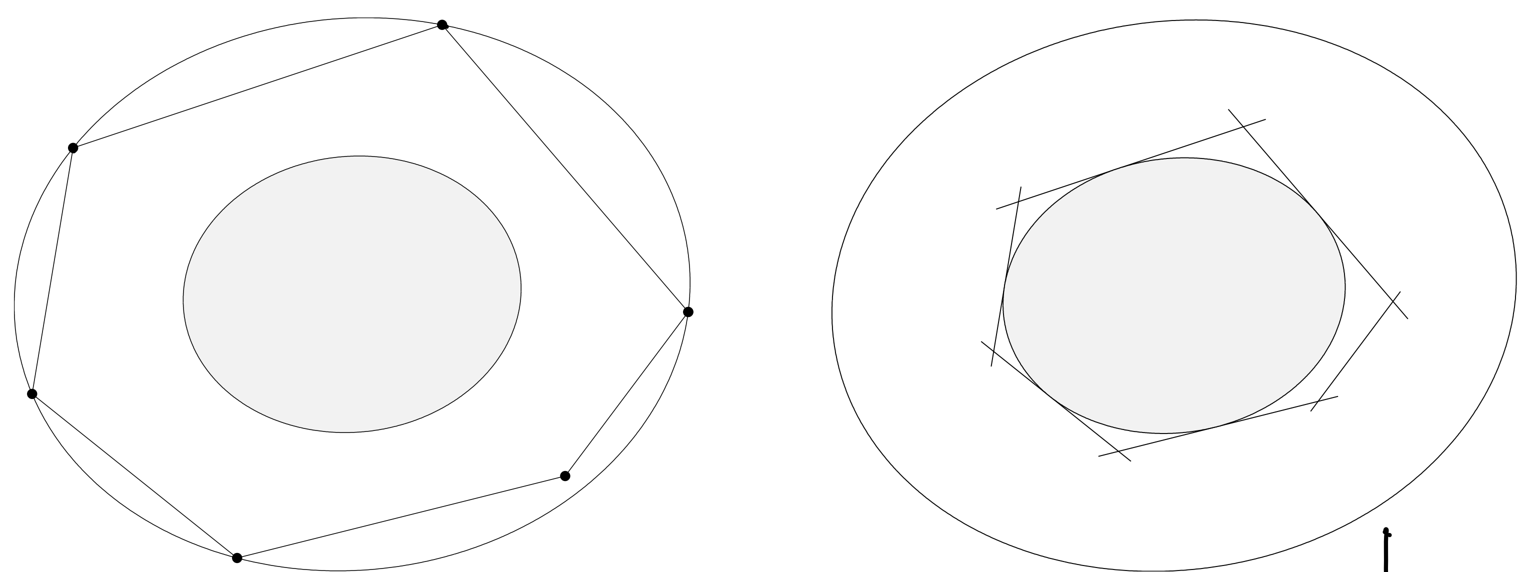
Efficiency of ellipsoidal approximations

$C \subseteq \mathbf{R}^n$ convex, bounded, with nonempty interior

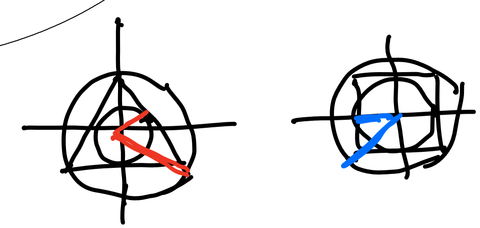
- Löwner-John ellipsoid, shrunk by a factor n, lies inside C
- maximum volume inscribed ellipsoid, expanded by a factor n, covers C

example (for two polyhedra in $\mathbf{R}^2$)

$n=2$



factor n can be improved to $\sqrt{n}$ if C is symmetric



(symm. wrt a center)

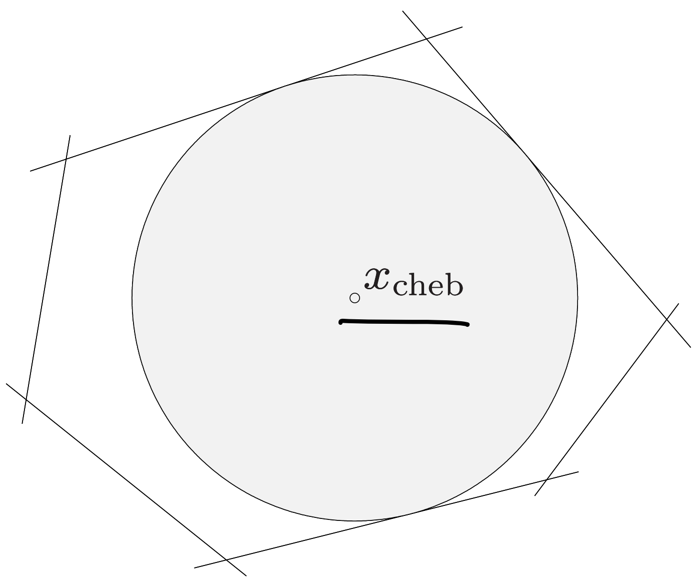
Centering

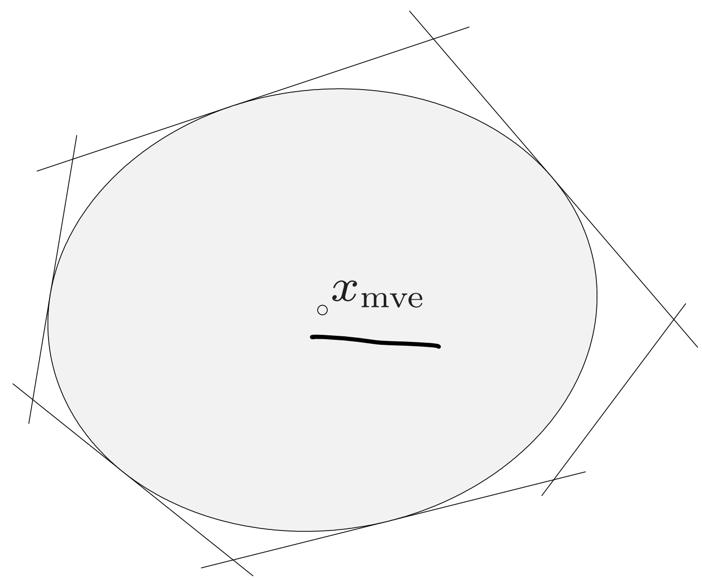
some possible definitions of 'center' of a convex set C:

- center of largest inscribed ball ('Chebyshev center')

  for polyhedron, can be computed via linear programming (see LP lecture)

- center of maximum volume inscribed ellipsoid (page 8–3)

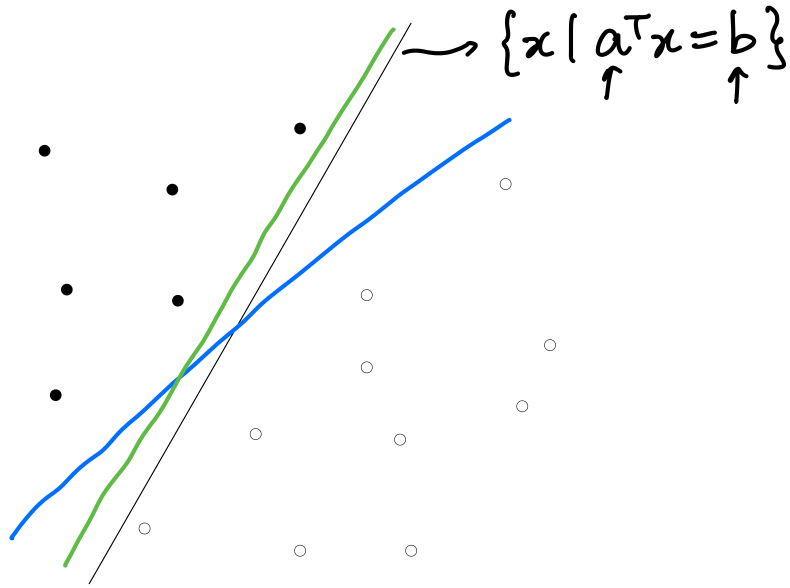




Linear discrimination

separate two sets of points $\{x_1, \ldots, x_N\}$, $\{y_1, \ldots, y_M\}$ by a hyperplane:

$$a^T x_i + b > 0, \quad i = 1, \ldots, N, \qquad a^T y_i + b < 0, \quad i = 1, \ldots, M$$



- training a classifier based on data
- use it to classify new data
- spam filter

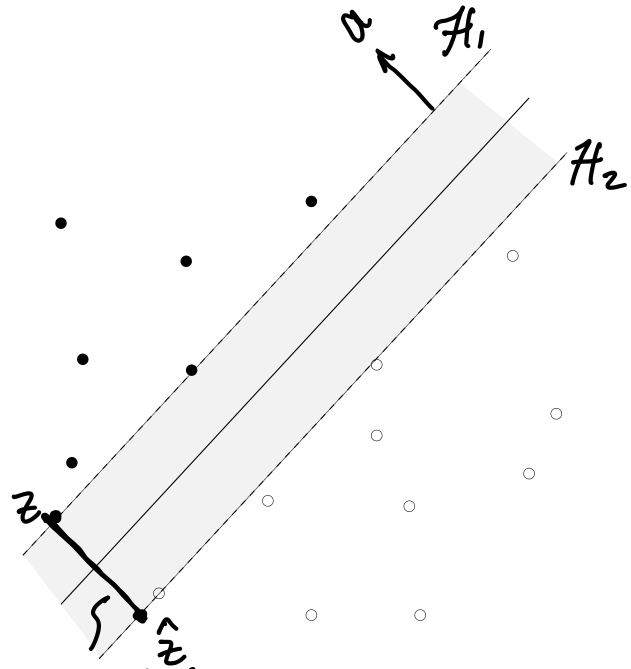
homogeneous in a, b, hence equivalent to

$$a^T x_i + b \geq 1, \quad i = 1, \ldots, N, \qquad a^T y_i + b \leq -1, \quad i = 1, \ldots, M$$

a set of linear inequalities in a, b

Robust linear discrimination

(Euclidean) distance between hyperplanes



parallel to $a = \beta a = \frac{2}{\|a\|_2^2} a$

$$
\begin{aligned}
\mathcal{H}_1 &= \{z \mid a^T z + b = 1\} \\
\mathcal{H}_2 &= \{z \mid a^T z + b = -1\}
\end{aligned}
$$

is $\mathbf{dist}(\mathcal{H}_1, \mathcal{H}_2) = 2/\|a\|_2$

$z \in \mathcal{H}_1 \quad a^T z + b = 1 \qquad a^T(\underbrace{z - \hat{z}}_{\beta a}) = 2 \Rightarrow \beta = \frac{2}{a^T a} = \frac{2}{\|a\|_2^2}$

$\hat{z} \in \mathcal{H}_2 \quad a^T \hat{z} + b = -1$

to separate two sets of points by maximum margin,

$$
\begin{array}{ll}
\underset{a,b}{\text{minimize}} & (1/2)\|a\|_2^2 \\
\text{subject to} & a^T x_i + b \geq 1, \quad i = 1, \ldots, N \\
 & a^T y_i + b \leq -1, \quad i = 1, \ldots, M
\end{array} \qquad (1)
$$

(after squaring objective) a QP in a, b

$$\left[\begin{array}{lll} \min\limits_{a,b} \ \frac{1}{2}\|a\|_2 & & \\ \text{s.t.} \quad a^T x_i + b \geq 1 & i=1,\dots,N & \lambda_i \\ \qquad\ \ a^T y_i + b \leq -1 & i=1,\dots,M & \mu_i \end{array}\right.$$

$$L(a,b,\lambda,\mu) = \frac{1}{2}\|a\|_2 + \sum_{i=1}^{N} \lambda_i (1 - a^T x_i - b) + \sum_{i=1}^{M} \mu_i (1 + a^T y_i + b)$$

$$g(\lambda,\mu) = \inf_{a,b} L$$

$$\inf_b \left(-b\sum\lambda_i + b\sum\mu_i\right) = b\left(\sum\mu_i - \sum\lambda_i\right) = \begin{cases} 0 & , \sum\mu_i = \sum\lambda_i \quad (1^T\mu = 1^T\lambda) \\ -\infty & , \text{else} \end{cases}$$

$$\inf_a \left(\frac{1}{2}\|a\| - a^T \underbrace{\left(\sum\lambda_i x_i - \sum\mu_i y_i\right)}_{}\right) = \inf_a \|a\|_2 \left(\frac{1}{2} - \left\|\sum_{i=1}^{N}\lambda_i x_i - \sum_{i=1}^{M}\mu_i y_i\right\|\right)$$

$$= \begin{cases} 0 & , \left\|\sum\lambda_i x_i - \sum\mu_i y_i\right\| \leq \frac{1}{2} \\ -\infty & , \text{else} \end{cases}$$

putting together:

$$\left[\begin{array}{l} \min\limits_{\lambda,\mu} \quad 1^T\lambda + 1^T\mu \\ 1^T\mu = 1^T\lambda, \ \lambda \geq 0, \ \mu \geq 0 \\ 2\left\|\sum\lambda_i x_i + \sum\mu_i y_i\right\| \leq 1 \end{array}\right.$$

Lagrange dual of maximum margin separation problem (1)

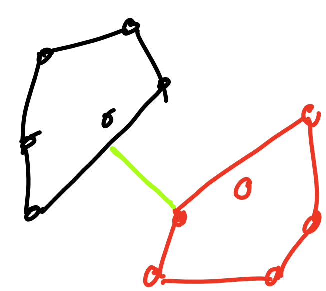
$$
\begin{array}{ll}
\text{maximize} & \mathbf{1}^T\lambda + \mathbf{1}^T\mu \\
\text{subject to} & 2\left\|\sum_{i=1}^N \dfrac{\overset{\theta_i}{\widetilde{\lambda_i}} x_i}{\mathbf{1}^T\lambda} - \sum_{i=1}^M \dfrac{\overset{\gamma_i}{\widetilde{\mu_i}} y_i}{\mathbf{1}^T\mu}\right\|_2 \le 1 \\
& \mathbf{1}^T\lambda = \mathbf{1}^T\mu, \quad \lambda \succeq 0, \quad \mu \succeq 0
\end{array}
\qquad (2)
$$

from duality, optimal value is inverse of maximum margin of separation

interpretation

- change variables to $\underline{\theta_i = \lambda_i/\mathbf{1}^T\lambda}$, $\underline{\gamma_i = \mu_i/\mathbf{1}^T\mu}$, $\underline{t = 1/(\mathbf{1}^T\lambda + \mathbf{1}^T\mu)} = \dfrac{1}{2(\mathbf{1}^T\lambda)}$
- invert objective to minimize $1/(\mathbf{1}^T\lambda + \mathbf{1}^T\mu) = t$

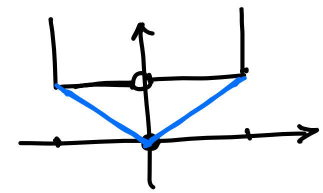
$$
\left[
\begin{array}{ll}
\underset{\theta,\gamma,t}{\text{minimize}} & t \\
\text{subject to} & \left\|\overbrace{\sum_{i=1}^N \theta_i x_i}^{\text{convex comb. of } x_i} - \overbrace{\sum_{i=1}^M \gamma_i y_i}^{\text{convex comb of } y_i}\right\|_2 \le t \\
& \theta \succeq 0, \quad \mathbf{1}^T\theta = 1, \quad \gamma \succeq 0, \quad \mathbf{1}^T\gamma = 1
\end{array}
\right.
$$



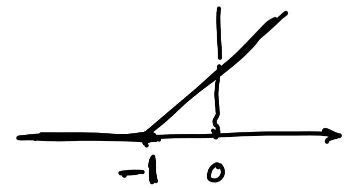
optimal value is distance between convex hulls

Approximate linear separation of non-separable sets

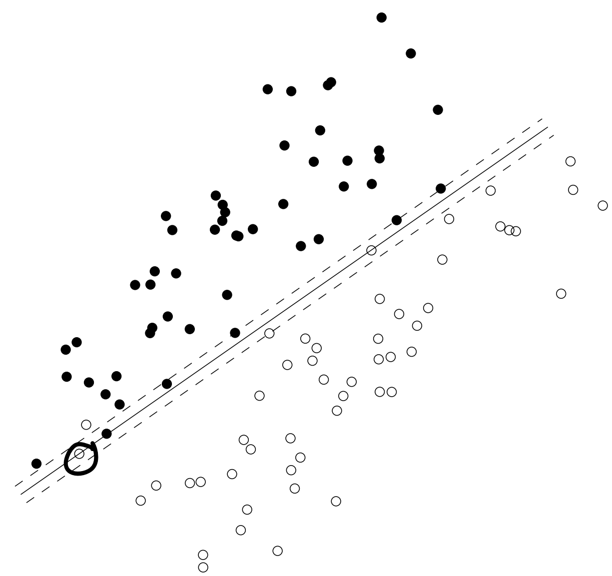
$$
\begin{array}{ll}
\underset{a,b,u,v}{\text{minimize}} & \mathbf{1}^T u + \mathbf{1}^T v \\
\text{subject to} & a^T x_i + b \geq 1 - u_i, \quad i = 1, \ldots, N \\
& a^T y_i + b \leq -1 + v_i, \quad i = 1, \ldots, M \\
& u \succeq 0, \quad v \succeq 0
\end{array}
$$



hinge-loss



- an LP in a, b, u, v
- at optimum, $u_i = \max\{0, 1 - a^T x_i - b\}$, $v_i = \max\{0, 1 + a^T y_i + b\}$
- can be interpreted as a heuristic for minimizing #misclassified points

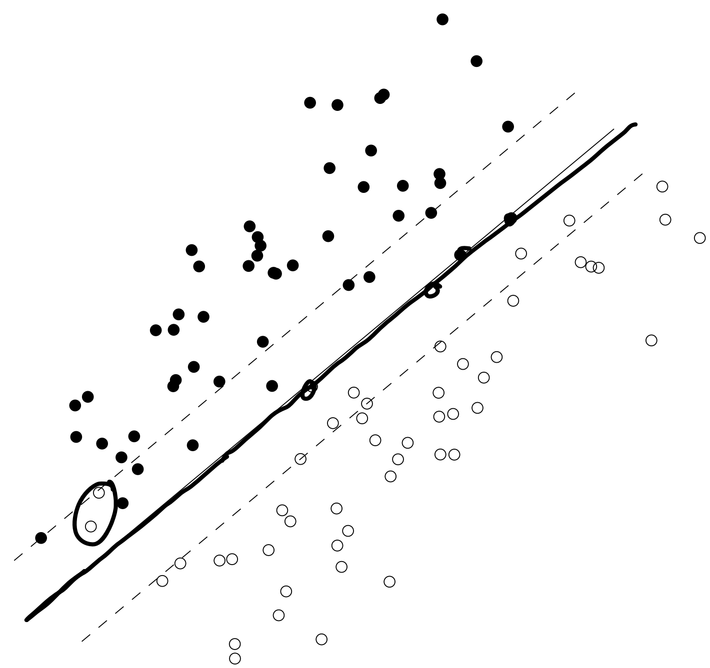


Support vector classifier (SVM)

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

produces point on trade-off curve between inverse of margin $2/\|a\|_2$ and classification error, measured by total slack $\mathbf{1}^T u + \mathbf{1}^T v$

same example as previous page,
with $\gamma = 0.1$:



Nonlinear discrimination

separate two sets of points by a nonlinear function:

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

- choose a linearly parametrized family of functions

$$f(z) = \overset{\text{linear in } \theta_i}{\underline{\underline{\theta}}^T} F(z)$$

$F = (F_1, \ldots, F_k) : \mathbf{R}^n \to \mathbf{R}^k$ are basis functions

- solve a set of linear inequalities in θ:

$$\theta^T F(\overset{\text{data}}{x_i}) \geq 1, \quad i = 1, \ldots, N, \qquad \theta^T F(y_i) \leq -1, \quad i = 1, \ldots, M$$

quadratic discrimination: $f(z) = z^T P z + q^T z + r$ (P, q, r)

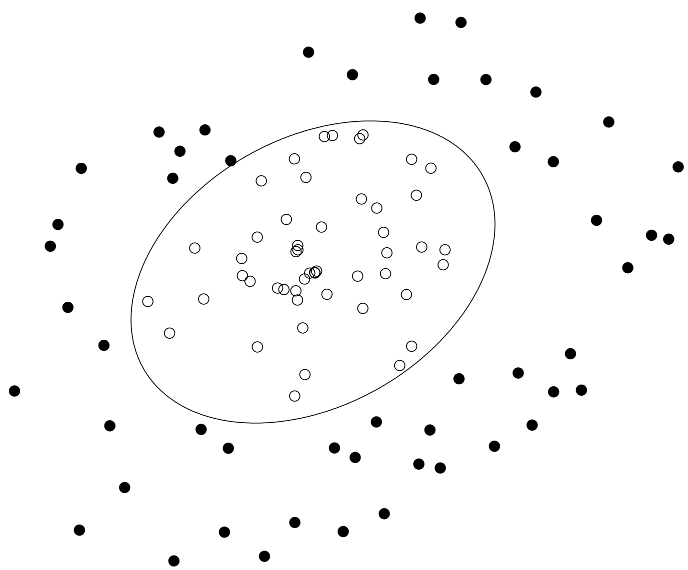
$$x_i^T P x_i + q^T x_i + r \geq 1, \qquad y_i^T P y_i + q^T y_i + r \leq -1$$

y_i's lie inside an ellipsoid

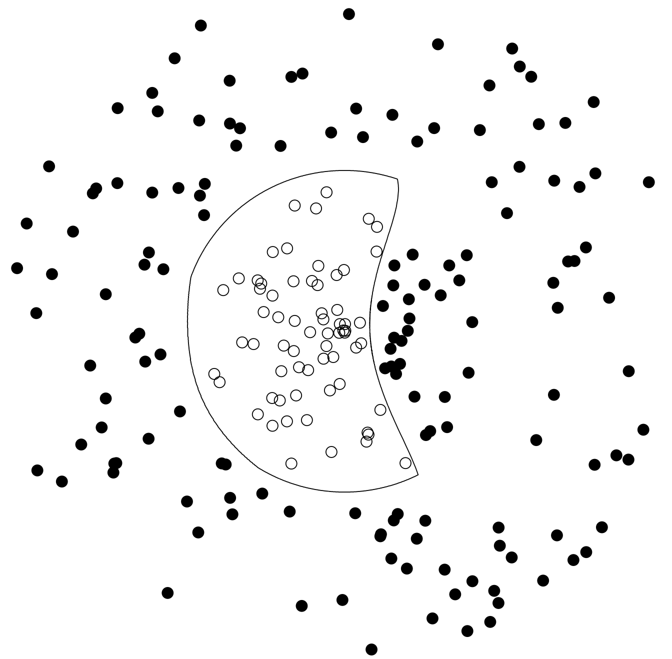
can add additional constraints (*e.g.*, $P \preceq -I$ to separate by an ellipsoid)

polynomial discrimination: $F(z)$ are all monomials up to a given degree

(Kernel methods)



separation by ellipsoid



separation by 4th degree polynomial

Placement and facility location (skipped)

- N points with coordinates $x_i \in \mathbf{R}^2$ (or $\mathbf{R}^3$)
- some positions x_i are given; the other x_i's are variables
- for each pair of points, a cost function $f_{ij}(x_i, x_j)$

placement problem

$$\text{minimize} \quad \sum_{i \neq j} f_{ij}(x_i, x_j)$$

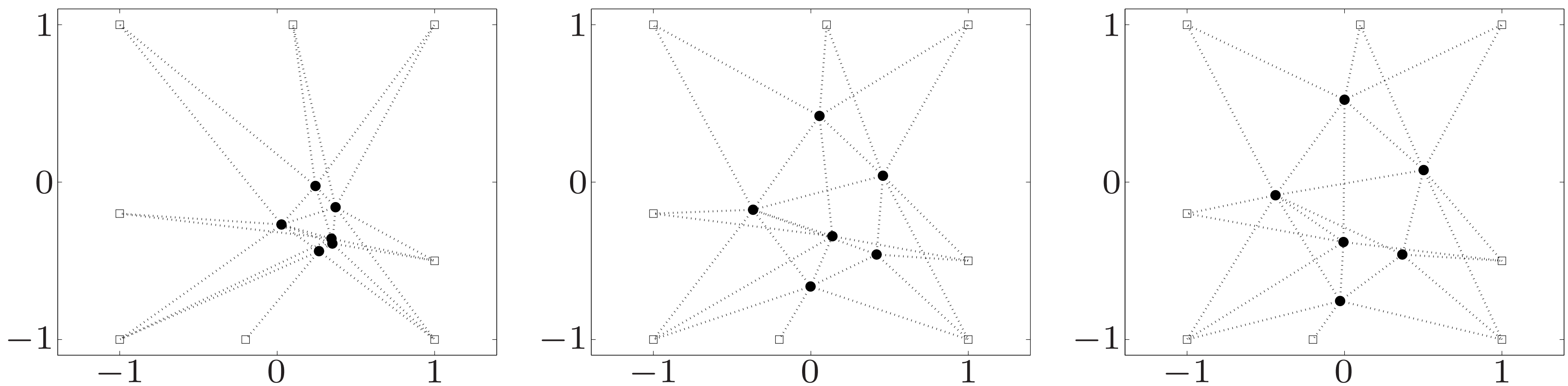
variables are positions of free points

interpretations

- points represent plants or warehouses; f_{ij} is transportation cost between facilities i and j
- points represent cells on an IC; f_{ij} represents wirelength

example: minimize $\sum_{(i,j)\in\mathcal{A}} h(\|x_i - x_j\|_2)$, with 6 free points, 27 links

optimal placement for $h(z) = z$, $h(z) = z^2$, $h(z) = z^4$



histograms of connection lengths $\|x_i - x_j\|_2$