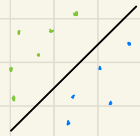
· Linear discrimination

separate 2 sets of points $\{x_1, \cdots x_N\}$ $\{y_1 \cdots y_M\}$ by a hyperplane

$a^T x_i + b > 0 \qquad a^T y_i + b < 0$

(homogeneous in a and b
hence equal to)



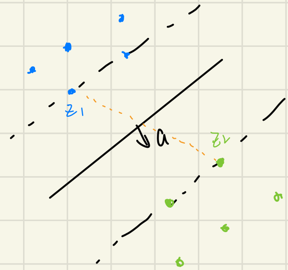
$a^T x_i + b \geq 1 \qquad a^T y_i + b \leq -1$ } LP feasibility problem

· Robust linear discrimination

(Euclidean) distance between hyperplane

$H_1 = \{z \mid a^T z + b = 1\}$

$H_2 = \{z \mid a^T z + b = -1\}$



$\text{dist}(H_1, H_2) = 2/\|a\|_2$

$a^T z_1 + b = 1 \qquad a^T z_2 + b = -1$

$a^T(z_1 - z_2) = 2 = \|a\| \cdot u_a^T(z_1 - z_2)$

$\text{gap} = u_a^T(z_1 - z_2) = 2/\|a\|_2$

$$\begin{aligned} \min.\quad & \tfrac{1}{2}\|a\|_2 \\ \text{s.t.}\quad & a^T x_i + b \geq 1 \\ & a^T y_i + b \leq -1 \end{aligned}$$

(a QP in a and b)

thickest margin between 2 sets of points

$\Updownarrow$ dual

smallest distance between the convex hulls of 2 set of points

$$\begin{aligned} \mathcal{L}(a, b, \lambda, u) &= \tfrac{1}{2}\|a\|_2 + \sum_i \lambda_i (1 - a^T x_i - b) + \sum_j u_j (a^T y_j + b + 1) \\ &= \tfrac{1}{2}\|a\|_2 + a^T\Big(\sum_j u_j y_j - \sum_i \lambda_i x_i\Big) + b(\mathbf{1}^T u - \mathbf{1}^T \lambda) + (\mathbf{1}^T u + \mathbf{1}^T \lambda) \end{aligned}$$

$$g(\lambda, u) = \inf_{a,b} \mathcal{L}(a, b, \lambda, u) = \begin{cases} -\infty & \mathbf{1}^T u \neq \mathbf{1}^T \lambda \\ -\infty & \|\sum_j u_j y_j - \sum_i \lambda_i x_i\| > \tfrac{1}{2} \\ \mathbf{1}^T u + \mathbf{1}^T \lambda & (a = 0) \end{cases}$$

$$\begin{aligned} \max.\ & \mathbb{1}^T\lambda + \mathbb{1}^T u \\ \text{s.t}\ & \|\sum_j u_j y_j - \sum_i \lambda_i x_i\|_2 \le 1/2 \\ & \mathbb{1}^T u - \mathbb{1}^T \lambda = 0 \quad \lambda \succeq 0 \quad u \succeq 0 \end{aligned}$$

$\Downarrow$ change of variable $\theta_i = \lambda_i / \mathbb{1}^T\lambda \quad \gamma_i = u_i/\mathbb{1}^T u$

$t = 1/(\mathbb{1}^T\lambda + \mathbb{1}^T u)$

$$\begin{aligned} \min.\ & t \\ \text{s.t.}\ & \|\sum_i \theta_i x_i - \sum_j \gamma_j y_j\| \le t \\ & \theta \succeq 0 \qquad \mathbb{1}^T\theta = 1 \\ & \gamma \succeq 0 \qquad \mathbb{1}^T\gamma = 1 \end{aligned}$$

θ and γ are probability distribution

$\left.\begin{array}{l} \sum_i \theta_i x_i \text{ is a point in convex hull of } x_i\text{'s} \\ \sum_j \gamma_j y_j \text{ is a point in convex hull of } y_j\text{'s} \end{array}\right\}$ distance between the 2 convex hulls

- Approximate linear separation of non-separable sets

$$\begin{aligned} \min.\ & \mathbb{1}^T u + \mathbb{1}^T v \\ \text{s.t.}\ & a^T x_i + b \ge 1 - u_i \\ & a^T y_i + b \le -1 + v_i \\ & u \succeq 0 \quad v \succeq 0 \end{aligned}$$

- SVM.

$$\begin{aligned} \min.\ & \|a\|_2 + \gamma(\mathbb{1}^T u + \mathbb{1}^T v) \\ \text{s.t.}\ & a^T x_i + b \ge 1 - u_i \\ & a^T y_i + b \le -1 + v_i \\ & u \succeq 0 \quad v \succeq 0 \end{aligned}$$

- Nonlinear discrimination

separate 2 sets of points by a non-linear function

$f(x_i) > 0 \quad f(y_i) < 0$

- choose a function that's linear in parameters

$$f(z) = \theta^T F(z)$$

$F = (F_1, \cdots F_k) : \mathbb{R}^n \to \mathbb{R}^k$ are function basis

- Solve a set of linear inequalities in θ

$$\theta^T F(x_i) \geq 1 \qquad \theta^T F(y_i) \leq -1$$

eg. quadratic discrimination

$f(z) = x^T P x + q^T x + r$ (linear in P, q, r)

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

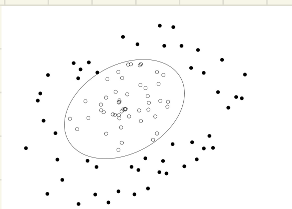
($P < 0$ to make the separating surface an ellipsoid)

find $P\ q\ r$

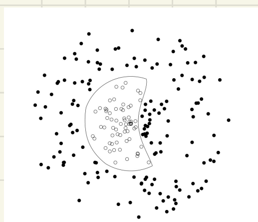
s.t. $x_i^T P x_i + q^T x_i + r \geq 1$

$y_i^T P y_i + q^T y_i + r \leq -1$

$P \leq -I$ (homogeneous in P)



quadratic with $P \succeq 0$



minimum degree polynomial discrimination in $\mathbb{R}^2$

eg. Placement and facility location

- N points with coordinates x_i
- Some position x_i's are given ; others are variables
- for each pair of points, a cost function $f_{ij}(x_i\ x_j)$

$$\min. \sum_{i \neq j} f_{ij}(x_i, x_j)$$

- Matrix structure and algorithm complexity

  cost (executing time) of solving $Ax=b$ with $A \in \mathbb{R}^{n\times n}$

  for general method, grows as n^3

  less if A is structured (banded, sparse, Toeplitz ...)

  flop counts

  flop: floating-point operations $(+\ -\ \times\ \div)$

  - Vector-vector operations $(x, y \in \mathbb{R}^n)$

    $x^T y$: $2n-1$ flops

    $x+y$, αx : n flops

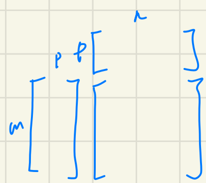
  - matrix-vector product $b = Ax$ $A \in \mathbb{R}^{m\times n}$

    $m(2n-1)$ flops

    $2N$ if A is sparse with N nonzero elements

    $2p(n+m)$ if A is given as $A = UV^T$ $U \in \mathbb{R}^{m\times p}$ $V \in \mathbb{R}^{n\times p}$

    (low rank structure)



  - matrix-matrix product $C = AB$ $A \in \mathbb{R}^{m\times n}$ $B \in \mathbb{R}^{n\times p}$

    $mp(2n-1)$ flops

    less if A and/or B sparse

- Linear equations that are easy to solve

  1. diagonal matrices

     $x = A^{-1}b$

  2. lower triangular : n^2 flops

     $x_1 := b_1/a_{11}$

     $x_2 = (b_2 - a_{21}x_1)/a_{22}$

     $x_3 = (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33}$

     (forward substitution)

     (can be written in block-lower-triangular)

$$\begin{bmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

3. orthogonal matrices: $A^{-1} = A^T$

$2n^2$ flops to compute $x = A^T b$

less with structure

eg. $A = I - 2uu^T$ with $\|u\|_2 = 1$

$x = A^T b = b - 2uu^T b$

$= b - 2(u^T b)$ 4n flops

4. permutation matrices:

$A^{-1} = A^T$, hence cost of solving $Ax = b$ is 0 flops

(data movement)

- Factor-solve method for solving $Ax = b$

1. factor A as a product of simple matrices (usually 2 or 3)

$A = A_1 A_2 \cdots A_k$

(A_i: diagonal, upper or lower triangular, etc)

2. $x = A^{-1} b = A_k^{-1} \cdots A_2^{-1} A_1^{-1} b$ by solving k 'easy' equations

$A_1 x_1 = b$ $A_2 x_2 = x_1$

cost of factorization usually dominates the cost

eg. $A \backslash b$ $n^3 + n^2$

$A \backslash [b_1 \ b_2]$ $n^3 + 2n^2$