Lecture 2
Piecewise-linear optimization

- piecewise-linear minimization
- ℓ_1- and ℓ_∞-norm approximation
- examples
- modeling software

Linear and affine functions

linear function: a function $f : \mathbf{R}^n \to \mathbf{R}$ is linear if

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \qquad \forall x, y \in \mathbf{R}^n, \alpha, \beta \in \mathbf{R}$$

property: f is linear if and only if $f(x) = a^T x$ for some a

affine function: a function $f : \mathbf{R}^n \to \mathbf{R}$ is affine if

$$f(\alpha x + (1-\alpha) y) = \alpha f(x) + (1-\alpha) f(y) \qquad \forall x, y \in \mathbf{R}^n, \alpha \in \mathbf{R}$$

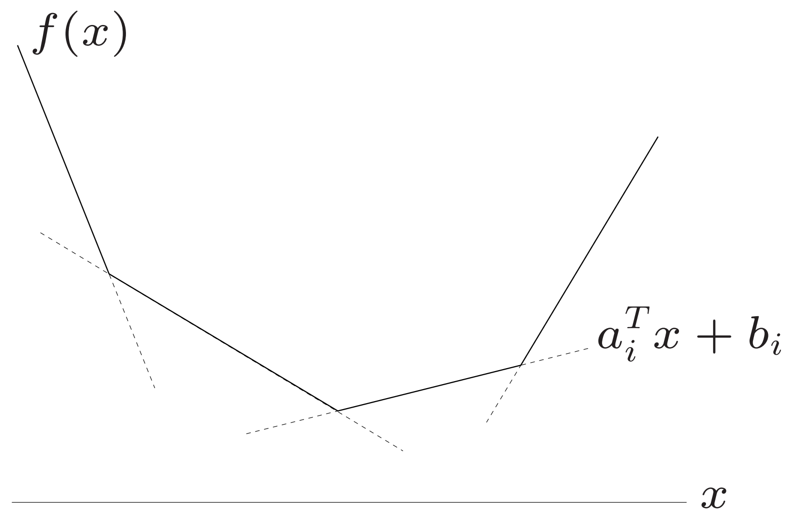
property: f is affine if and only if $f(x) = a^T x + b$ for some a, b

Piecewise-linear function

$f : \mathbf{R}^n \to \mathbf{R}$ is (convex) **piecewise-linear** if it can be expressed as

$$f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i)$$

f is parameterized by m n-vectors a_i and m scalars b_i



(the term *piecewise-affine* is more accurate but less common)

Piecewise-linear minimization

$$\text{minimize} \quad f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i)$$

- **equivalent LP** (with variables x and auxiliary scalar variable t)

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

  to see equivalence, note that for fixed x the optimal t is $t = f(x)$

- **LP in matrix notation:** minimize $\tilde{c}^T \tilde{x}$ subject to $\tilde{A}\tilde{x} \leq \tilde{b}$ with

$$\tilde{x} = \begin{bmatrix} x \\ t \end{bmatrix}, \qquad \tilde{c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad \tilde{A} = \begin{bmatrix} a_1^T & -1 \\ \vdots & \vdots \\ a_m^T & -1 \end{bmatrix}, \qquad \tilde{b} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \end{bmatrix}$$

Minimizing a sum of piecewise-linear functions

$$\text{minimize} \quad f(x) + g(x) = \max_{i=1,\dots,m} (a_i^T x + b_i) + \max_{i=1,\dots,p} (c_i^T x + d_i)$$

- **cost function is piecewise-linear:** maximum of mp affine functions

$$f(x) + g(x) = \max_{\substack{i=1,\dots,m \\ j=1,\dots,p}} \left((a_i + c_j)^T x + (b_i + d_j) \right)$$

- **equivalent LP** with $m + p$ inequalities

$$\begin{array}{ll} \text{minimize} & t_1 + t_2 \\ \text{subject to} & a_i^T x + b_i \le t_1, \quad i = 1, \dots, m \\ & c_i^T x + d_i \le t_2, \quad i = 1, \dots, p \end{array}$$

note that for fixed x, optimal t_1, t_2 are $t_1 = f(x)$, $t_2 = g(x)$

- **equivalent LP in matrix notation**

$$
\begin{array}{ll}
\text{minimize} & \tilde{c}^T \tilde{x} \\
\text{subject to} & \tilde{A}\tilde{x} \leq \tilde{b}
\end{array}
$$

with

$$
\tilde{x} = \begin{bmatrix} x \\ t_1 \\ t_2 \end{bmatrix}, \quad \tilde{c} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} a_1^T & -1 & 0 \\ \vdots & \vdots & \vdots \\ a_m^T & -1 & 0 \\ c_1^T & 0 & -1 \\ \vdots & \vdots & \vdots \\ c_p^T & 0 & -1 \end{bmatrix}, \quad \tilde{b} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \\ -d_1 \\ \vdots \\ -d_p \end{bmatrix}
$$

ℓ_∞-Norm (Cheybshev) approximation

$$\text{minimize} \quad \|Ax - b\|_\infty$$

with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$

- **ℓ_∞-norm (Chebyshev norm)** of m-vector y is

$$\|y\|_\infty = \max_{i=1,\ldots,m} |y_i| = \max_{i=1,\ldots,m} \max\{y_i, -y_i\}$$

- **equivalent LP** (with variables x and auxiliary scalar variable t)

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t\mathbf{1} \le Ax - b \le t\mathbf{1} \end{array}$$

(for fixed x, optimal t is $t = \|Ax - b\|_\infty$)

- **equivalent LP in matrix notation**

$$
\begin{array}{ll}
\text{minimize} & \left[\begin{array}{c} 0 \\ 1 \end{array}\right]^T \left[\begin{array}{c} x \\ t \end{array}\right] \\
\text{subject to} & \left[\begin{array}{cc} A & -\mathbf{1} \\ -A & -\mathbf{1} \end{array}\right] \left[\begin{array}{c} x \\ t \end{array}\right] \leq \left[\begin{array}{c} b \\ -b \end{array}\right]
\end{array}
$$

ℓ_1-Norm approximation

$$\text{minimize} \quad \|Ax - b\|_1$$

- ℓ_1-**norm** of m-vector y is

$$\|y\|_1 = \sum_{i=1}^{m} |y_i| = \sum_{i=1}^{m} \max\{y_i, -y_i\}$$

- **equivalent LP** (with variable x and auxiliary vector variable u)

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} u_i \\ \text{subject to} & -u \leq Ax - b \leq u \end{array}$$

(for fixed x, optimal u is $u_i = |(Ax - b)_i|$, $i = 1, \ldots, m$)

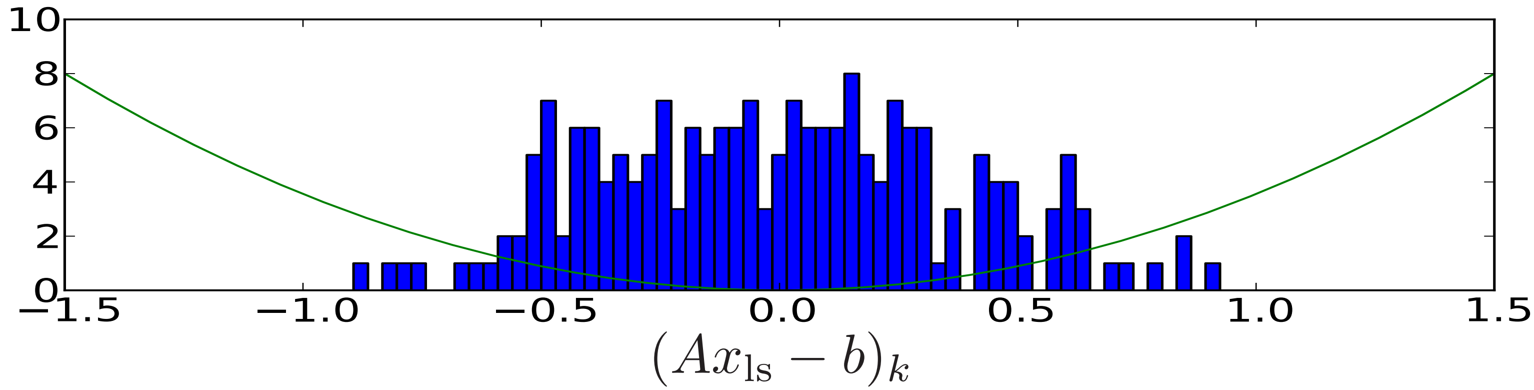
- **equivalent LP in matrix notation**

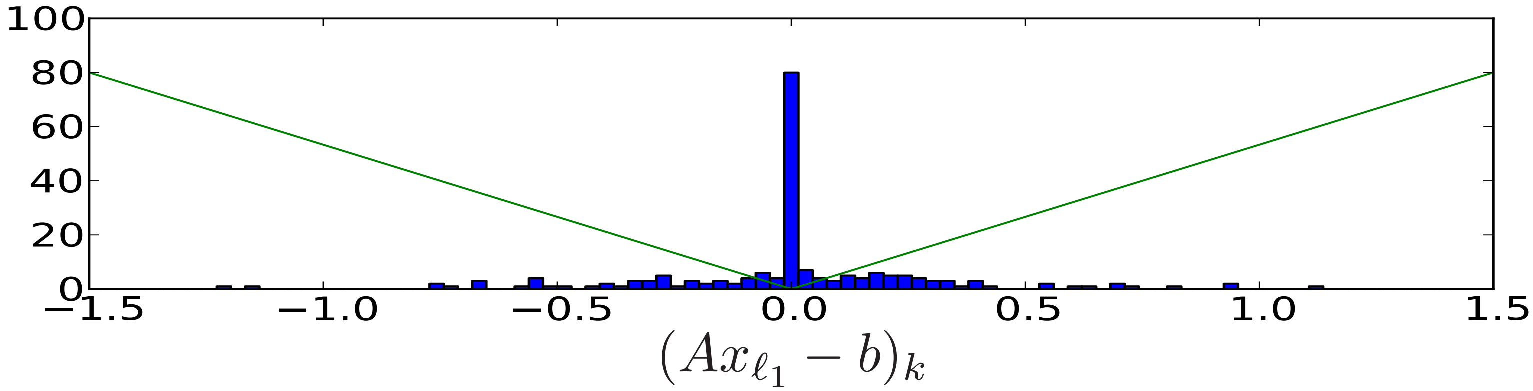
$$\begin{array}{ll} \text{minimize} & \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}^T \begin{bmatrix} x \\ u \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} A & -I \\ -A & -I \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{array}$$

Comparison with least-squares solution

histograms of residuals $Ax - b$, with randomly generated $A \in \mathbf{R}^{200\times 80}$, for

$$x_{\rm ls} = \operatorname{argmin} \|Ax - b\|, \qquad x_{\ell_1} = \operatorname{argmin} \|Ax - b\|_1$$



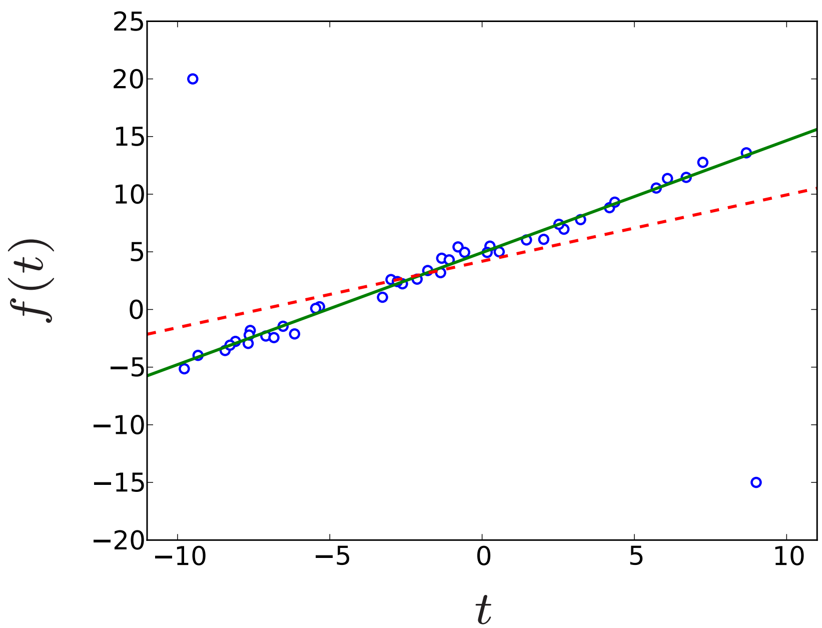


ℓ_1-norm distribution is **wider** with a **high peak at zero**

Robust curve fitting

- fit affine function $f(t) = \alpha + \beta t$ to m points (t_i, y_i)
- an approximation problem $Ax \approx b$ with

$$A = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \qquad x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \qquad b = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$



- dashed: minimize $\|Ax - b\|$
- solid: minimize $\|Ax - b\|_1$

ℓ_1-norm approximation is more robust against outliers

Sparse signal recovery via ℓ_1-norm minimization

- $\hat{x} \in \mathbf{R}^n$ is unknown signal, known to be very sparse
- we make linear measurements $y = A\hat{x}$ with $A \in \mathbf{R}^{m \times n}$, $m < n$

estimation by ℓ_1-norm minimization: compute estimate by solving

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = y \end{array}$$

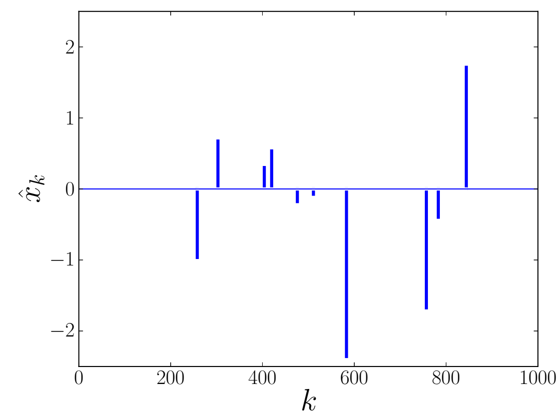
estimate is signal with smallest ℓ_1-norm, consistent with measurements

equivalent LP (variables x, $u \in \mathbf{R}^n$)

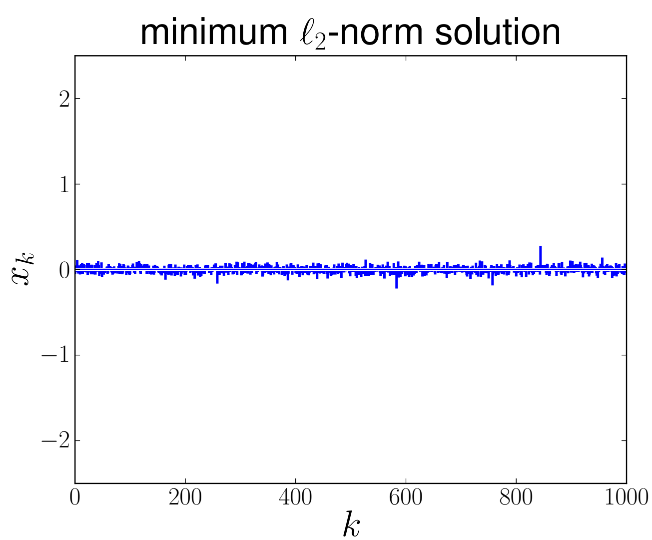
$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T u \\ \text{subject to} & -u \leq x \leq u \\ & Ax = y \end{array}$$

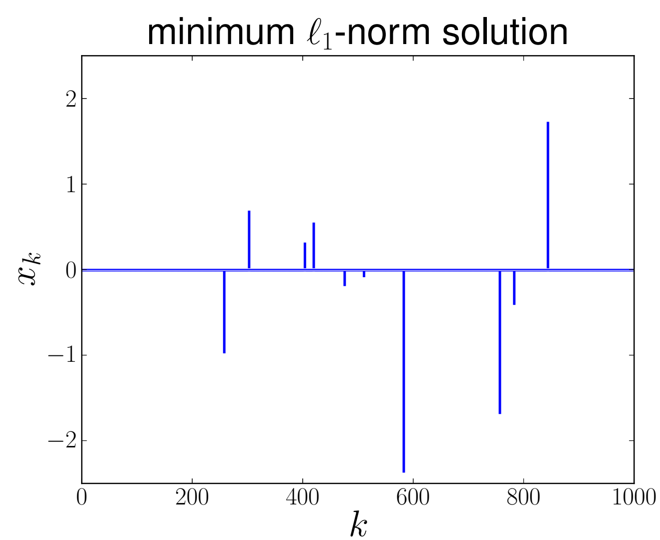
Example

- exact signal $\hat{x} \in \mathbf{R}^{1000}$
- 10 nonzero components



least-norm solutions (randomly generated $A \in \mathbf{R}^{100 \times 1000}$)





ℓ_1-norm estimate is **exact**

Exact recovery

when are the following problems equivalent?

$$\begin{array}{ll} \text{minimize} & \mathbf{card}(x) \\ \text{subject to} & Ax = y \end{array} \qquad\qquad \begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & Ax = y \end{array}$$

- $\mathbf{card}(x)$ is cardinality (number of nonzero components) of x
- depends on A and cardinality of sparsest solution of $Ax = y$

we say A allows **exact recovery** of k-sparse vectors if

$$\hat{x} = \underset{Ax=y}{\operatorname{argmin}} \|x\|_1 \qquad \text{when } y = A\hat{x} \text{ and } \mathbf{card}(\hat{x}) \leq k$$

- here, $\operatorname{argmin} \|x\|_1$ denotes the unique minimizer
- a property of (the nullspace) of the 'measurement matrix' A

'Nullspace condition' for exact recovery

necessary and sufficient condition for exact recovery of k-sparse vectors[1]

$$|z^{(1)}| + \cdots + |z^{(k)}| < \frac{1}{2}\|z\|_1 \qquad \forall z \in \mathrm{nullspace}(A) \setminus \{0\}$$

here, $z^{(i)}$ denotes component z_i in order of decreasing magnitude

$$|z^{(1)}| \geq |z^{(2)}| \geq \cdots \geq |z^{(n)}|$$

- a bound on how 'concentrated' nonzero vectors in $\mathrm{nullspace}(A)$ can be
- implies $k < n/2$
- difficult to verify for general A
- holds with high probability for certain distributions of random A

[1] Feuer & Nemirovski (IEEE Trans. IT, 2003) and several other papers on compressed sensing.

Proof of nullspace condition

notation

- x has support $I \subseteq \{1, 2, \ldots, n\}$ if $x_i = 0$ for $i \notin I$
- $|I|$ is number of elements in I
- P_I is projection matrix on n-vectors with support I: P_I is diagonal with

$$(P_I)_{jj} = \begin{cases} 1 & j \in I \\ 0 & \text{otherwise} \end{cases}$$

- A satisfies the nullspace condition if

$$\|P_I z\|_1 < \frac{1}{2}\|z\|_1$$

for all nonzero z in $\mathrm{nullspace}(A)$ and for all support sets I with $|I| \leq k$

sufficiency: suppose A satisfies the nullspace condition

- let $\hat{x}$ be k-sparse with support I ($i.e.$, with $P_I\hat{x} = \hat{x}$); define $y = A\hat{x}$
- consider any feasible x ($i.e.$, satisfying $Ax = y$), different from $\hat{x}$
- define $z = x - \hat{x}$; this is a nonzero vector in $\mathrm{nullspace}(A)$

$$\begin{aligned}
\|x\|_1 &= \|\hat{x} + z\|_1 \\
&\geq \|\hat{x} + z - P_I z\|_1 - \|P_I z\|_1 \\
&= \sum_{k\in I} |\hat{x}_k| + \sum_{k\notin I} |z_k| - \|P_I z\|_1 \\
&= \|\hat{x}\|_1 + \|z\|_1 - 2\|P_I z\|_1 \\
&> \|\hat{x}\|_1
\end{aligned}$$

(line 2 is the triangle inequality; the last line is the nullspace condition)

therefore $\hat{x} = \mathrm{argmin}_{Ax=y} \|x\|_1$

necessity: suppose A does not satisfy the nullspace condition

- for some nonzero $z \in \mathrm{nullspace}(A)$ and support set I with $|I| \leq k$,

$$\|P_I z\|_1 \geq \frac{1}{2}\|z\|_1$$

- define a k-sparse vector $\hat{x} = -P_I z$ and $y = A\hat{x}$
- the vector $x = \hat{x} + z$ satisfies $Ax = y$ and has ℓ_1-norm

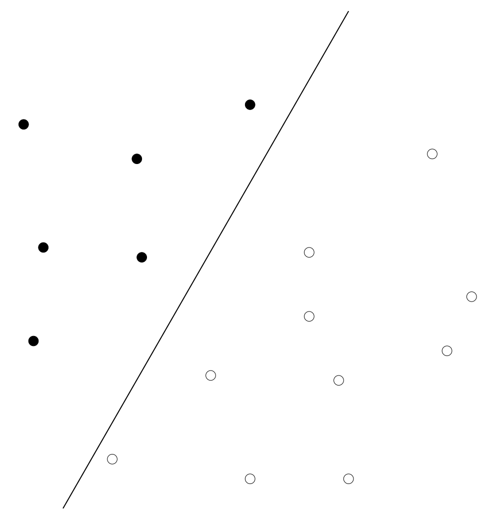
$$\begin{aligned} \|x\|_1 &= \|-P_I z + z\|_1 \\ &= \|z\|_1 - \|P_I z\|_1 \\ &\leq 2\|P_I z\|_1 - \|P_I z\|_1 \\ &= \|\hat{x}\|_1 \end{aligned}$$

therefore $\hat{x}$ is not the unique ℓ_1-minimizer

Linear classification

- given a set of points $\{v_1, \ldots, v_N\}$ with binary labels $s_i \in \{-1, 1\}$
- find hyperplane that strictly separates the two classes

$$a^T v_i + b > 0 \quad \text{if } s_i = 1$$
$$a^T v_i + b < 0 \quad \text{if } s_i = -1$$

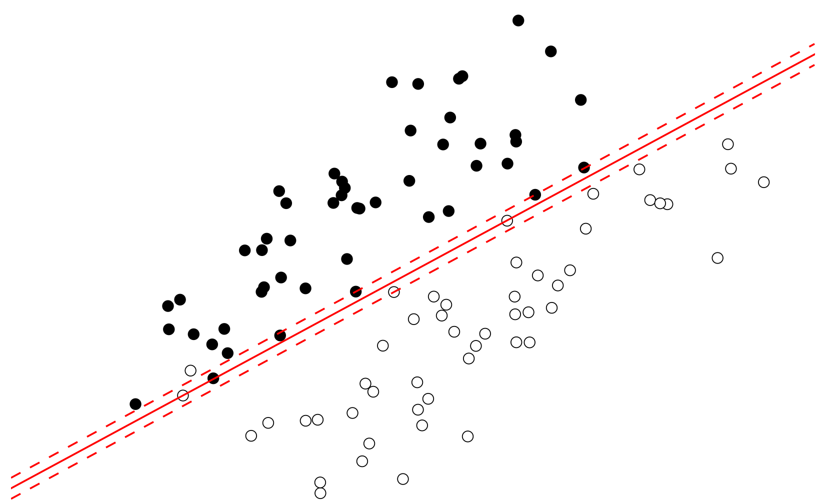


homogeneous in a, b, hence equivalent to the linear inequalities (in a, b)

$$s_i(a^T v_i + b) \geq 1, \quad i = 1, \ldots, N$$

Approximate linear separation of non-separable sets

$$\text{minimize} \quad \sum_{i=1}^{N} \max\{0, 1 - s_i(a^T v_i + b)\}$$



- penalty $1 - s_i(a_i^T v_i + b)$ for misclassifying point v_i
- can be interpreted as a heuristic for minimizing #misclassified points
- a piecewise-linear minimization problem with variables a, b

equivalent LP (variables $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, $u \in \mathbf{R}^N$)

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} u_i \\ \text{subject to} & 1 - s_i(v_i^T a + b) \leq u_i, \quad i = 1, \ldots, N \\ & u_i \geq 0, \quad i = 1, \ldots, N \end{array}$$

in matrix notation:

$$\begin{array}{ll} \text{minimize} & \left[\begin{array}{c} 0 \\ 0 \\ \mathbf{1} \end{array}\right]^T \left[\begin{array}{c} a \\ b \\ u \end{array}\right] \\ \text{subject to} & \left[\begin{array}{cccccc} -s_1 v_1^T & -s_1 & -1 & 0 & \cdots & 0 \\ -s_2 v_2^T & -s_2 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -s_N v_N^T & -s_N & 0 & 0 & \cdots & -1 \\ 0 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 \end{array}\right] \left[\begin{array}{c} a \\ b \\ u_1 \\ u_2 \\ \vdots \\ u_N \end{array}\right] \leq \left[\begin{array}{c} -1 \\ -1 \\ \vdots \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{array}\right] \end{array}$$

Modeling software

modeling tools simplify the formulation of LPs (and other problems)

- accept optimization problem in standard notation ($\max$, $\|\cdot\|_1$, . . .)
- recognize problems that can be converted to LPs
- express the problem in the input format required by a specific LP solver

examples of modeling packages

- AMPL, GAMS
- CVX, YALMIP (MATLAB)
- CVXPY, Pyomo, CVXOPT (Python)

CVX example

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_1 \\ \text{subject to} & 0 \leq x_k \leq 1, \quad k = 1, \ldots, n \end{array}$$

MATLAB code

```
cvx_begin
    variable x(n);
    minimize( norm(A*x - b, 1) )
    subject to
        x >= 0
        x <= 1
cvx_end
```

- between `cvx_begin` and `cvx_end`, `x` is a CVX variable
- after execution, `x` is MATLAB variable with optimal solution